Data Analysis: A 3-D Perspective

Balasubramaniam Jayaram

Department of Mathematics

Indian Institute of Technology Hyderabad

FSTA 2014 - 30, January

भारतीय प्रौद्योगिकी संस्थान हैदराबाद

Data Analysis: A 3-D Perspective

Data Analysis: A 3-D Perspective

Data Analysis: A 3-D Perspective

Dimension,

Data Analysis: A 3-D Perspective

Dimension, **D**istance and

Data Analysis: A 3-D Perspective

Dimension, Distance and Distribution.

Data Analysis: A 3-D Perspective

Dimension, **D**istance and **D**istribution.

Who is your nearest neighbour?

Who is your closest friend?

Who is your closest friend?

Does it even make sense?

Who is your closest friend?

Does it even make sense?

3-D's in Data Analysis

3-D's in Data Analysis

1. Dimensions

High Dimensional Data

High Dimensional Data

Occurence

High Dimensional Data

Occurence

- Images -

High Dimensional Data

Occurence

- Images - 256×256 resolution -

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions $\Longrightarrow$ More information

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions $\Longrightarrow$ More information
  $\Longrightarrow$ Better Inference

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions $\Longrightarrow$ More information

  $\Longrightarrow$ Better Inference

- Ex: Diagnosis of a disease.

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions $\Longrightarrow$ More information
  $\Longrightarrow$ Better Inference
- Ex: Diagnosis of a disease.

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions $\Longrightarrow$ More information
  $\Longrightarrow$ Better Inference
- Ex: Diagnosis of a disease.

Why is large m a problem?

High Dimensional Data

Occurence

- Images - 256×256 resolution -

  No. of dimensions $m = 65536$

- Gene Expression data - $m = 10$'s of thousands

Need

- More dimensions $\Longrightarrow$ More information
  $\Longrightarrow$ Better Inference
- Ex: Diagnosis of a disease.

Why is large m a problem?

This talk: At a Glance!

This talk: At a Glance!

Dimension

This talk: At a Glance!

Dimension

Distribution

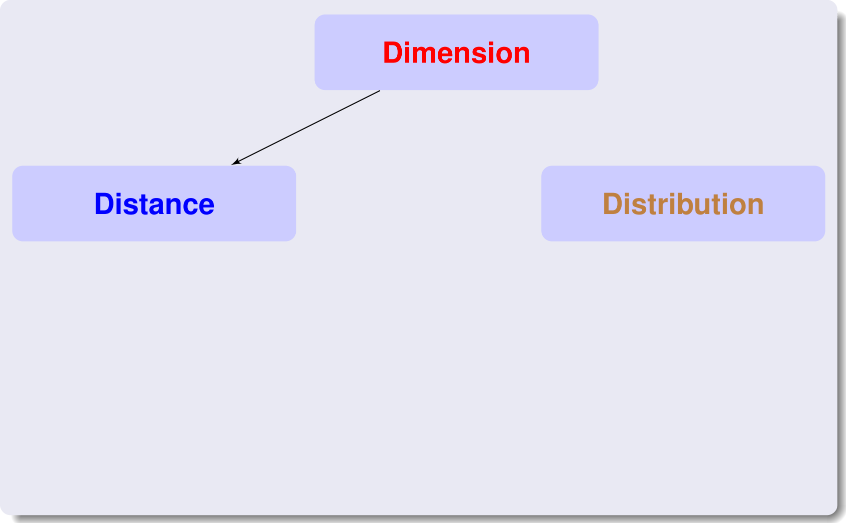
This talk: At a Glance!

Dimension

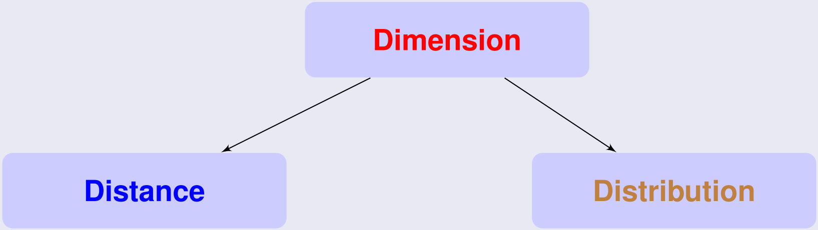
Distance

Distribution

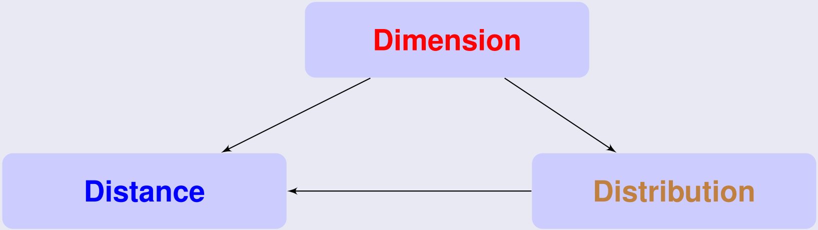
This talk: At a Glance!



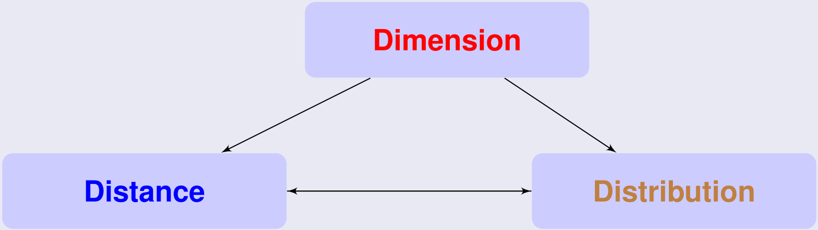
This talk: At a Glance!



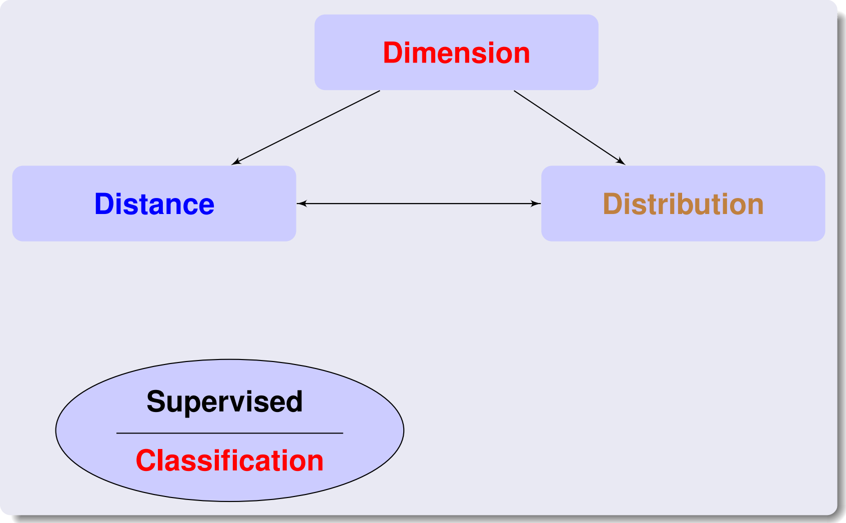
This talk: At a Glance!



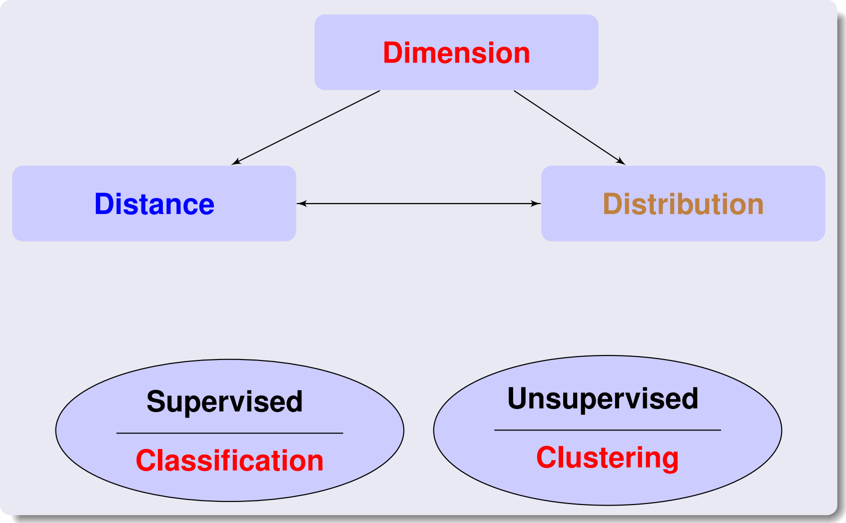
This talk: At a Glance!



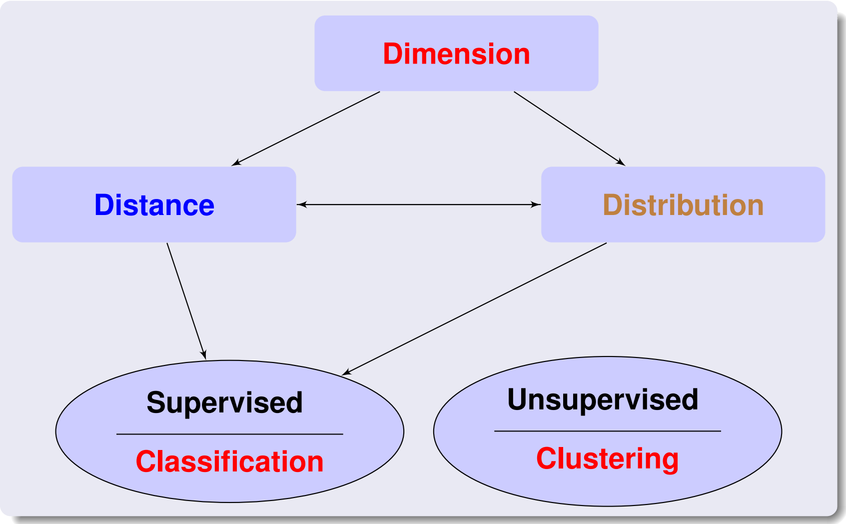
This talk: At a Glance!



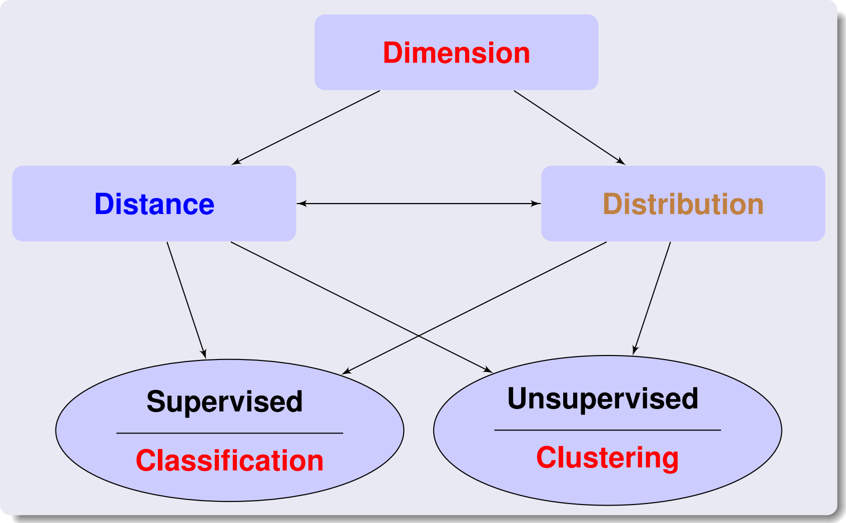
This talk: At a Glance!



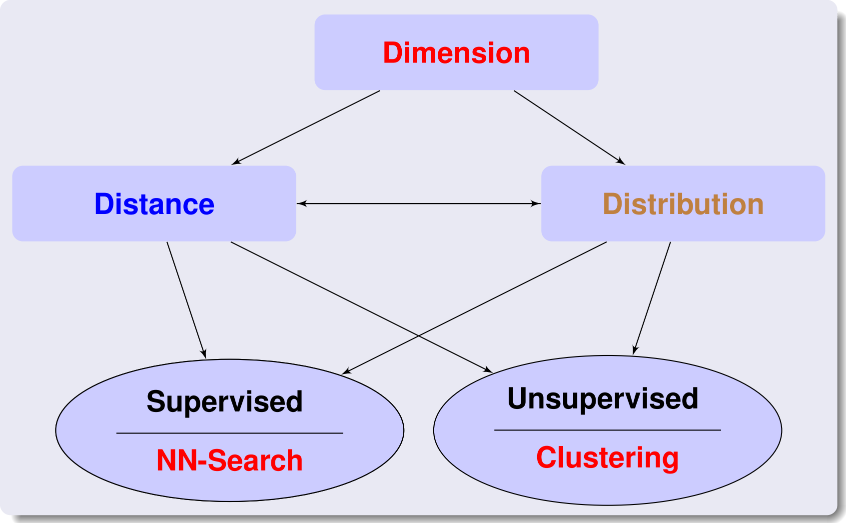
This talk: At a Glance!



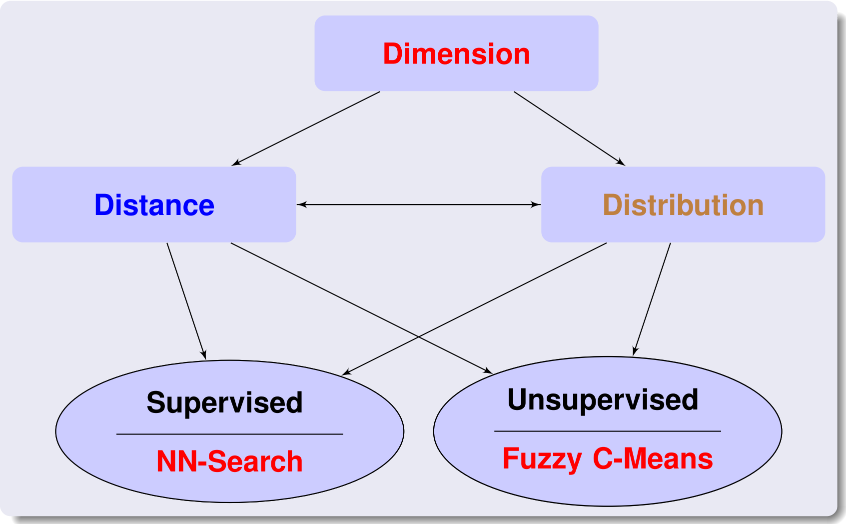
This talk: At a Glance!



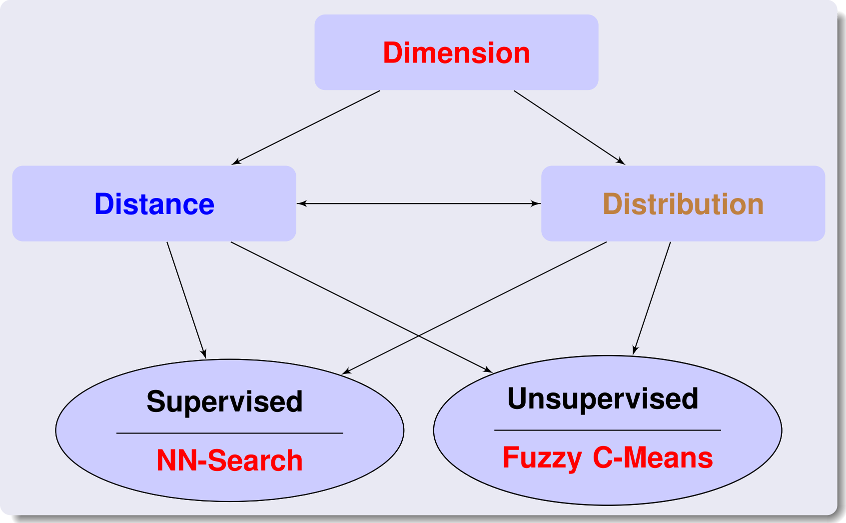
This talk: At a Glance!



This talk: At a Glance!



This talk: At a Glance!



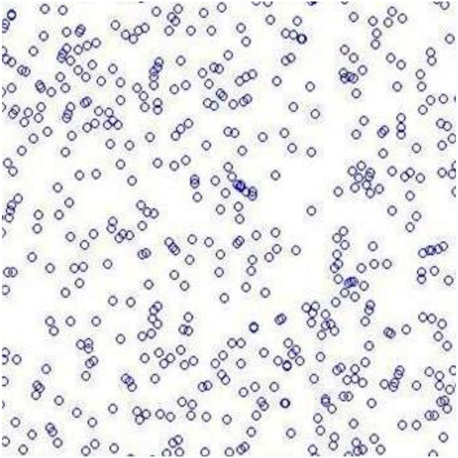
Algorithms in Data Analysis

Algorithms in Data Analysis

Query Searching & Clustering

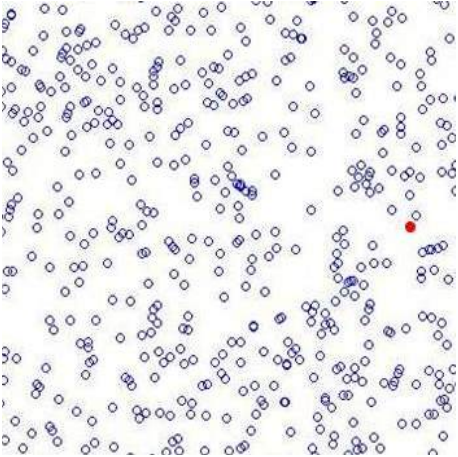
Nearest Neighbour Search - A Pictorial Example

Consider the data points in the figure.



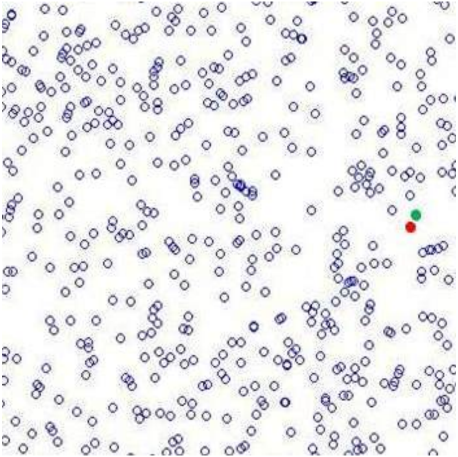
Nearest Neighbour Search - A Pictorial Example

Consider a query point in **red**.



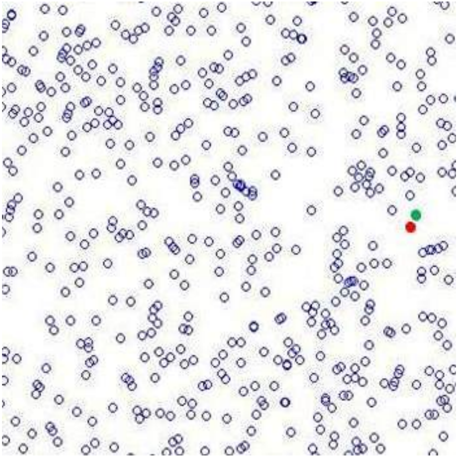
Nearest Neighbour Search - A Pictorial Example

Its nearest neighbour is the point in **green**



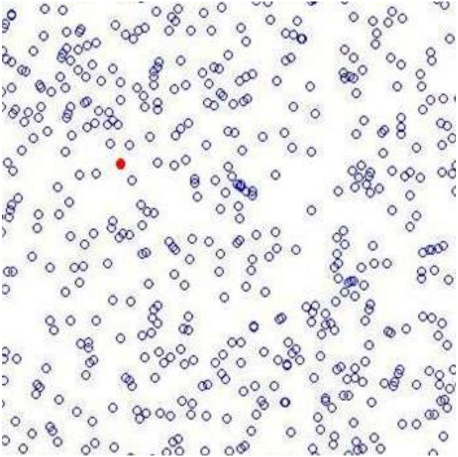
Nearest Neighbour Search - A Pictorial Example

$$NN_1(\bar{x}) = \underset{\bar{x}_i \in \mathcal{X}}{\arg\min}\{\|\bar{x} - \bar{x}_i\|\}.$$



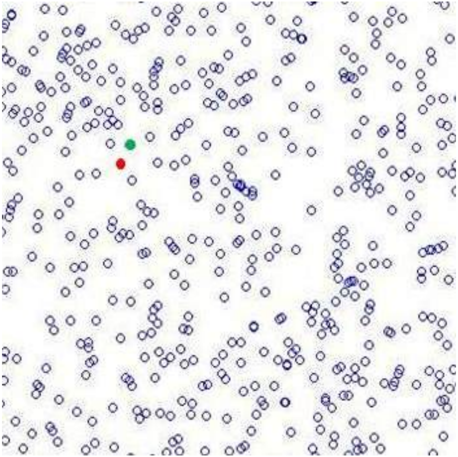
Nearest Neighbour Search - A Pictorial Example

Consider another query point in **red**.



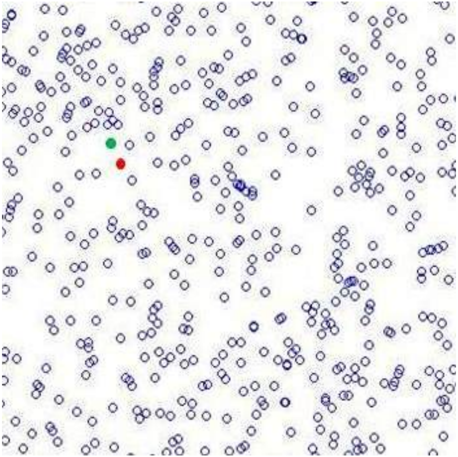
Nearest Neighbour Search - A Pictorial Example

Who is his nearest neighbour?



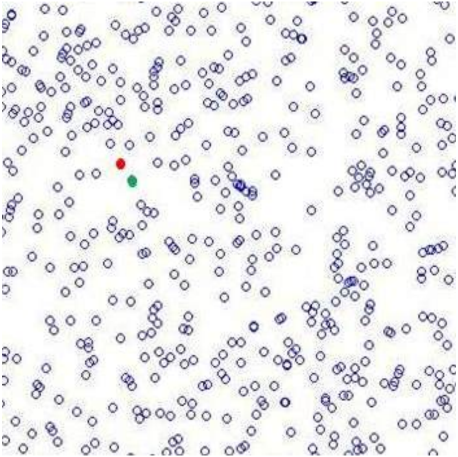
Nearest Neighbour Search - A Pictorial Example

Who is his nearest neighbour?



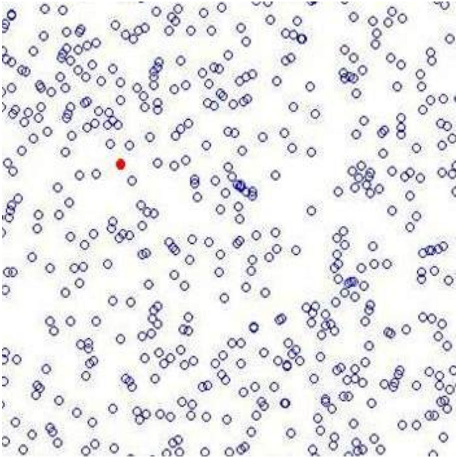
Nearest Neighbour Search - A Pictorial Example

Who is his nearest neighbour?



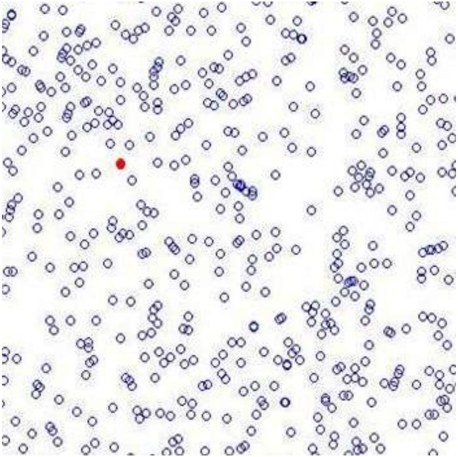
k-NN Search - A Pictorial Example

Find k-nearest neighbours of the query point in **red**.



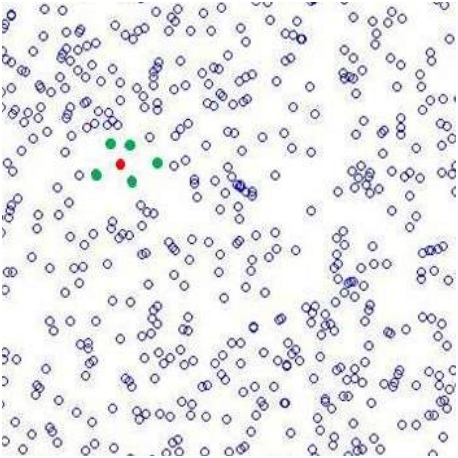
k-NN Search - A Pictorial Example

$NN_k(\bar{x})$ - k-nearest neighbours of $\bar{x}$.



k-NN Search - A Pictorial Example

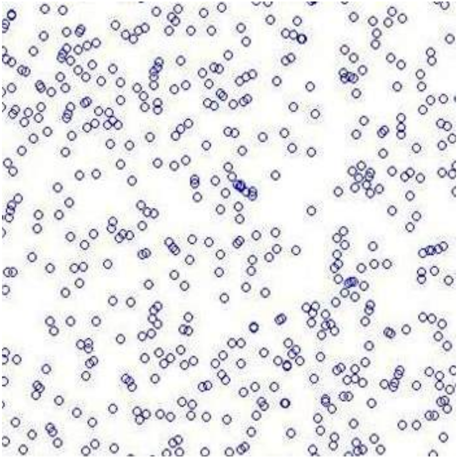
Find 5-nearest neighbours of the query point in **red**.



ε-Neighbourhood - A Pictorial Example

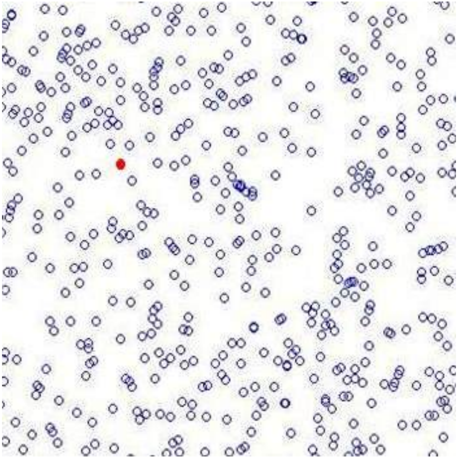
ε-Neighbourhood - A Pictorial Example

$$NN^{\varepsilon}(\bar{x}) = \{\bar{y} \in \mathcal{X} \mid \|\bar{x} - \bar{y}\| < \varepsilon\}.$$



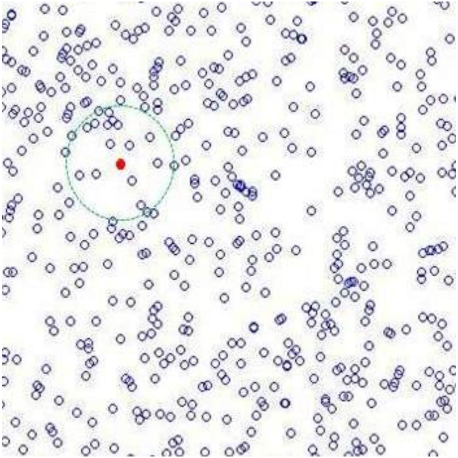
ε-Neighbourhood - A Pictorial Example

$$NN^{\varepsilon}(\bar{x}) = \{\bar{y} \in \mathcal{X} \mid \|\bar{x} - \bar{y}\| < \varepsilon\}.$$



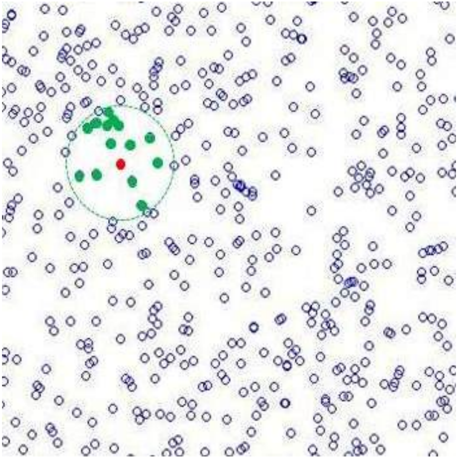
ε-Neighbourhood - A Pictorial Example

$$NN^{\varepsilon}(\bar{x}) = \{\bar{y} \in \mathcal{X} \mid \|\bar{x} - \bar{y}\| < \varepsilon\}.$$



ε-Neighbourhood - A Pictorial Example

$$NN^{\varepsilon}(\bar{x}) = \{\bar{y} \in \mathcal{X} \mid \|\bar{x} - \bar{y}\| < \varepsilon\}.$$



Data Clustering

Clustering - General Idea

Clustering - General Idea

- Minimise intra-cluster distances.

Data Clustering

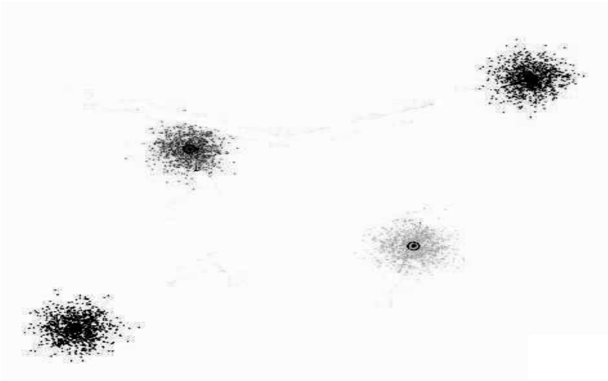
Clustering - General Idea

- Minimise intra-cluster distances.
- Maximise inter-cluster distances.

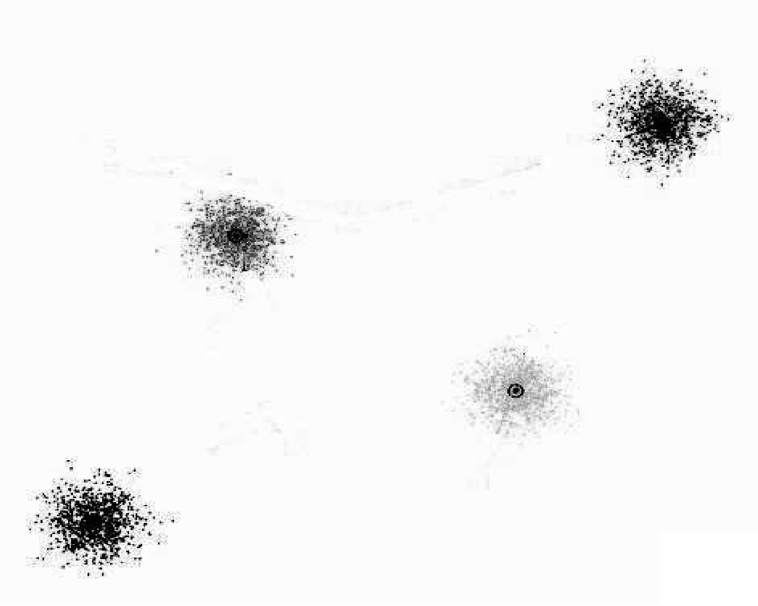
Data Clustering

Clustering - General Idea

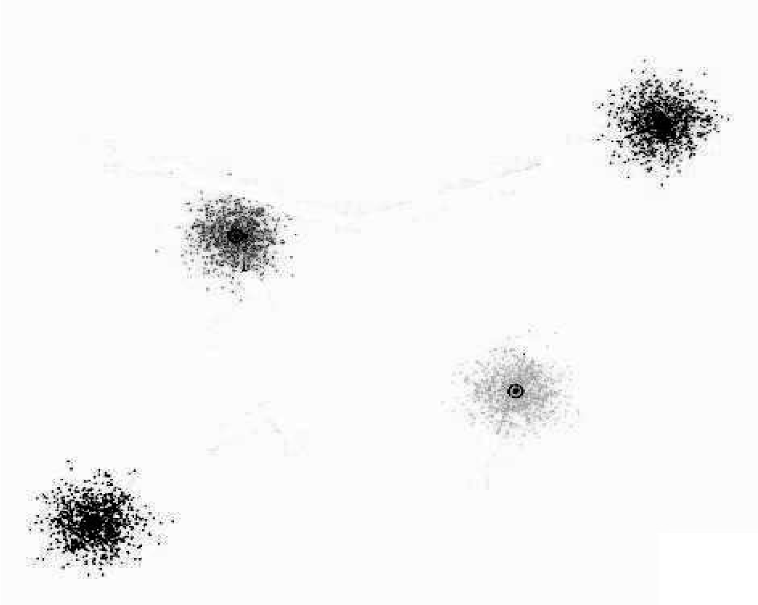
- Minimise intra-cluster distances.
- Maximise inter-cluster distances.



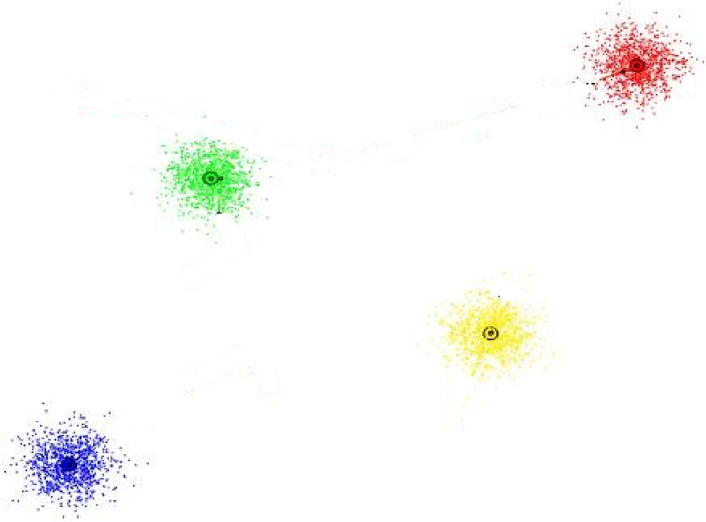
Data to be clustered



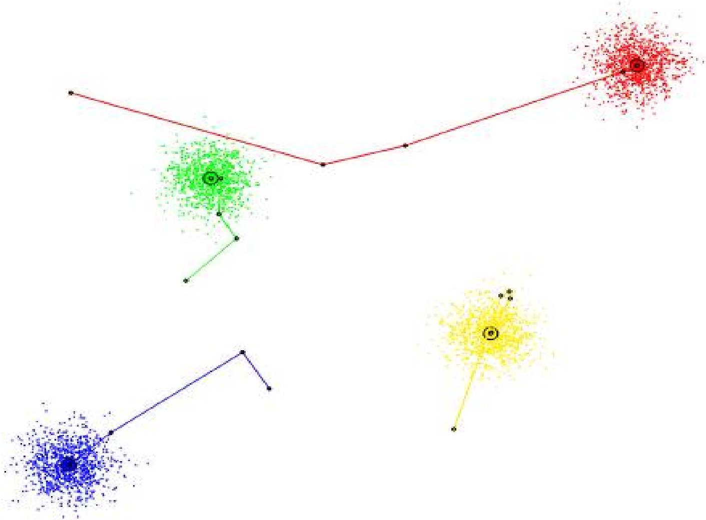
Fuzzy C-Means



Fuzzy C-Means



Fuzzy C-Means - In Low Dimensions



Issues with Fuzzy C-Means

FCM Collapses in High Dimensions

Issues with Fuzzy C-Means

FCM Collapses in High Dimensions

- Winkler et al. (2010) investigated FCM in High-Dimension

Issues with Fuzzy C-Means

FCM Collapses in High Dimensions

- Winkler et al. (2010) investigated FCM in High-Dimension
- **It collapses** !!!

Issues with Fuzzy C-Means

FCM Collapses in High Dimensions

- Winkler et al. (2010) investigated FCM in High-Dimension
- **It collapses** !!!
- All the prototypes move towards the center of the data

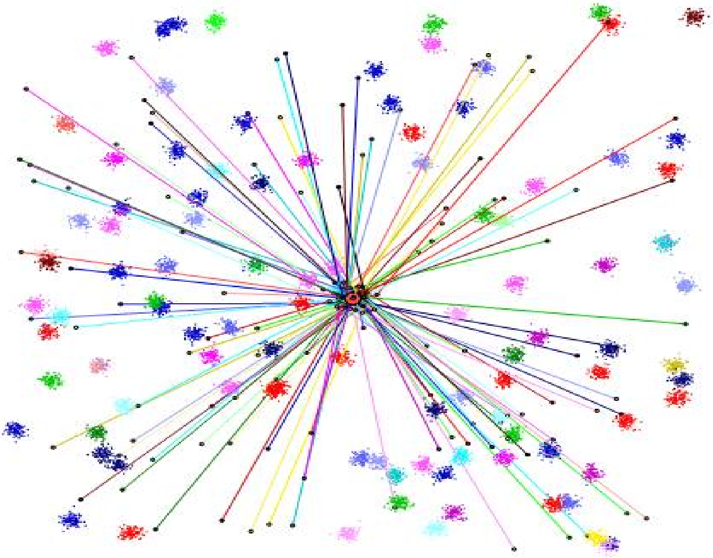
FCM Collapses in High Dimensions

- Winkler et al. (2010) investigated FCM in High-Dimension
- **It collapses** !!!
- All the prototypes move towards the center of the data
- Converges only if properly initialised !!

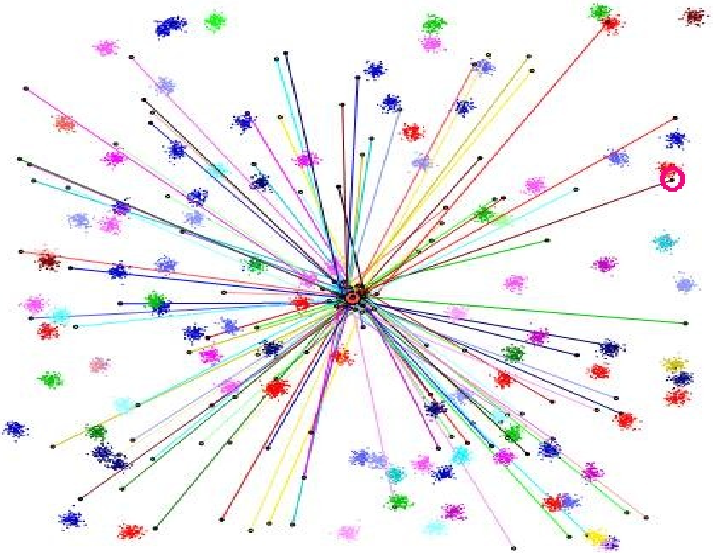
FCM Collapses in High Dimensions

- Winkler et al. (2010) investigated FCM in High-Dimension
- **It collapses** !!!
- All the prototypes move towards the center of the data
- Converges only if properly initialised !!

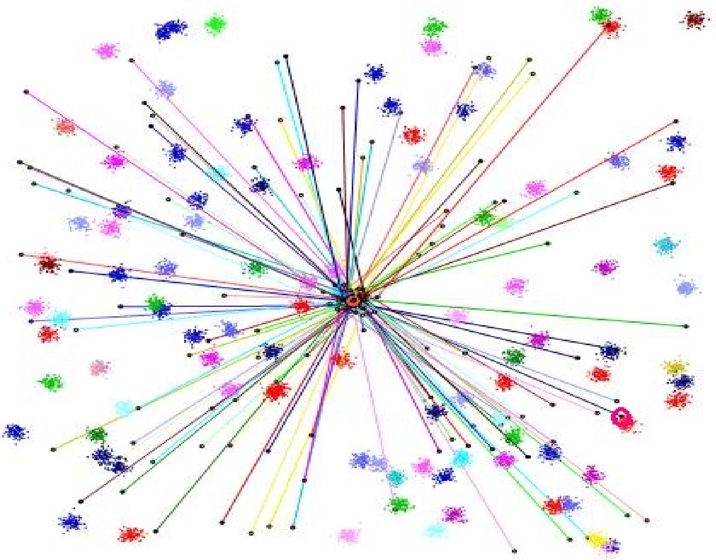
Fuzzy C-Means - In High Dimensions



Fuzzy C-Means - In High Dimensions



Fuzzy C-Means - In High Dimensions



Data Analysis - A Math Challenge!!

Issues in High Dimensional Data

Data Analysis - A Math Challenge!!

- Hilbert's Lecture at ICM **1900**.

Issues in High Dimensional Data

Data Analysis - A Math Challenge!!

- Hilbert's Lecture at ICM **1900**.
- **Prof. David Donoho** - August 8, **2000** by **AMS**.

Issues in High Dimensional Data

Data Analysis - A Math Challenge!!

- Hilbert's Lecture at ICM **1900**.
- **Prof. David Donoho** - August 8, **2000** by **AMS**.
- Math Challenges of the 21st Century.

Issues in High Dimensional Data

Data Analysis - A Math Challenge!!

- Hilbert's Lecture at ICM **1900**.
- **Prof. David Donoho** - August 8, **2000** by **AMS**.
- Math Challenges of the 21st Century.
- **Curse of Dimensionality**.

Data Analysis - A Math Challenge!!

- Hilbert's Lecture at ICM **1900**.
- **Prof. David Donoho** - August 8, **2000** by **AMS**.
- Math Challenges of the 21st Century.
- **Curse of Dimensionality**.

Curse of Dimensionality

1.

Curse of Dimensionality

1. Combinatorial Explosion in Search Space

Curse of Dimensionality - Aspect 1

Curse of Dimensionality - Aspect 1

Combinatorial Explosion in Search Space

Combinatorial Explosion in Search Space

- Richard Bellman (1961).

Curse of Dimensionality - Aspect 1

Combinatorial Explosion in Search Space

- Richard Bellman (1961).
- If we have m variables each of which can take n_i values ...

Combinatorial Explosion in Search Space

- Richard Bellman (1961).
- If we have m variables each of which can take n_i values ...
- ... then there is a total of $n_1 \times \ldots \times n_m$ possibilities.

Curse of Dimensionality - Aspect 1

Combinatorial Explosion in Search Space

- Richard Bellman (1961).
- If we have m variables each of which can take n_i values ...
- ... then there is a total of $n_1 \times \ldots \times n_m$ possibilities.
- Think of finding the joint probabilities $P(X_1, X_2, \ldots, X_m)$!!

Curse of Dimensionality - Aspect 1

Combinatorial Explosion in Search Space

- Richard Bellman (1961).
- If we have m variables each of which can take n_i values ...
- ... then there is a total of $n_1 \times \ldots \times n_m$ possibilities.
- Think of finding the joint probabilities $P(X_1, X_2, \ldots, X_m)$!!
- *Think of a complete fuzzy If-Then rule base* !!

Curse of Dimensionality - Aspect 1

Combinatorial Explosion in Search Space

- Richard Bellman (1961).
- If we have m variables each of which can take n_i values ...
- ... then there is a total of $n_1 \times \ldots \times n_m$ possibilities.
- Think of finding the joint probabilities $P(X_1, X_2, \ldots, X_m)$!!
- *Think of a complete fuzzy If-Then rule base* !!

What does it affect?

Increases Computational Complexity.

Curse of Dimensionality

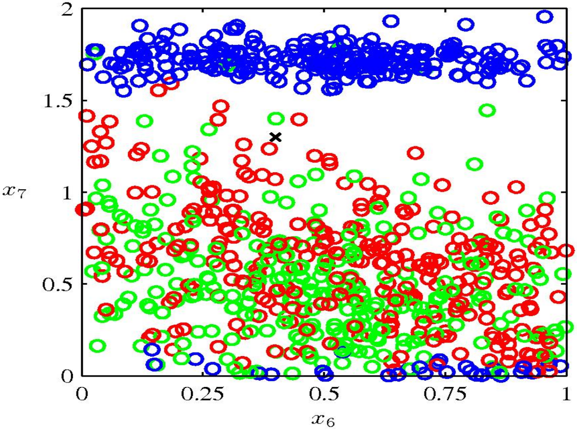
2.

Curse of Dimensionality

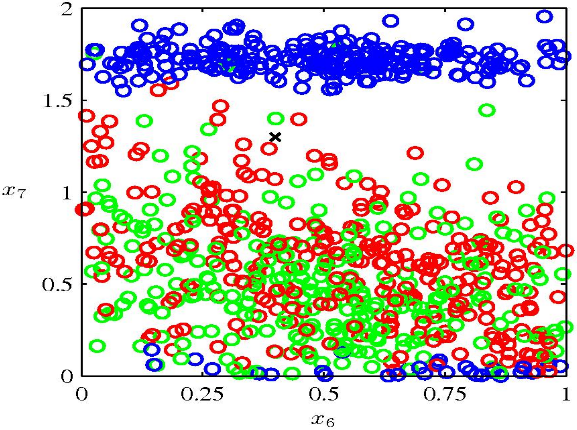
2. Need for Greed

Curse of Dimensionality - Aspect 2

Classify the point marked x.

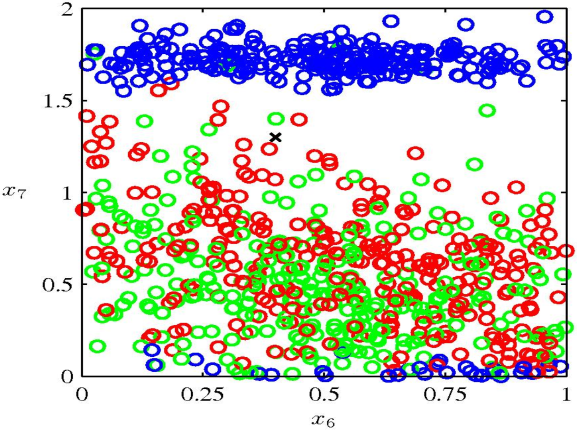


Classify the point marked x.



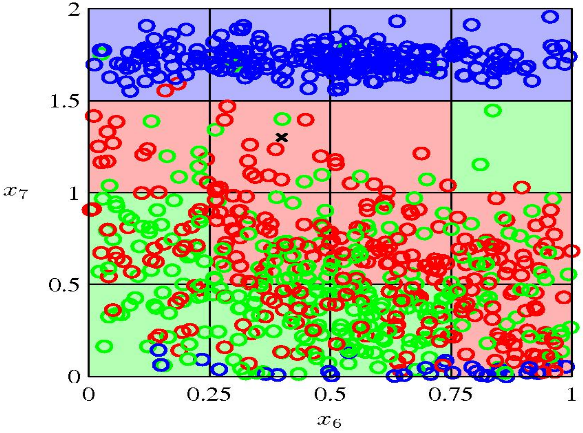
Curse of Dimensionality - Aspect 2

Partition the space into grids.



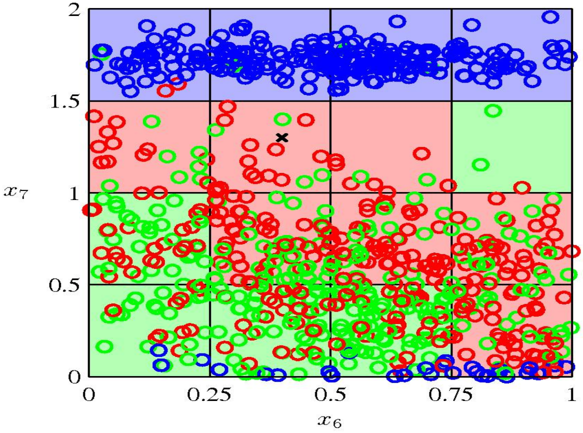
Curse of Dimensionality - Aspect 2

Partition the space into grids.

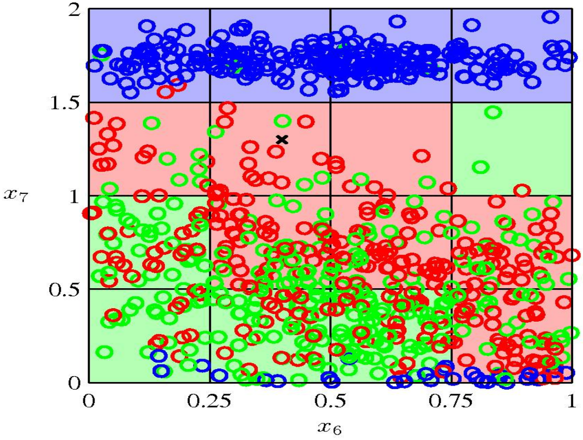


Curse of Dimensionality - Aspect 2

Partition the space into grids.

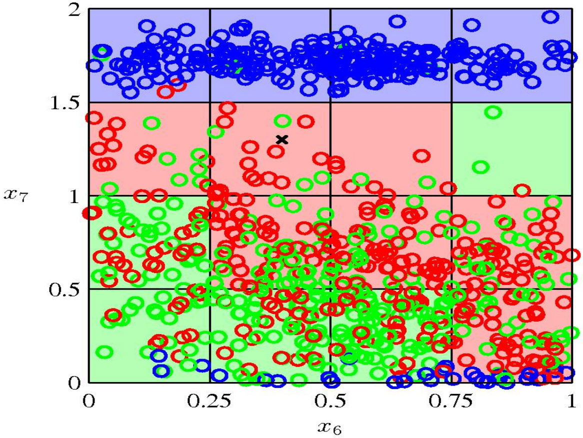


Count the no. of points in each class in the grid containing x.



Curse of Dimensionality - Aspect 2

Assign x to the class having most points in the same grid.

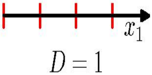


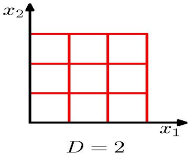
2. Need for Greed - $(N \gg m)$

2. Need for Greed - $(N \gg m)$

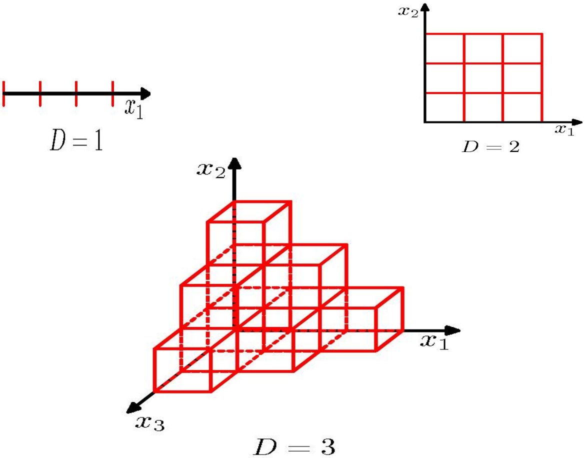
$D = 1$

2. Need for Greed - $(N \gg m)$

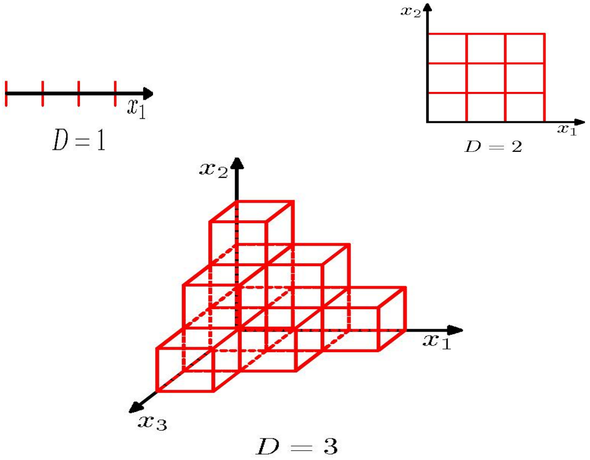




2. Need for Greed - $(N \gg m)$

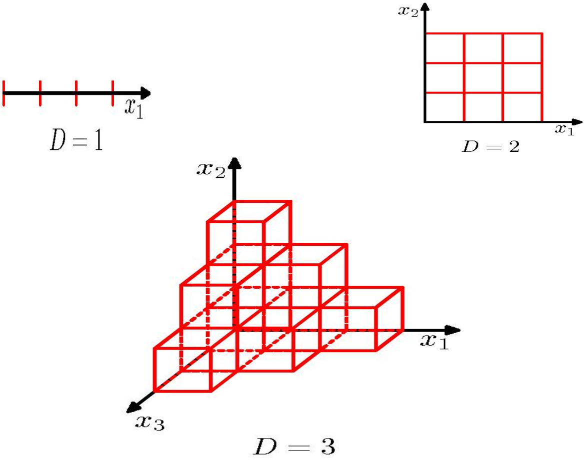


2. Need for Greed - $(N \gg m)$



As m increases, the no. of points N also should increase!

2. Need for Greed - $(N \gg m)$



Increases Storage Complexity.

Curse of Dimensionality

3.

Curse of Dimensionality

3. The Empty Space Phenomenon

Curse of Dimensionality - Aspect 3

Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]$.

Curse of Dimensionality - Aspect 3

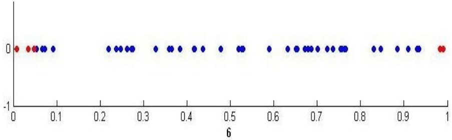
- Consider uniformly distributed data on $[0, 1]$.
- Consider data at the edges,

Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]$.
- Consider data at the edges, (distance < 0.05 from the edges).

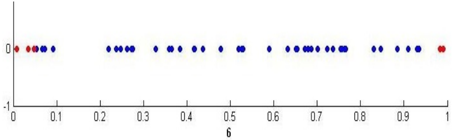
Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]$.
- Consider data at the edges, (distance < 0.05 from the edges).



Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]$.
- Consider data at the edges, (distance < 0.05 from the edges).



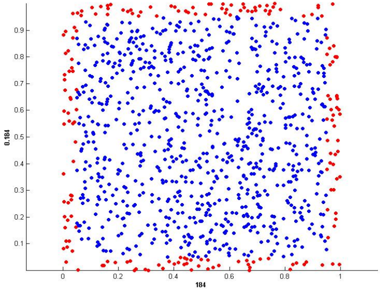
Roughly 10% of the data are in the edges.

Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]^2$.

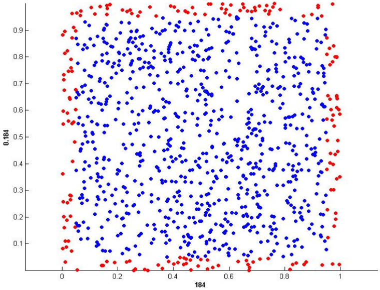
Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]^2$.



Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]^2$.



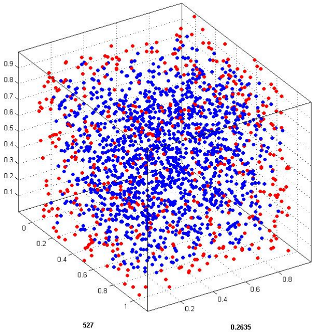
Roughly 19% of the data are in the edges.

Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0,1]^3$.

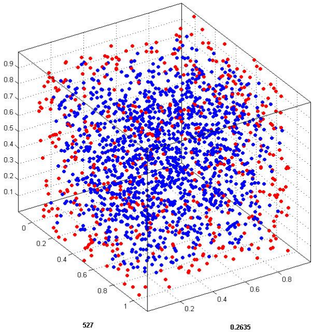
Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]^3$.



Curse of Dimensionality - Aspect 3

- Consider uniformly distributed data on $[0, 1]^3$.



Roughly 27% of the data are in the edges.

3. The Empty Space Phenomenon

3. The Empty Space Phenomenon

- How many data points are close to the edges?

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D:

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

- In 2-D:

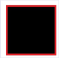
- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$



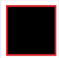
3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D:

3. The Empty Space Phenomenon

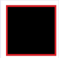
- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$



- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$

3. The Empty Space Phenomenon

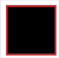
- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = \color{blue}{10\%}$



- In 2-D: $1 - (0.9)^2 = 0.19 = \color{blue}{19\%}$
- In 3-D: $1 - (0.9)^3 = 0.271 = \color{blue}{27.1\%}$
- In 50-D:

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$



- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} =$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 =$

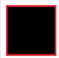
3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 = 99.5\%$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$

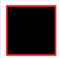


- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 = 99.5\%$

B_ϵ^m be the volume of the *edge set* in $[0, 1]^m$ for a fixed ϵ.

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$



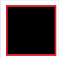
- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 = 99.5\%$

B_ϵ^m be the volume of the *edge set* in $[0, 1]^m$ for a fixed ϵ.

$$\lim_{m \to \infty} B_\epsilon^m =$$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$



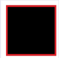
- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 = 99.5\%$

B_ϵ^m be the volume of the *edge set* in $[0,1]^m$ for a fixed ϵ.

$$\lim_{m\to\infty} B_\epsilon^m = \lim_{m\to\infty} [1 - (1-2\epsilon)^m] =$$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$



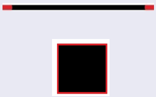
- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 = 99.5\%$

B_ϵ^m be the volume of the *edge set* in $[0, 1]^m$ for a fixed ϵ.

$$\lim_{m \to \infty} B_\epsilon^m = \lim_{m \to \infty} [1 - (1 - 2\epsilon)^m] = 1 .$$

3. The Empty Space Phenomenon

- How many data points are close to the edges?
- In 1-D: $1 - 0.9 = 0.1 = 10\%$
- In 2-D: $1 - (0.9)^2 = 0.19 = 19\%$
- In 3-D: $1 - (0.9)^3 = 0.271 = 27.1\%$
- In 50-D: $1 - (0.9)^{50} = 0.995 = 99.5\%$



B_ϵ^m be the volume of the *edge set* in $[0,1]^m$ for a fixed ϵ.

$$\lim_{m\to\infty} B_\epsilon^m = \lim_{m\to\infty} \left[1 - (1-2\epsilon)^m\right] = 1 \; .$$

The Empty Space Phenomenon !!

Curse of Dimensionality

4.

Curse of Dimensionality

4. Relations among the Dimensions

Curse of Dimensionality - Aspect 4

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

- Let the data set $\mathcal{X} \subset \mathbb{R}^6$.

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

- Let the data set $\mathcal{X} \subset \mathbb{R}^6$.
- Dimensionality of $\mathcal{X}$, $m = 6$.

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

- Let the data set $\mathcal{X} \subset \mathbb{R}^6$.
- Dimensionality of $\mathcal{X}$, $m = 6$.
- However, if $\bar{x} = (x_1, x_2, 3x_1, x_1 + x_2, 2x_2, x_1^2) \in \mathcal{X}$, then ...

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

- Let the data set $\mathcal{X} \subset \mathbb{R}^6$.
- Dimensionality of $\mathcal{X}$, $m = 6$.
- However, if $\bar{x} = (x_1, x_2, 3x_1, x_1 + x_2, 2x_2, x_1^2) \in \mathcal{X}$, then ...
- ... $\bar{x}$ essentially depends only on x_1, x_2.

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

- Let the data set $\mathcal{X} \subset \mathbb{R}^6$.
- Dimensionality of $\mathcal{X}$, $m = 6$.
- However, if $\bar{x} = (x_1, x_2, 3x_1, x_1 + x_2, 2x_2, x_1^2) \in \mathcal{X}$, then ...
- ... $\bar{x}$ essentially depends only on x_1, x_2.
- **Embedding** Dimn m vs. **Intrinsic** Dimn ℓ.

Curse of Dimensionality - Aspect 4

Correlation among the Dimensions

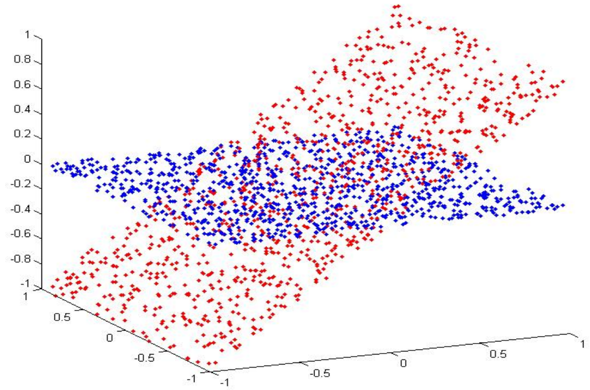
- Let the data set $\mathcal{X} \subset \mathbb{R}^6$.
- Dimensionality of $\mathcal{X}$, $m = 6$.
- However, if $\bar{x} = (x_1, x_2, 3x_1, x_1 + x_2, 2x_2, x_1^2) \in \mathcal{X}$, then ...
- ... $\bar{x}$ essentially depends only on x_1, x_2.
- **Embedding** Dimn m vs. **Intrinsic** Dimn ℓ.

Data may lie in a low-dimensional manifold, i.e., $\ell \ll m$.

4. Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

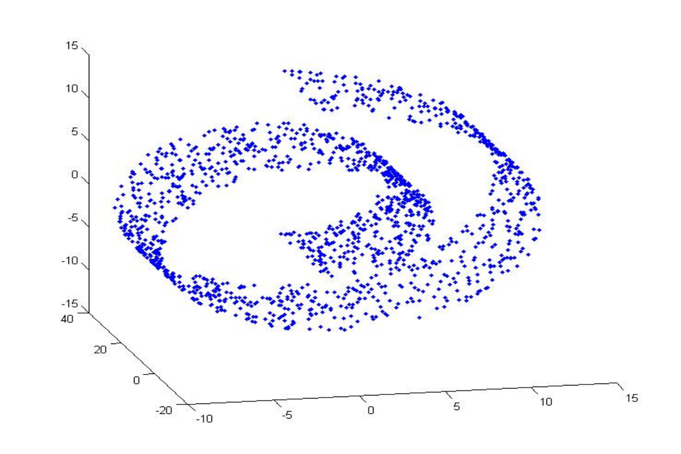
4. Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

Linear Manifolds - $\ell = 2 < m = 3$



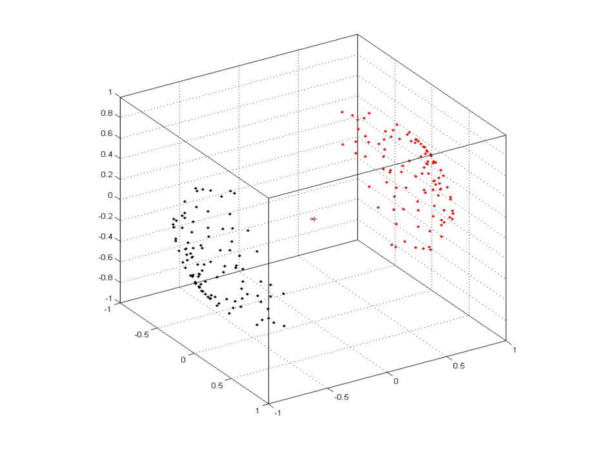
4. Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

Non-Linear Manifold - $\ell = 2 < m = 3$



4. Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

Non-Linear Manifold - $\ell = 2 < m = 3$

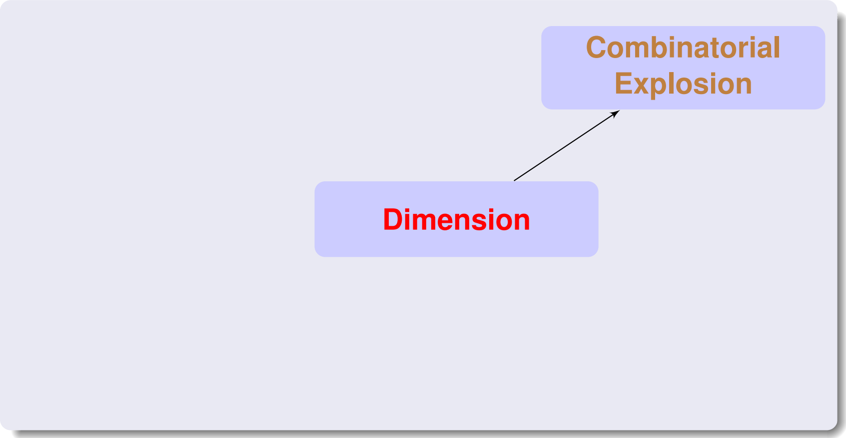


The story so far ...

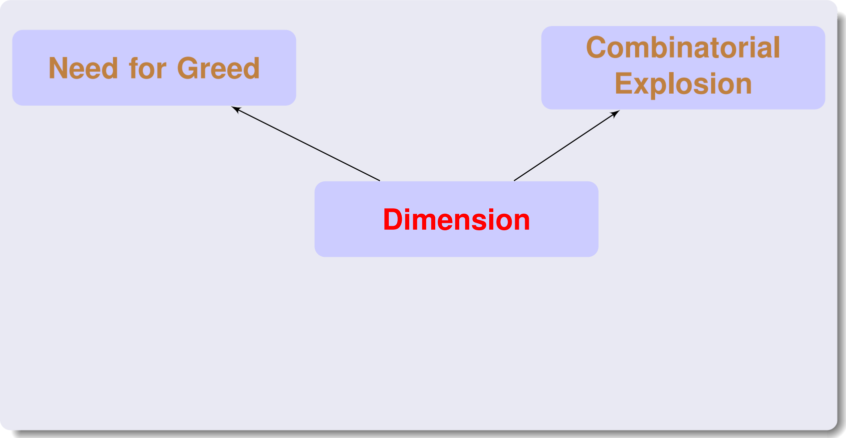
The story so far ...

Dimension

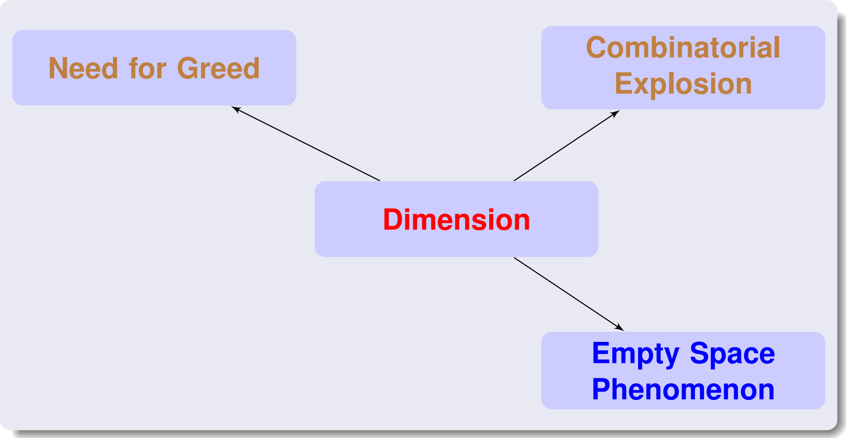
The story so far ...



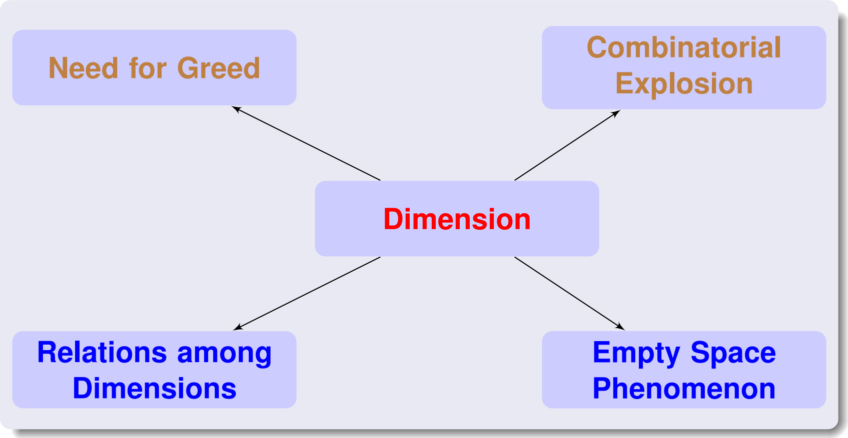
The story so far ...



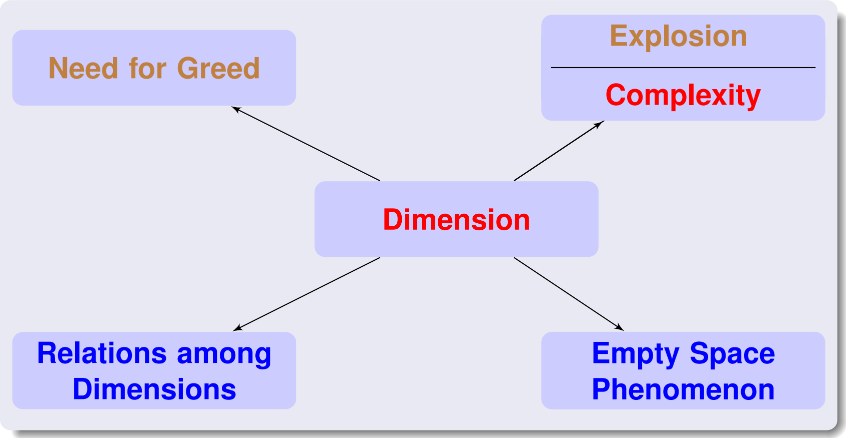
The story so far ...



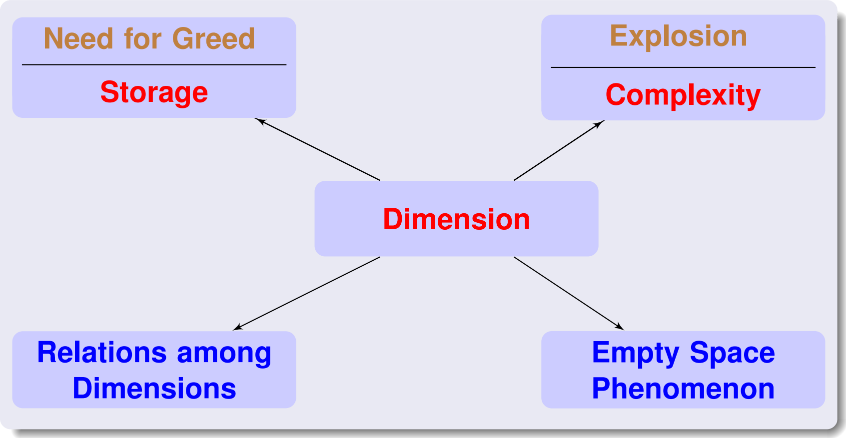
The story so far ...



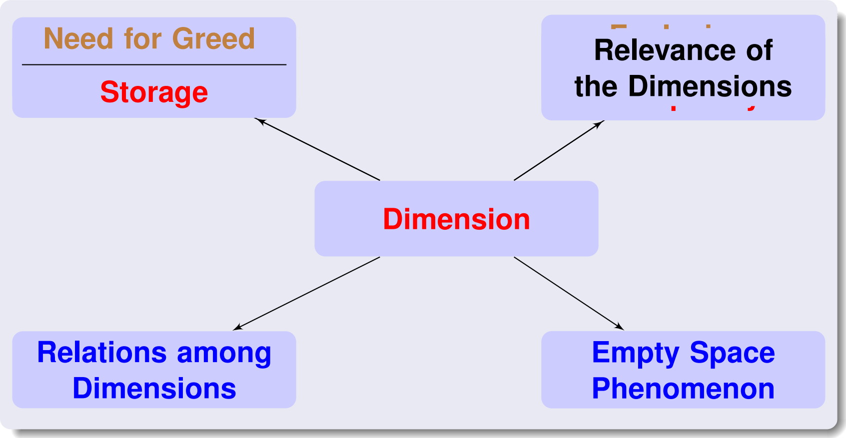
The story so far ...



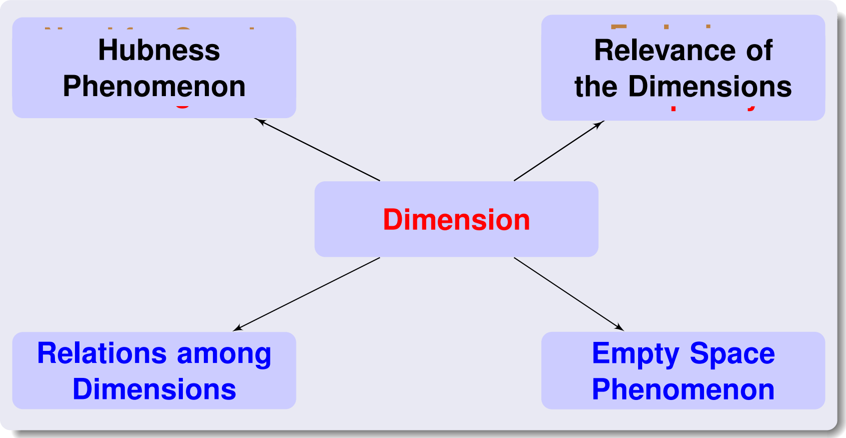
The story so far ...



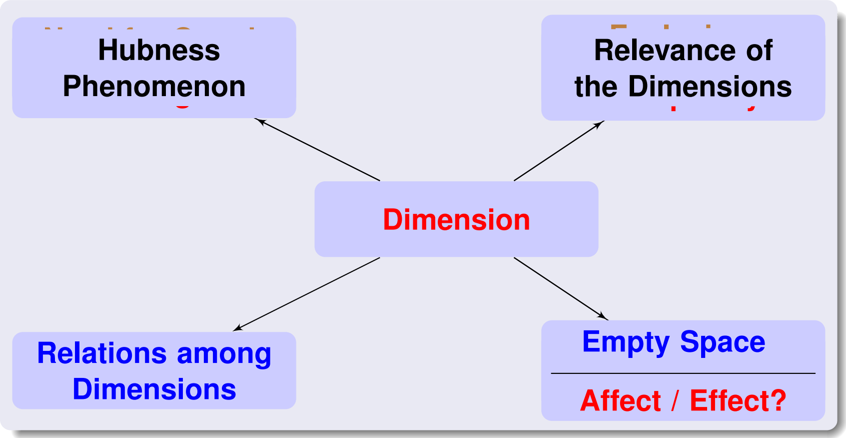
The story so far ...



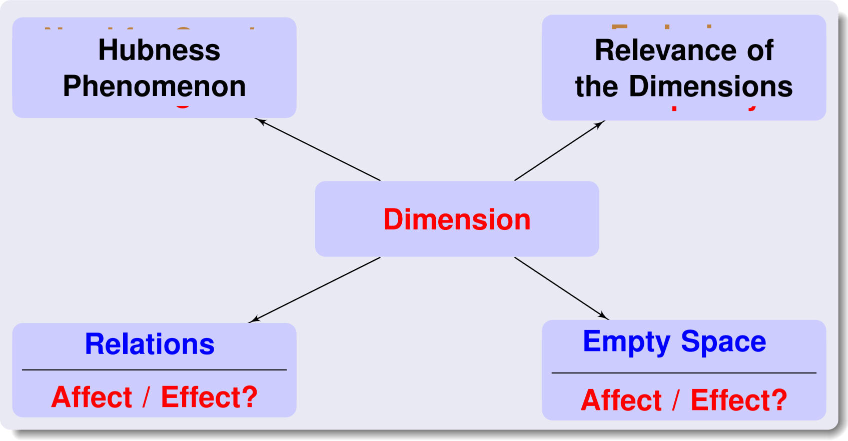
The story so far ...



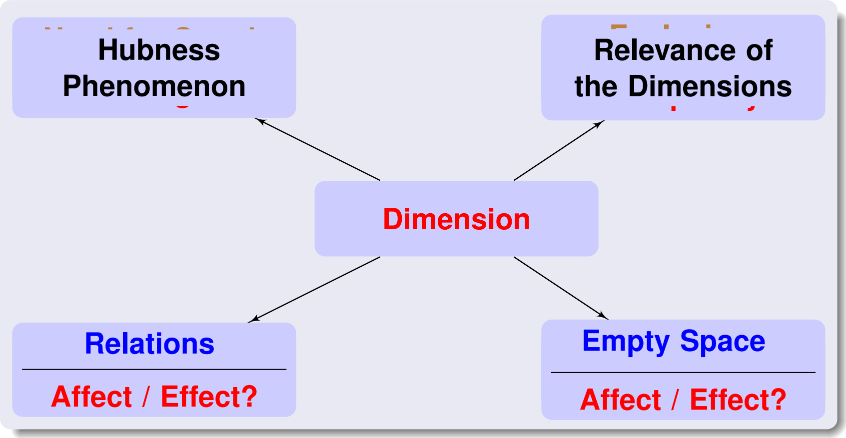
The story so far ...



The story so far ...

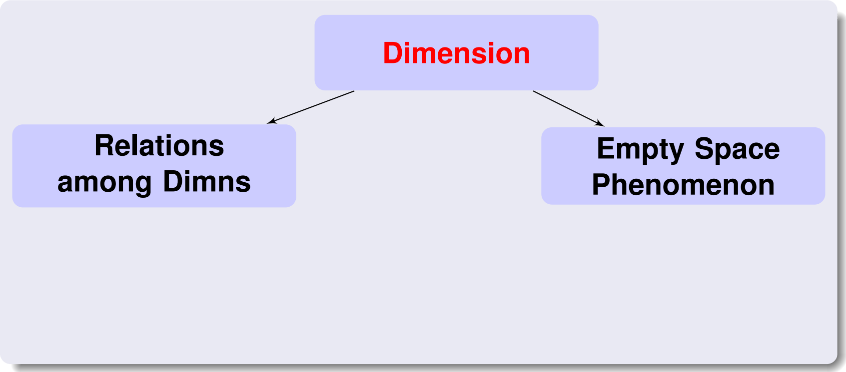


The story so far ...

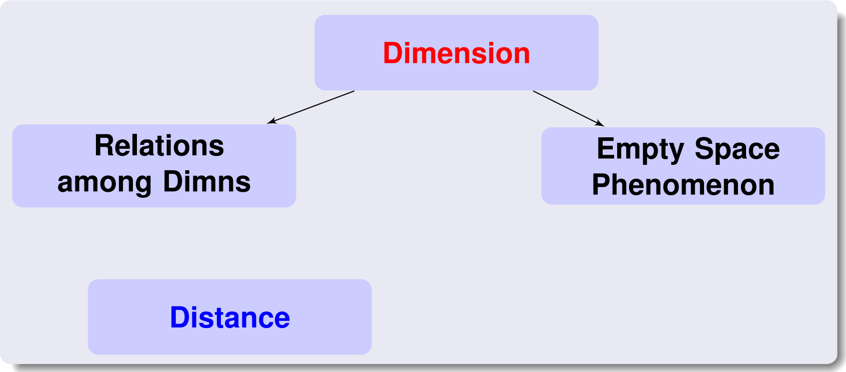


The story ahead ...

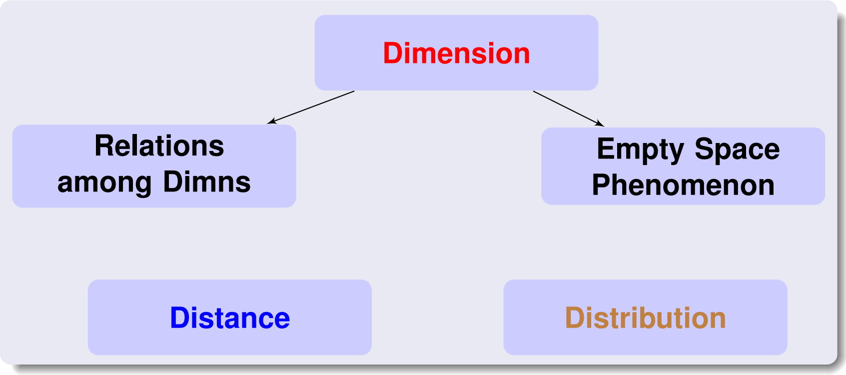
The story ahead ...



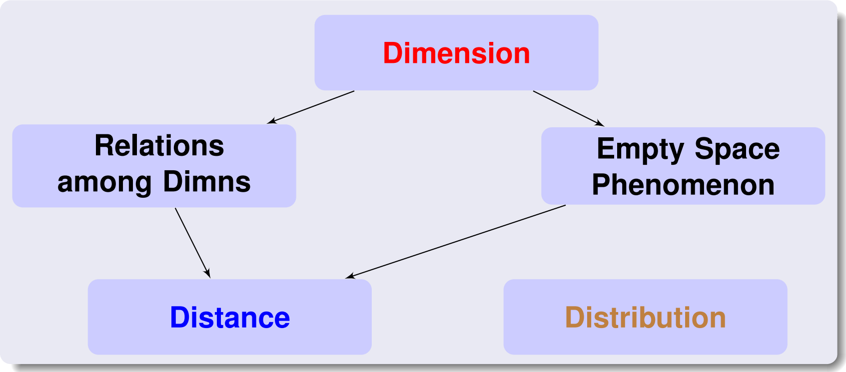
The story ahead ...



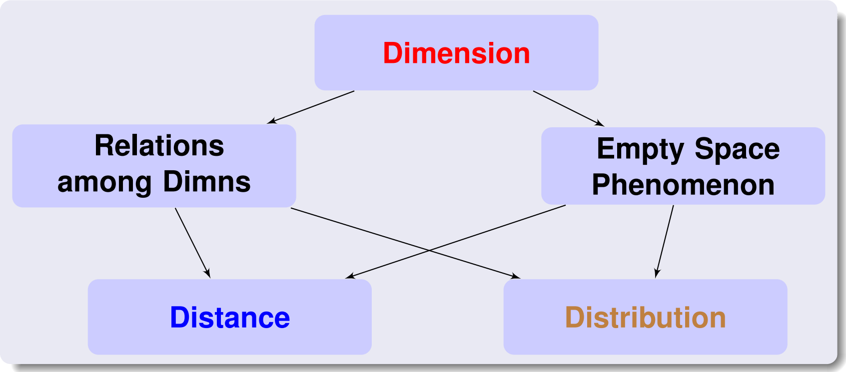
The story ahead ...



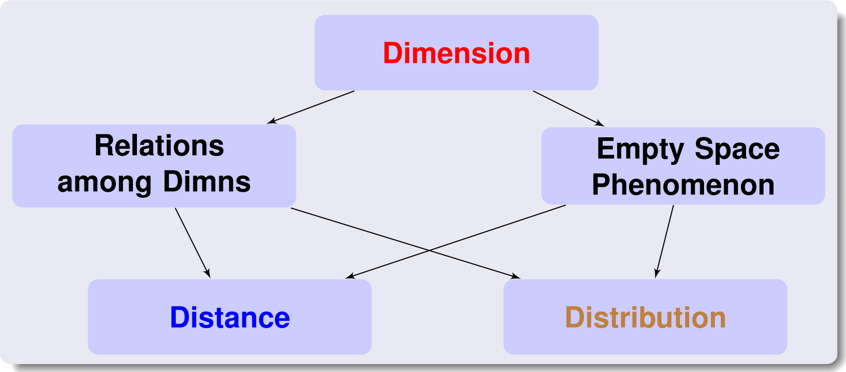
The story ahead ...



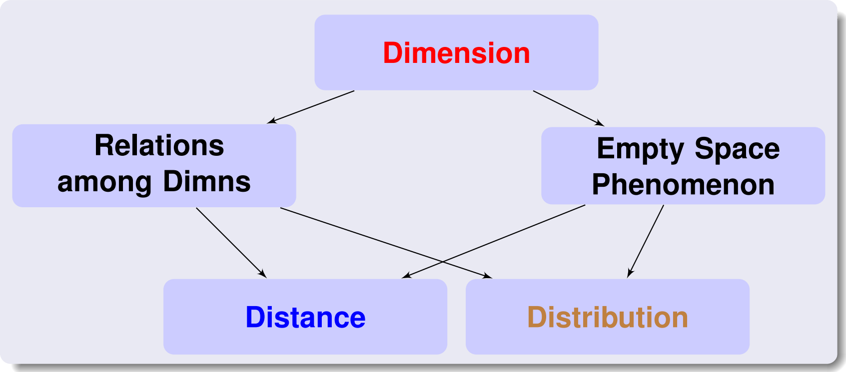
The story ahead ...



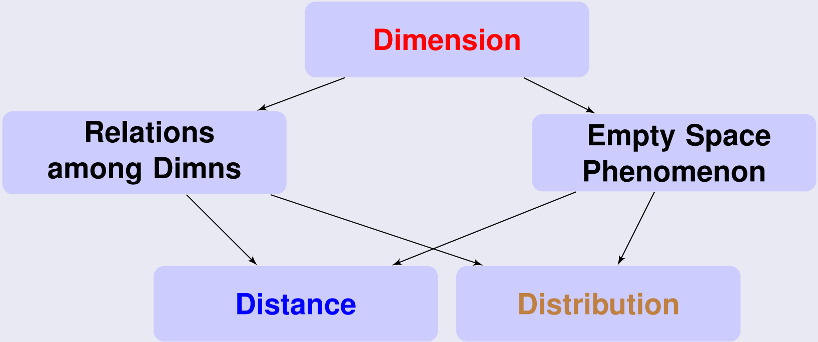
The story ahead ...



The story ahead ...

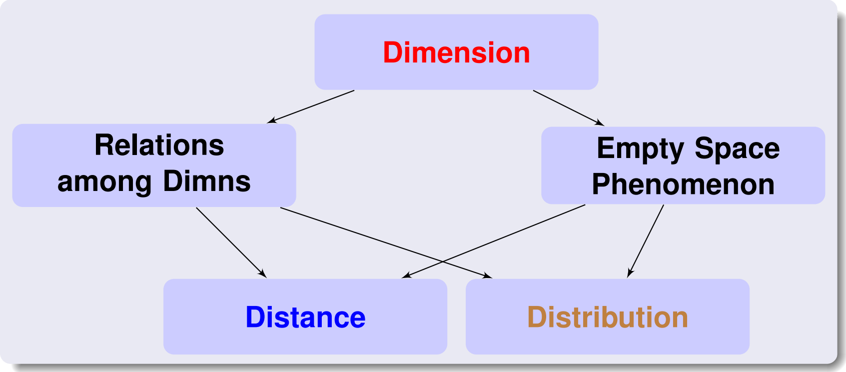


The story ahead ...



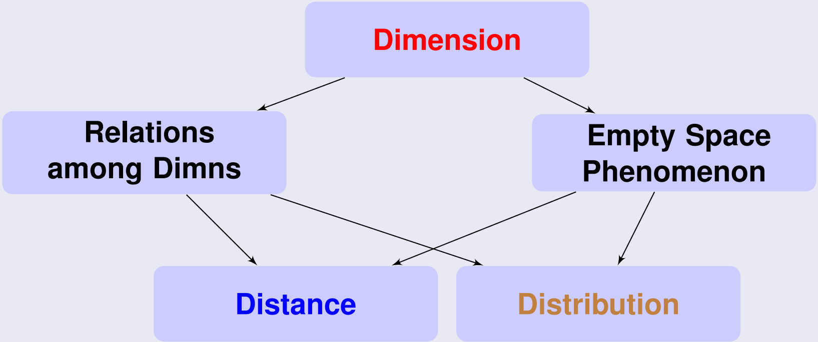
- How **D** & **D** are affected by the above issues.

The story ahead ...



- How **D** & **D** are affected by the above issues.
- In turn, can **D** & **D** mitigate the above effects?

The story ahead ...



- How **D** & **D** are affected by the above issues.
- In turn, can **D** & **D** mitigate the above effects?

3-D's in Data Analysis

3-D's in Data Analysis

2. Distance

Importance of Distance Measures

Importance of Distance Measures

In Similarity Searches

Importance of Distance Measures

In Similarity Searches

- Similarity and Distance are in some sense dual concepts.

Importance of Distance Measures

In Similarity Searches

- Similarity and Distance are in some sense dual concepts.
- k-**nearest** neighbours of a point, $NN_k(\bar{x})$.

Importance of Distance Measures

In Similarity Searches

- Similarity and Distance are in some sense dual concepts.
- k-**nearest** neighbours of a point, $NN_k(\bar{x})$.
- ε-**neighbourhood** of a point, $NN^{\varepsilon}(\bar{x})$.

Importance of Distance Measures

In Similarity Searches

- Similarity and Distance are in some sense dual concepts.
- k-**nearest** neighbours of a point, $NN_k(\bar{x})$.
- ε-**neighbourhood** of a point, $NN^{\varepsilon}(\bar{x})$.

In Clustering

Importance of Distance Measures

In Similarity Searches

- Similarity and Distance are in some sense dual concepts.
- k-**nearest** neighbours of a point, $NN_k(\bar{x})$.
- ε-**neighbourhood** of a point, $NN^{\varepsilon}(\bar{x})$.

In Clustering

- Most of the clustering algorithms are distance based.

Importance of Distance Measures

In Similarity Searches

- Similarity and Distance are in some sense dual concepts.
- k-**nearest** neighbours of a point, $NN_k(\bar{x})$.
- ε-**neighbourhood** of a point, $NN^{\varepsilon}(\bar{x})$.

In Clustering

- Most of the clustering algorithms are distance based.
- If not directly, *indirectly* !!

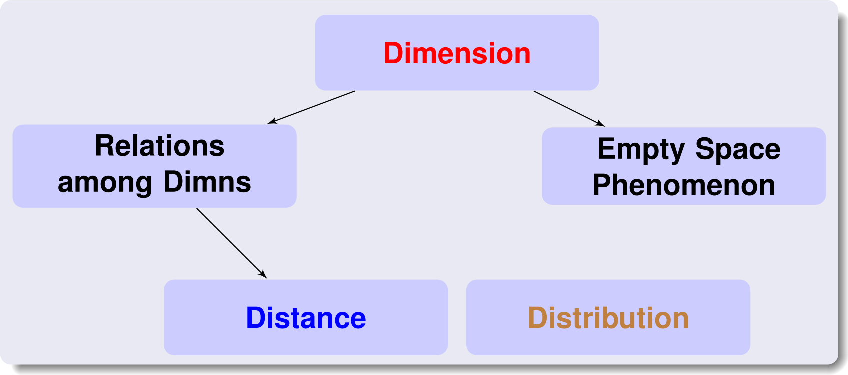
Importance of Distance Measures

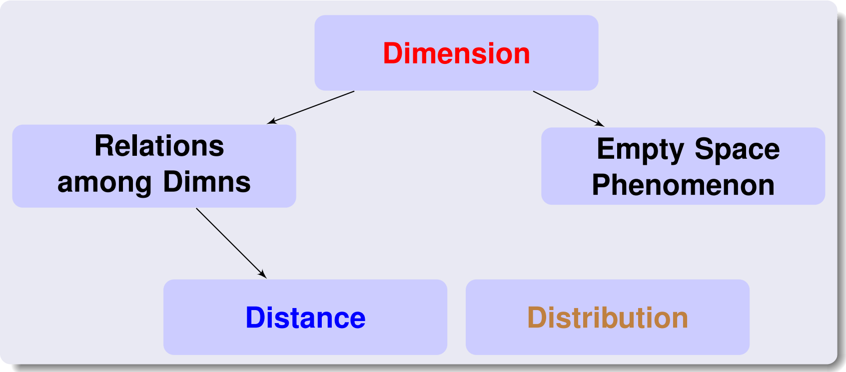
In Similarity Searches

- Similarity and Distance are in some sense dual concepts.
- k-**nearest** neighbours of a point, $NN_k(\bar{x})$.
- ε-**neighbourhood** of a point, $NN^{\varepsilon}(\bar{x})$.

In Clustering

- Most of the clustering algorithms are distance based.
- If not directly, *indirectly* !!


Dimension
Relations among Dimns
Empty Space Phenomenon
Distance
Distribution

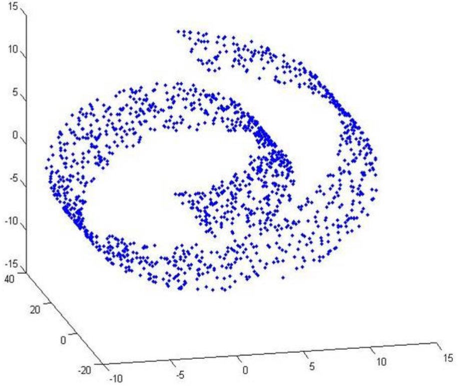


Appropriateness of Distance Functions

Appropriateness of Distance Functions

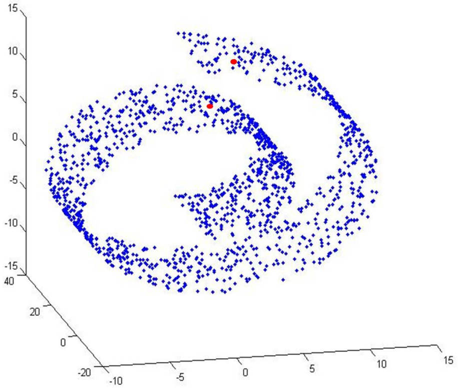
Appropriateness of Distance Functions

Consider the Swiss Roll data set.



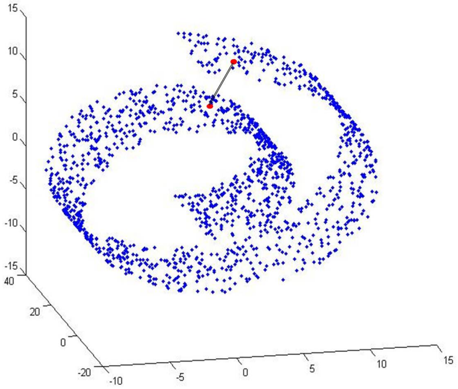
Appropriateness of Distance Functions

Are these two points close or far?



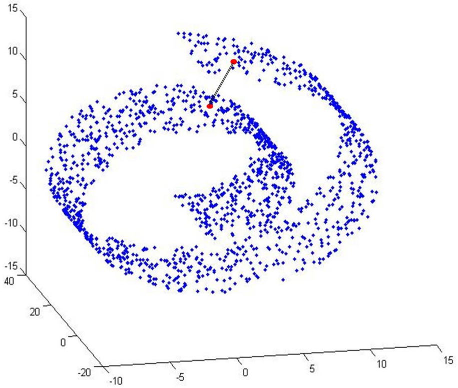
Appropriateness of Distance Functions

Are these two points close or far?



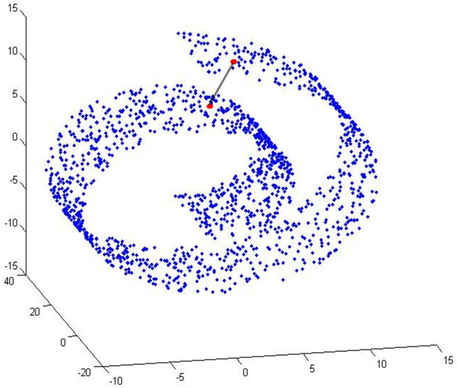
Appropriateness of Distance Functions

Intrinsic dimension $\ell = 2$, while Embedding dimension $m = 3$.



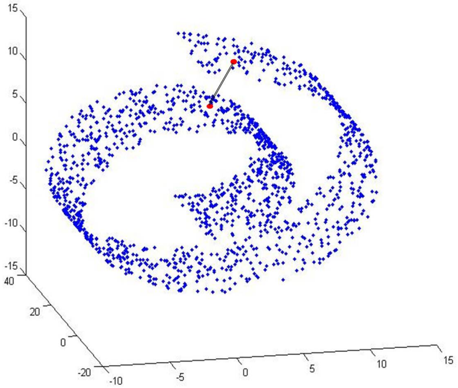
Appropriateness of Distance Functions

Distance in which dimension - Intrinsic or Embedding ?



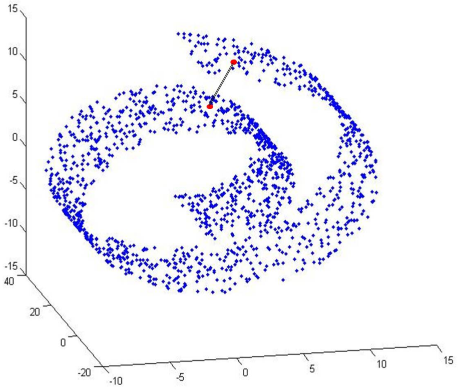
Relation between Dimensions - Revisited

Distance in which dimension - Intrinsic or Embedding ?



Relation between Dimensions - Revisited

Distance along which manifold ?



Appropriateness of Distance Functions

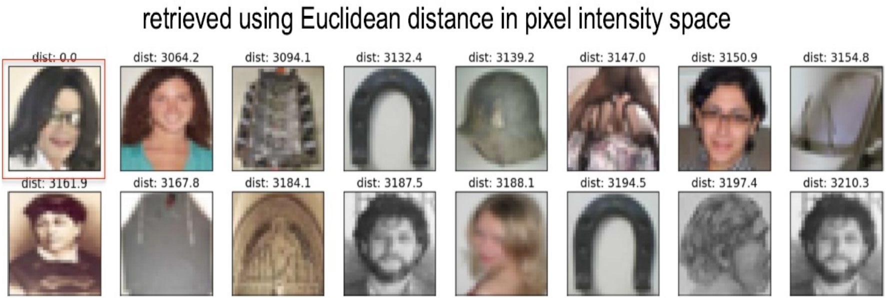
Appropriateness of Distance Functions

Euclidean distance in pixel intensity space

Appropriateness of Distance Functions

retrieved using Euclidean distance in pixel intensity space

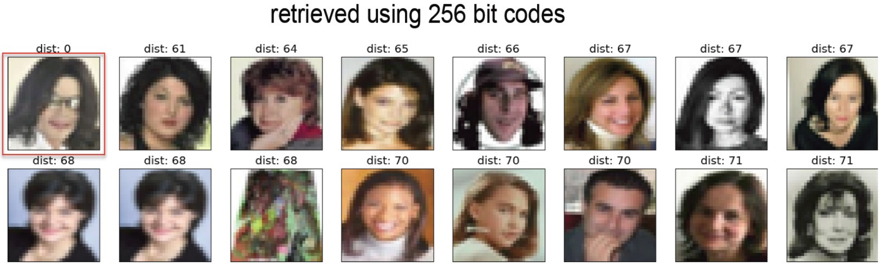
Appropriateness of Distance Functions



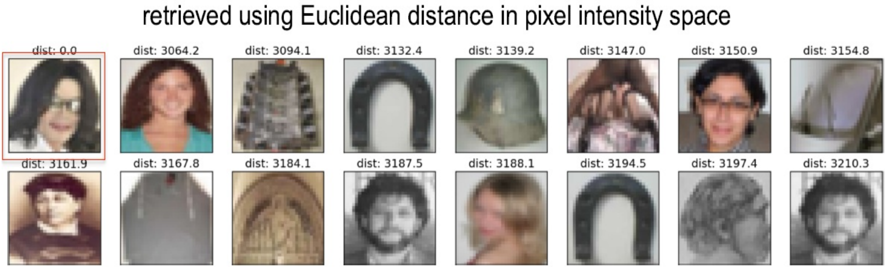
Appropriateness of Distance Functions

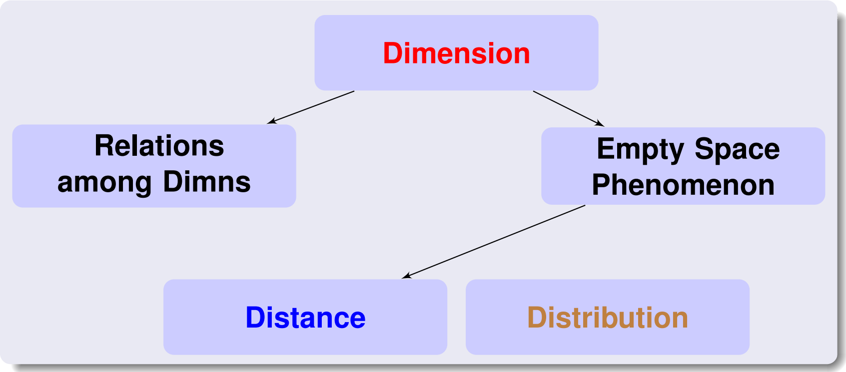
retrieved using 256 bit codes

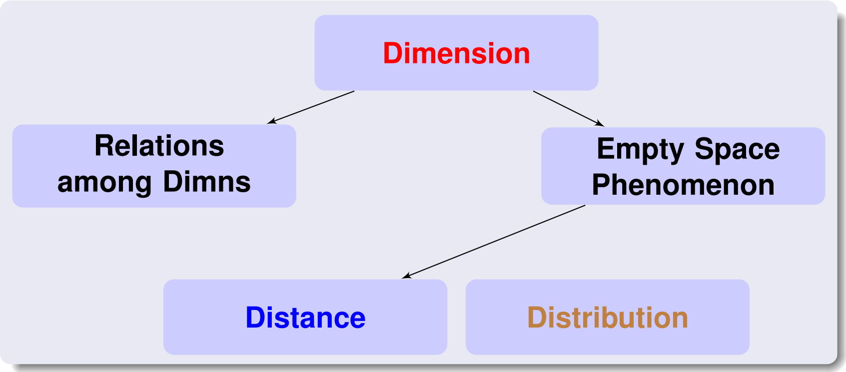
Appropriateness of Distance Functions



Appropriateness of Distance Functions




Dimension
Relations among Dimns
Empty Space Phenomenon
Distance
Distribution



Meaningfulness of Distances

Distribution of Euclidean Distances

Distribution of Euclidean Distances

- $\mathcal{X} = \{\bar{x}_i \in U\left([0,1]^m\right) \mid$

Distribution of Euclidean Distances

- $\mathcal{X} = \{\bar{x}_i \in U([0,1]^m) \mid 1 \leq i \leq 20,000\}$.

Distribution of Euclidean Distances

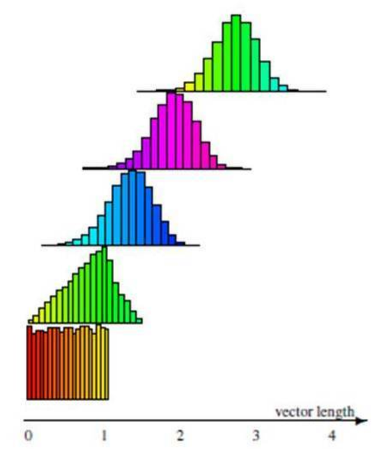
- $\mathcal{X} = \{\bar{x}_i \in U\left([0,1]^{m}\right) \mid 1 \leq i \leq 20,000\}$.
- Distribution of $\|\bar{x}_i\|$

Distribution of Euclidean Distances

- $\mathcal{X} = \{\bar{x}_i \in U\left([0,1]^m\right) \mid 1 \leq i \leq 20,000\}$.
- Distribution of $\|\bar{x}_i\|$ for $m = 1, 2, 5, 10, 20$.

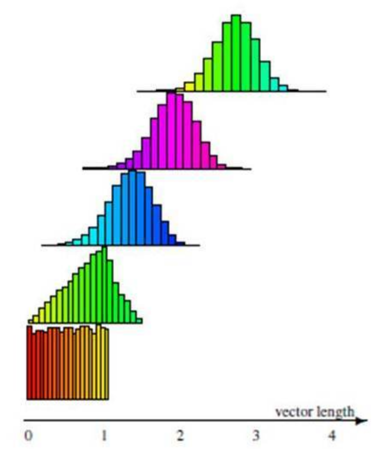
Distribution of Euclidean Distances

- $\mathcal{X} = \{\bar{x}_i \in U([0,1]^m) \mid 1 \le i \le 20,000\}$.
- Distribution of $\|\bar{x}_i\|$ for $m = 1, 2, 5, 10, 20$.



Distribution of Euclidean Distances

- $\|\bar{x}_i\|$ is small iff all m-components are small.
- When $m = 10$, $P\{\bar{x} \mid \|\bar{x}\| \leq 1\}$ is extremely small.



Distribution of lengths w.r.to any L_p Norm

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} \quad .$$

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} \quad .$$

- If $\bar{x} \in [0,1]^m$,

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} \quad .$$

- If $\bar{x} \in [0,1]^m$, $0 \le \|\bar{x}\|_2^2 \le m$.

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} \quad .$$

- If $\bar{x} \in [0,1]^m$, $0 \leq \|\bar{x}\|_2^2 \leq m$.

- Let us consider the normalised lengths, i.e., $\frac{1}{m}\|\bar{x}\|_2^2$.

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} .$$

- If $\bar{x} \in [0,1]^m$, $0 \leq \|\bar{x}\|_2^2 \leq m$.
- Let us consider the normalised lengths, i.e., $\frac{1}{m}\|\bar{x}\|_2^2$.
- Let us consider a general L_p norm, i.e.,

$$\|\bar{x}\|_p = \left(|x_1|^p + |x_2|^p + \ldots + |x_m|^p\right)^{\frac{1}{p}} .$$

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} \ .$$

- If $\bar{x} \in [0, 1]^m$, $0 \leq \|\bar{x}\|_2^2 \leq m$.
- Let us consider the normalised lengths, i.e., $\frac{1}{m}\|\bar{x}\|_2^2$.
- Let us consider a general L_p norm, i.e.,

$$\|\bar{x}\|_p = \left(|x_1|^p + |x_2|^p + \ldots + |x_m|^p\right)^{\frac{1}{p}} \ .$$

- Let us consider a modified general L_p norm, i.e.,

$$\|\bar{x}\|_p^p = |x_1|^p + |x_2|^p + \ldots + |x_m|^p \ .$$

Distribution of lengths w.r.to any L_p Norm

- Let $\bar{x} = (x_1, x_2, \ldots, x_m)$.

$$\|\bar{x}\|_2 = \left(x_1^2 + x_2^2 + \ldots + x_m^2\right)^{\frac{1}{2}} \; .$$

- If $\bar{x} \in [0,1]^m$, $0 \le \|\bar{x}\|_2^2 \le m$.
- Let us consider the normalised lengths, i.e., $\frac{1}{m}\|\bar{x}\|_2^2$.
- Let us consider a general L_p norm, i.e.,

$$\|\bar{x}\|_p = \left(|x_1|^p + |x_2|^p + \ldots + |x_m|^p\right)^{\frac{1}{p}} \; .$$

- Let us consider a modified general L_p norm, i.e.,

$$\|\bar{x}\|_p^p = |x_1|^p + |x_2|^p + \ldots + |x_m|^p \; .$$

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) =$$

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{p+1}\ .$$

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{p+1} \ .$$

- The corresponding variance is

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{p+1} \, .$$

- The corresponding variance is

$$Var\left(\frac{1}{m}\|\bar{x}\|_p^p\right) =$$

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{p+1}\ .$$

- The corresponding variance is

$$Var\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{m} \cdot \frac{p^2}{(2p+1)\,(p+1)^2}\ .$$

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{p+1} .$$

- The corresponding variance is

$$Var\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{m} \cdot \frac{p^2}{(2p+1)\,(p+1)^2} .$$

$$\lim_{m\to\infty} Var\left(\frac{1}{m}\|\bar{x}\|_p^p\right) =$$

Distribution of lengths of vectors in unit hypercube

- The expected value of the modified p-norm of a ...
- ... random vector $\bar{x}$ ...
- ... from the m-dimensional unit hypercube is

$$E\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{p+1} \; .$$

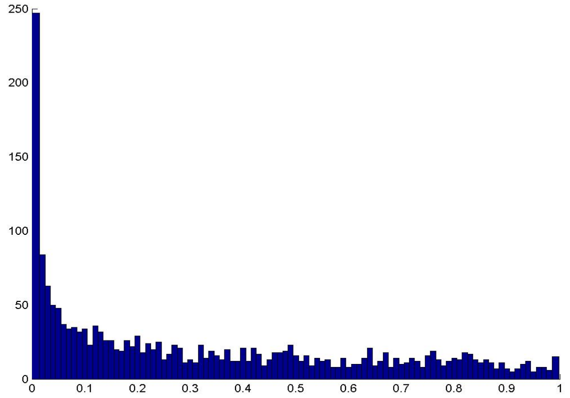
- The corresponding variance is

$$Var\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \frac{1}{m} \cdot \frac{p^2}{(2p+1)\,(p+1)^2} \; .$$

$$\lim_{m\to\infty} Var\left(\frac{1}{m}\|\bar{x}\|_p^p\right) = \mathbf{0} \; .$$

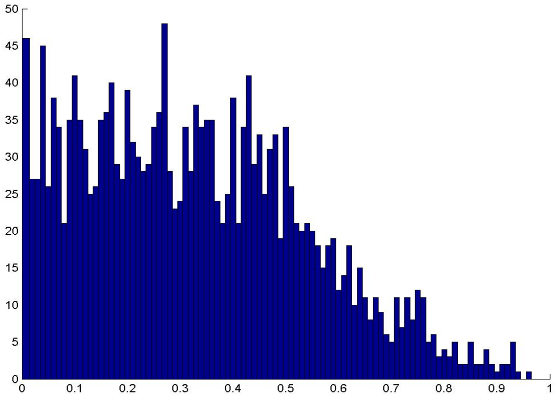
Distribution of distances to the Origin

$m = 1, p = 2$



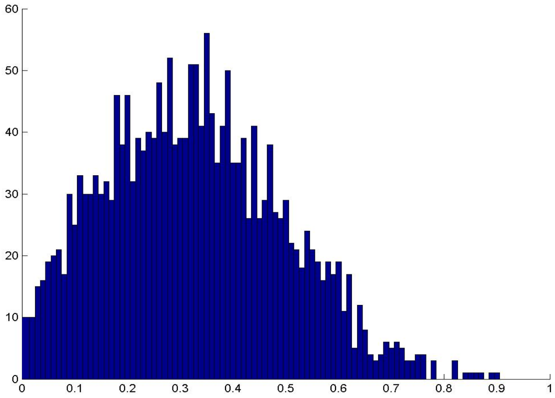
Distribution of distances to the Origin

$m = 2, p = 2$



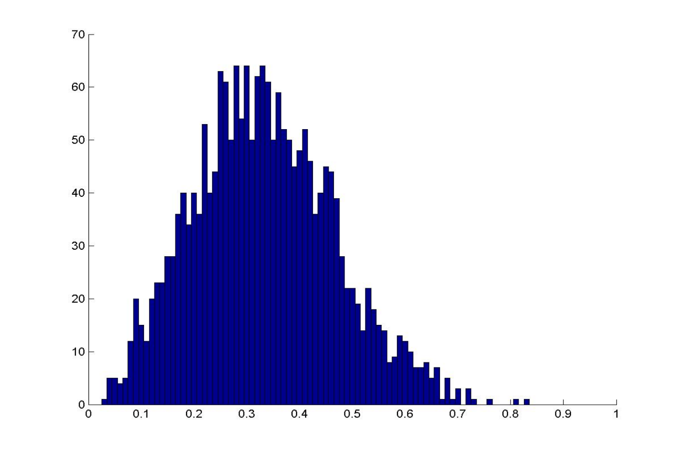
Distribution of distances to the Origin

$m = 3, p = 2$



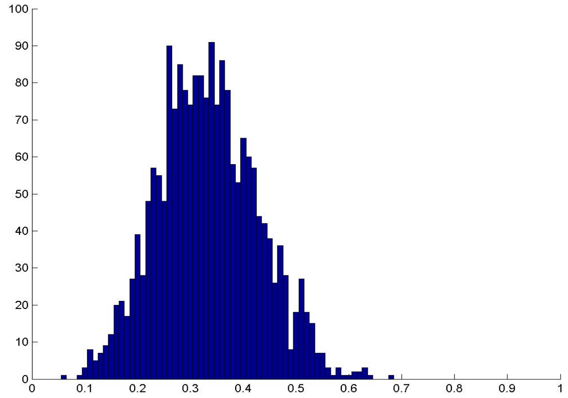
Distribution of distances to the Origin

$m = 5, p = 2$



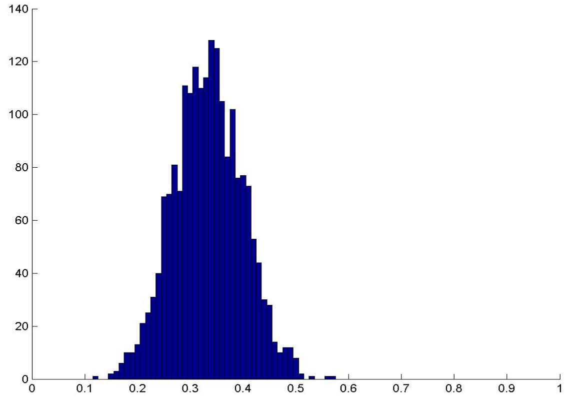
Distribution of distances to the Origin

$m = 10, p = 2$



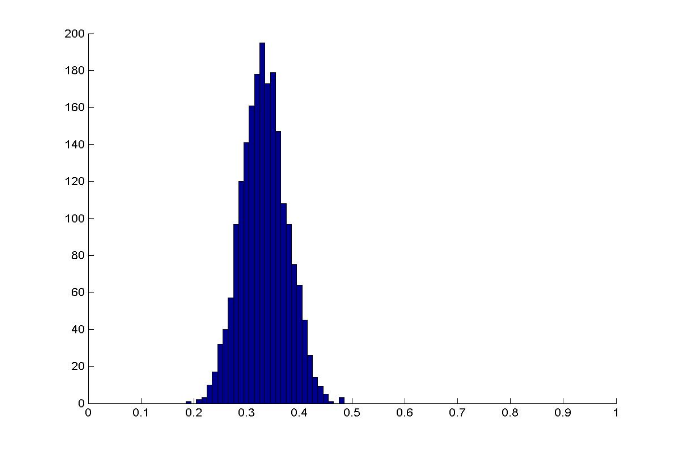
Distribution of distances to the Origin

$m = 20, p = 2$



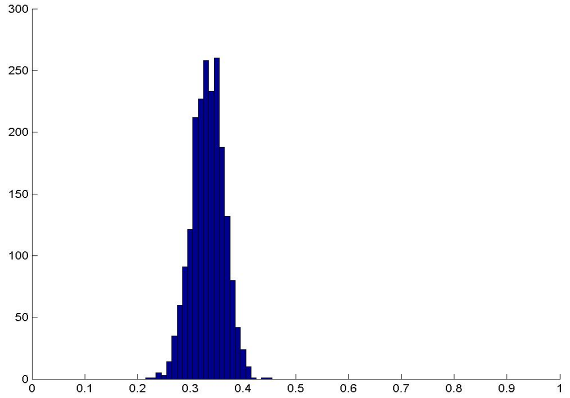
Distribution of distances to the Origin

$m = 50, p = 2$



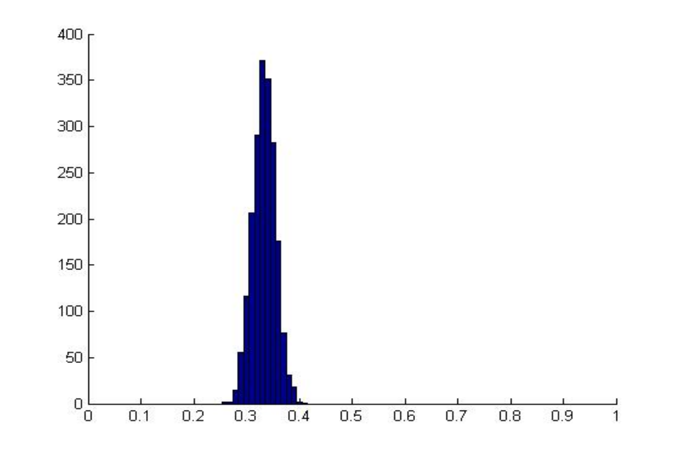
Distribution of distances to the Origin

$m = 100, p = 2$



Distribution of distances to the Origin

$m = 200, p = 2$



Distribution of lengths w.r.to any $\|\cdot\|$

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.

Distribution of lengths w.r.to any $\| \cdot \|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point.

Distribution of lengths w.r.to any $\| \cdot \|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\|\cdot\|$.

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\|\cdot\|$.
- $D_m^{(\min)}$ - smallest distance between any $\bar{x}_m$ to $\bar{q}_m$.

Distribution of lengths w.r.to any $\| \cdot \|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\| \cdot \|$.
- $D_m^{(\min)}$ - smallest distance between any $\bar{x}_m$ to $\bar{q}_m$.

Theorem [Beyer *et al*, 1999]

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\|\cdot\|$.
- $D_m^{(\min)}$ - smallest distance between any $\bar{x}_m$ to $\bar{q}_m$.

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right)$$

Distribution of lengths w.r.to any $\| \cdot \|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\| \cdot \|$.
- $D_m^{(\min)}$ - smallest distance between any $\bar{x}_m$ to $\bar{q}_m$.

Theorem [Beyer *et al*, 1999]

$$\lim_{m \to \infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0$$

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\|\cdot\|$.
- $D_m^{(\min)}$ - smallest distance between any $\bar{x}_m$ to $\bar{q}_m$.

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \Longrightarrow$$

Distribution of lengths w.r.to any $\|\cdot\|$

- $X_1, X_2, \ldots, X_m$ be i.i.d. random variables.
- $\bar{x}_m = (x_1, x_2, \ldots, x_m)$ be a r.$\bar{v}$, where $x_i \sim X_i$.
- $\bar{q}_m$ be the query point. W.l.o.g. $\bar{q}_m = \bar{0}_m$.
- $D_m^{(\max)}$ - largest distance between any $\bar{x}_m$ to $\bar{q}_m$ w.r.to $\|\cdot\|$.
- $D_m^{(\min)}$ - smallest distance between any $\bar{x}_m$ to $\bar{q}_m$.

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Meaningfulness of distance functions in HD

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \Longrightarrow \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

- When the relative variance (w.r.to the mean distance) ...

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

- When the relative variance (w.r.to the mean distance) ...
- ... of the distances to the origin converges to 0, then ...

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \, .$$

- When the relative variance (w.r.to the mean distance) ...
- ... of the distances to the origin converges to 0, then ...
- ... the relative difference between the ...

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

- When the relative variance (w.r.to the mean distance) ...
- ... of the distances to the origin converges to 0, then ...
- ... the relative difference between the ...
- ... closest and the farthest points to the origin ...

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

- When the relative variance (w.r.to the mean distance) ...
- ... of the distances to the origin converges to 0, then ...
- ... the relative difference between the ...
- ... closest and the farthest points to the origin ...
- ... goes to 0 with increasing dimensions.

Meaningfulness of distance functions in HD

Theorem [Beyer *et al*, 1999]

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

- When the relative variance (w.r.to the mean distance) ...
- ... of the distances to the origin converges to 0, then ...
- ... the relative difference between the ...
- ... closest and the farthest points to the origin ...
- ... goes to 0 with increasing dimensions.

Concentration of Norms phenomenon!!

Concentration of Norms

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0\ .$$

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \, .$$

Inability of a distance function to **separate** points well in HD.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \,.$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m$

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0\ .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.
- ... under some mild assumptions on X_i, e.g., $\mu_i, \sigma_i^2 < \infty$.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \,.$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.
- ... under some mild assumptions on X_i, e.g., $\mu_i, \sigma_i^2 < \infty$.
- ... for most distance functions.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.
- ... under some mild assumptions on X_i, e.g., $\mu_i, \sigma_i^2 < \infty$.
- ... for most distance functions.
  - L_p norms, $p \in [1, \infty)$ - Beyer *et al.*, ICDT 1999 .

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.
- ... under some mild assumptions on X_i, e.g., $\mu_i, \sigma_i^2 < \infty$.
- ... for most distance functions.
  - L_p norms, $p \in [1, \infty)$ - Beyer *et al.*, ICDT 1999 .
  - L_p norms, $p \in (0, 1)$ - Francois *et al.* 2007, IEEE TKDE.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.
- ... under some mild assumptions on X_i, e.g., $\mu_i, \sigma_i^2 < \infty$.
- ... for most distance functions.
  - L_p norms, $p \in [1, \infty)$ - Beyer *et al.*, ICDT 1999 .
  - L_p norms, $p \in (0, 1)$ - Francois *et al.* 2007, IEEE TKDE.
  - Aggarwal 2001, Hsu & Chen 2009, Jayaram & Klawonn 2012.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

Inability of a distance function to **separate** points well in HD.

Result is valid ...

- ... for any $\bar{q}_m \neq \bar{0}_m$.
- ... under some mild assumptions on X_i, e.g., $\mu_i, \sigma_i^2 < \infty$.
- ... for most distance functions.
  - L_p norms, $p \in [1, \infty)$ - Beyer *et al.*, ICDT 1999 .
  - L_p norms, $p \in (0, 1)$ - Francois *et al.* 2007, IEEE TKDE.
  - Aggarwal 2001, Hsu & Chen 2009, Jayaram & Klawonn 2012.

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

Inability of a distance function to **separate** points well in HD.

What is its effect?

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

Inability of a distance function to **separate** points well in HD.

What is its effect?

- Distance between any two points in HD seems the same!!

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Inability of a distance function to **separate** points well in HD.

What is its effect?

- Distance between any two points in HD seems the same!!
- Affects Query searches !

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Inability of a distance function to **separate** points well in HD.

What is its effect?

- Distance between any two points in HD seems the same!!
- Affects Query searches !
- Who is your nearest neighbour?

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \Longrightarrow \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \; .$$

Inability of a distance function to **separate** points well in HD.

What is its effect?

- Distance between any two points in HD seems the same!!
- Affects Query searches !
- Who is your nearest neighbour?

Cause?

Concentration of Norms

$$\lim_{m\to\infty} Var\left(\frac{\|\bar{x}_m\|}{E\left(\|\bar{x}_m\|\right)}\right) = 0 \implies \frac{D_m^{(\max)} - D_m^{(\min)}}{D_m^{(\min)}} \to_p 0 \ .$$

Inability of a distance function to **separate** points well in HD.

What is its effect?

- Distance between any two points in HD seems the same!!
- Affects Query searches !
- Who is your nearest neighbour?

Cause?

The Empty Space phenomenon?

Interplay between the 3D's - 1

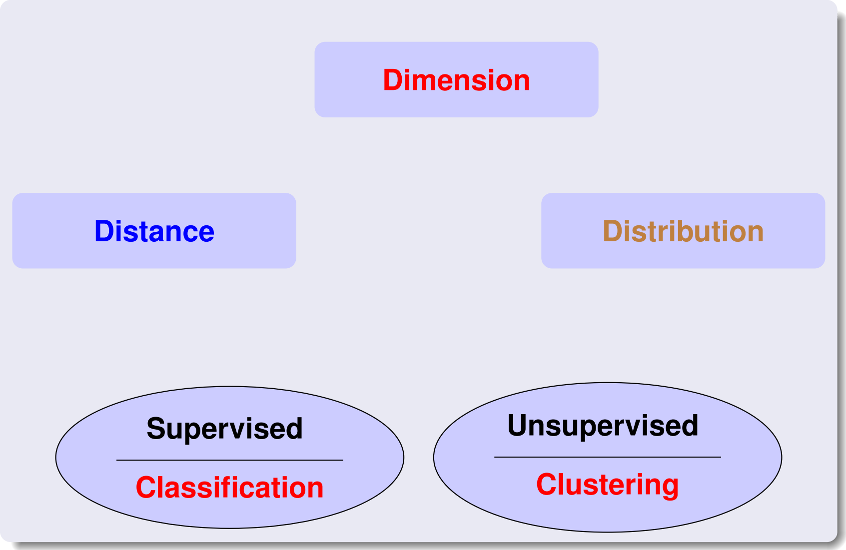
Interplay between the 3D's - 1

Dimension

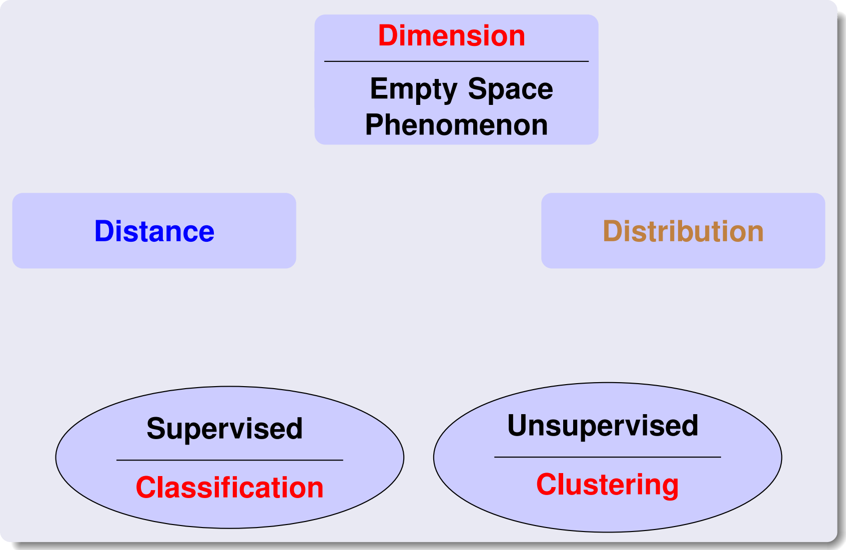
Distance

Distribution

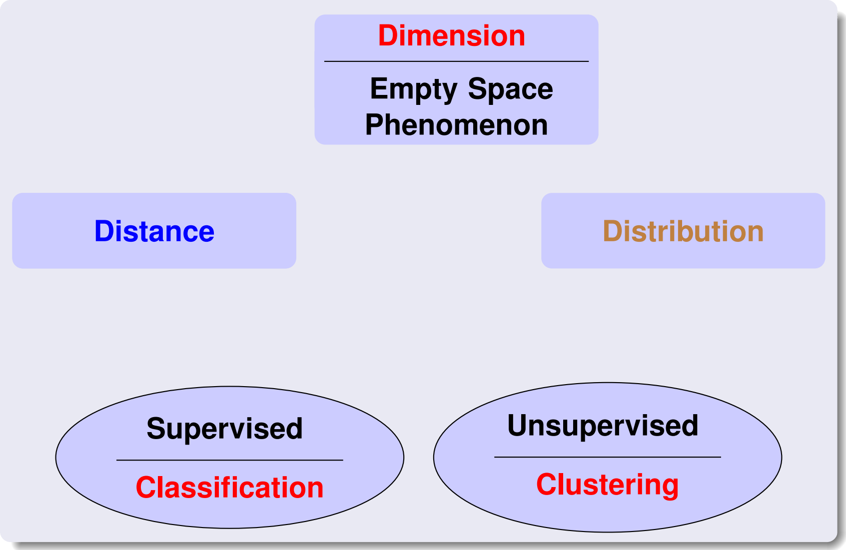
Interplay between the 3D's - 1



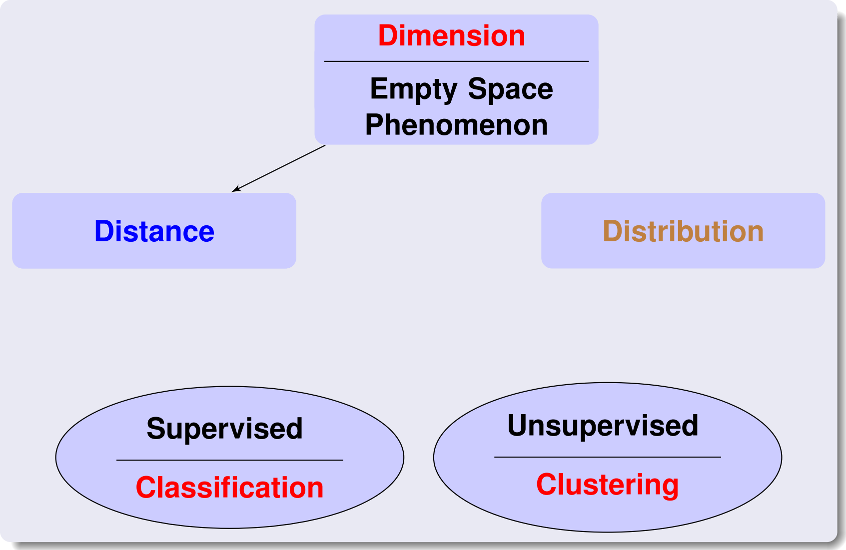
Interplay between the 3D's - 1



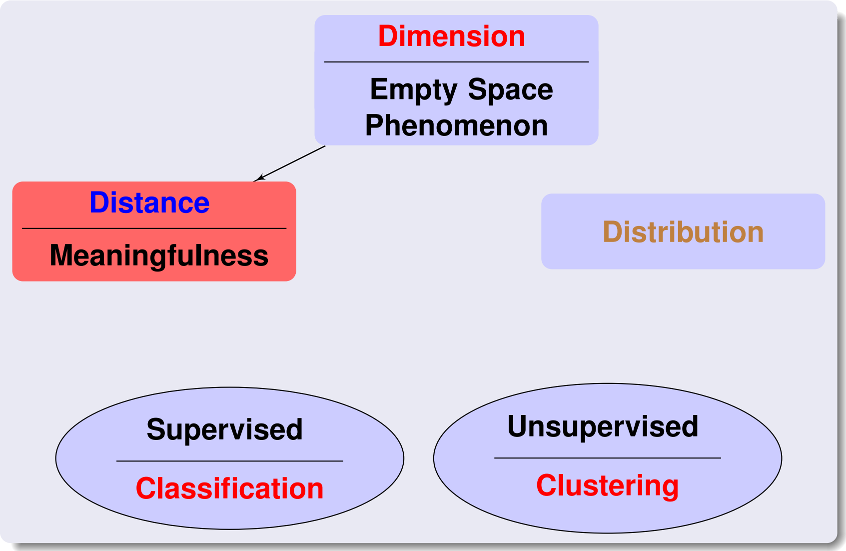
Interplay between the 3D's - 1



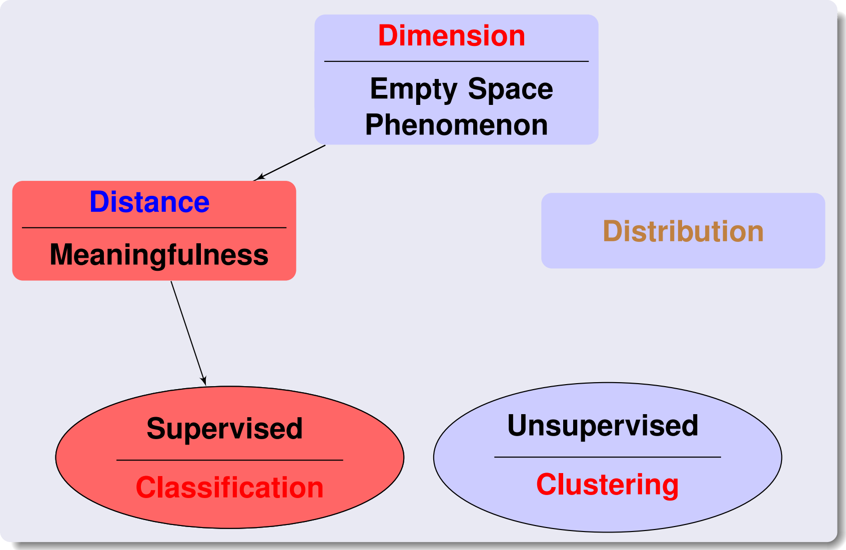
Interplay between the 3D's - 1



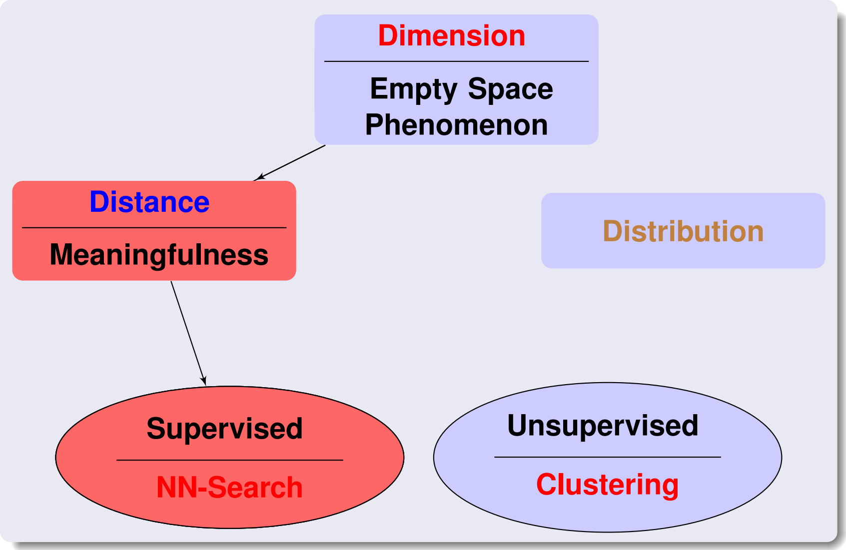
Interplay between the 3D's - 1



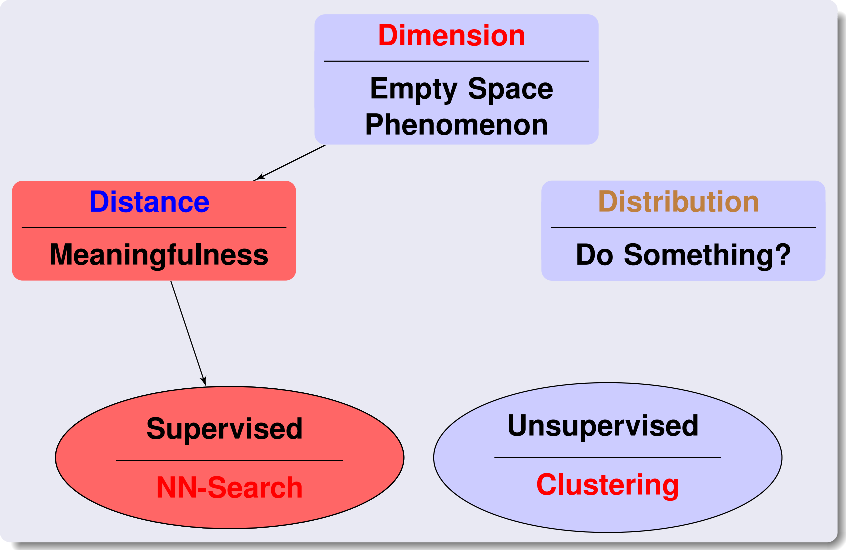
Interplay between the 3D's - 1



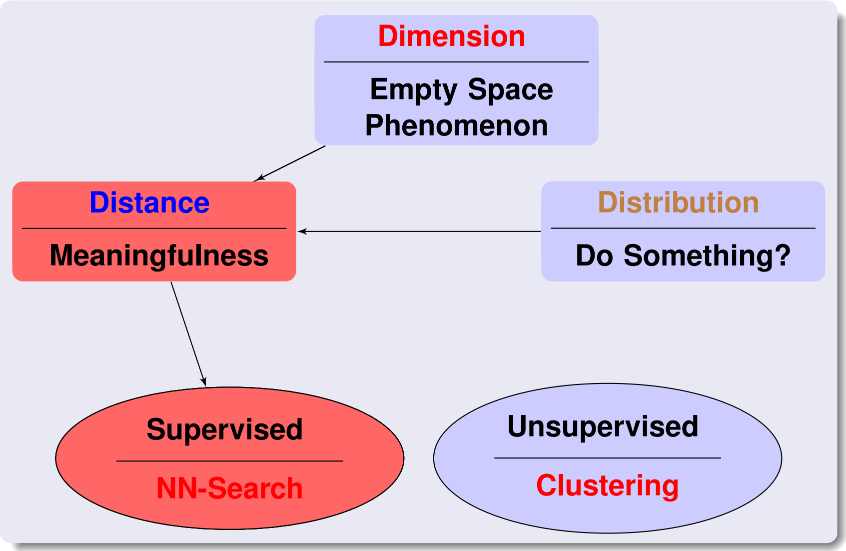
Interplay between the 3D's - 1



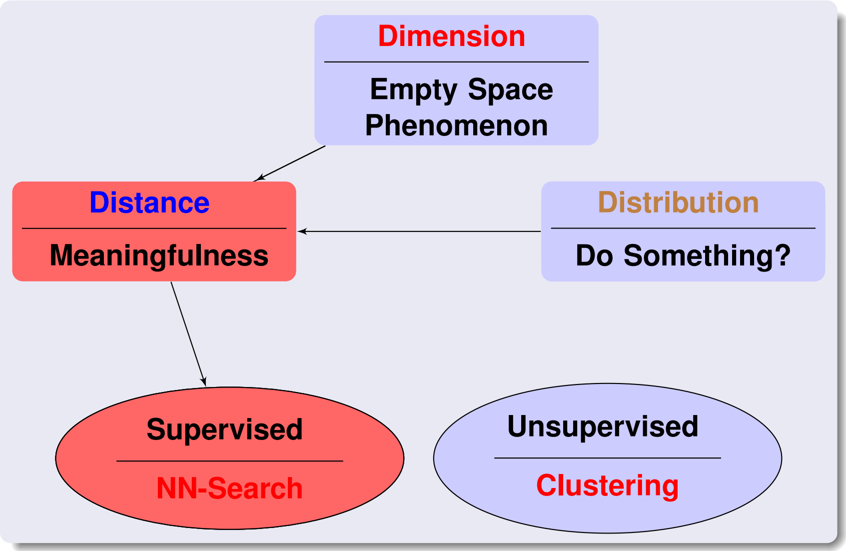
Interplay between the 3D's - 1



Interplay between the 3D's - 1



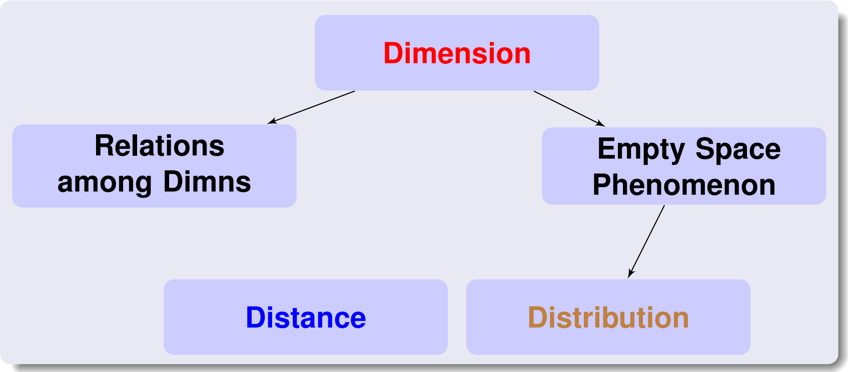
Interplay between the 3D's - 1

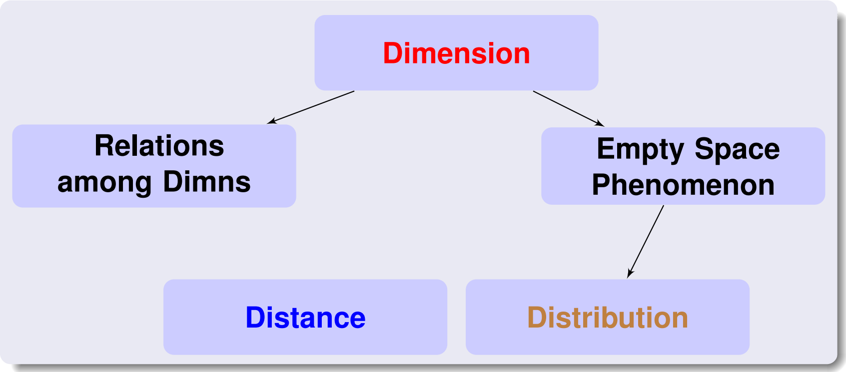


3-D’s in Data Analysis

3-D's in Data Analysis

3. Distribution


Dimension
Relations among Dimns
Empty Space Phenomenon
Distance
Distribution



Meaningfulness of Results ?

Data from a mixture of distributions

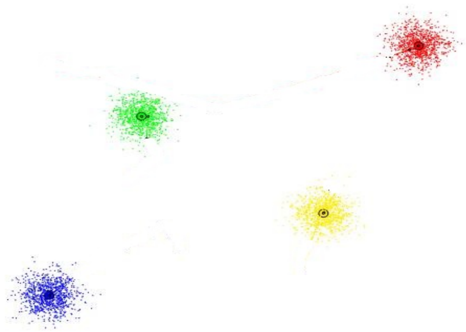
- So far, $\mathcal{X}$ contained r.$\bar{v}$ from a single distribution.

Data from a mixture of distributions

- So far, $\mathcal{X}$ contained r.$\bar{v}$ from a single distribution.
- Now, let, $\mathcal{X}$ contain r.$\bar{v}$ from a mixture of distributions.

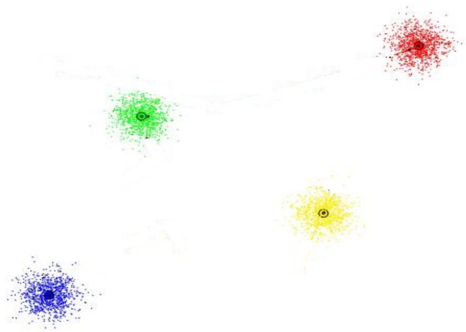
Data from a mixture of distributions

- So far, $\mathcal{X}$ contained r.$\bar{v}$ from a single distribution.
- Now, let, $\mathcal{X}$ contain r.$\bar{v}$ from a mixture of distributions.



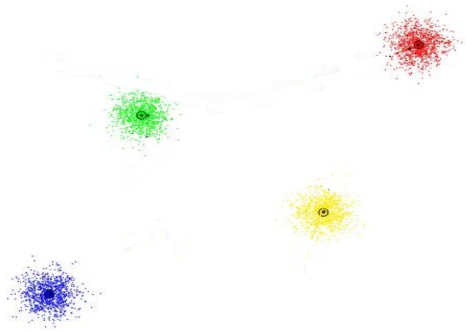
Data from a mixture of distributions

- Let there be c Gaussian clusters in m-dimensions.
-



Data from a mixture of distributions

- Let there be c Gaussian clusters in m-dimensions.
- **Is not nearest neighbour query meaningful now?**



NN-Queries can be meaningful!!

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...
- ... (as is usually the case with real data sets) ...

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...
- ... (as is usually the case with real data sets) ...
- ... nearest neighbour searches can be meaningful !

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...
- ... (as is usually the case with real data sets) ...
- ... nearest neighbour searches can be meaningful !
- Theoretical results do exist - Bennett *et al.* (2001).

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...
- ... (as is usually the case with real data sets) ...
- ... nearest neighbour searches can be meaningful !
- Theoretical results do exist - Bennett *et al.* (2001).
- The Empty Space or CoN phenomena do not mean much for NN now!!.

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...
- ... (as is usually the case with real data sets) ...
- ... nearest neighbour searches can be meaningful !
- Theoretical results do exist - Bennett *et al.* (2001).
- The Empty Space or CoN phenomena do not mean much for NN now!!.

Really?!

NN-Queries can be meaningful!!

- When data come from mixture of distributions ...
- ... (as is usually the case with real data sets) ...
- ... nearest neighbour searches can be meaningful !
- Theoretical results do exist - Bennett *et al.* (2001).
- The Empty Space or CoN phenomena do not mean much for NN now!!.

Really?!

Interplay between the 3D's - 1

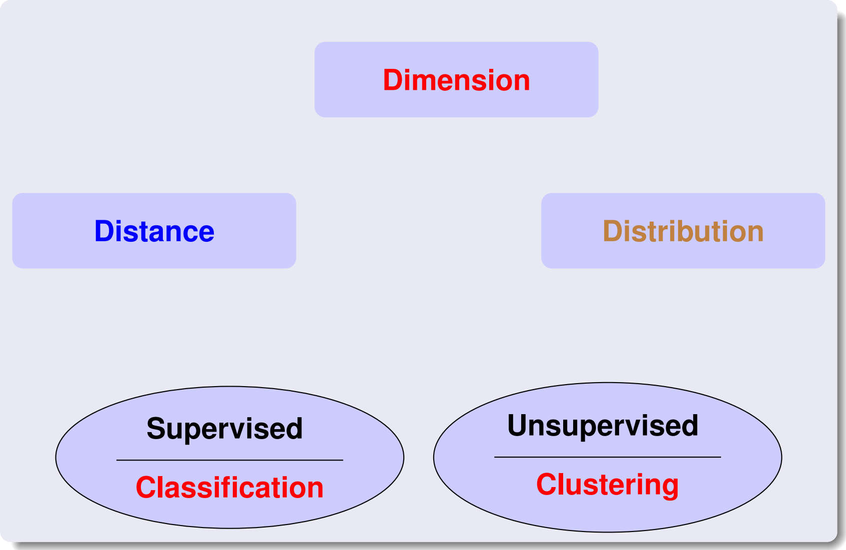
Interplay between the 3D's - 1

Dimension

Distance

Distribution

Interplay between the 3D's - 1



Interplay between the 3D's - 1

Dimension

Empty Space Phenomenon

Distance

Distribution

Supervised

Classification

Unsupervised

Clustering

Interplay between the 3D's - 1

Dimension

Empty Space Phenomenon

Distance

Distribution

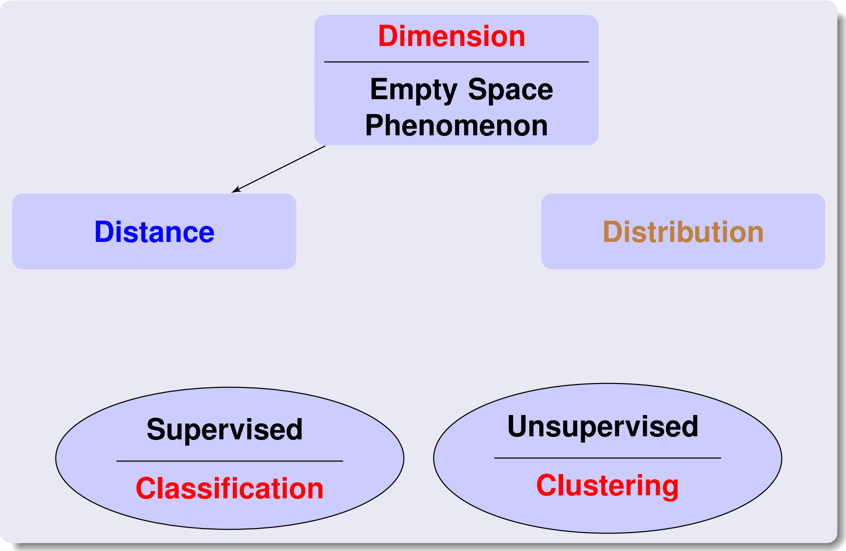
Supervised

Classification

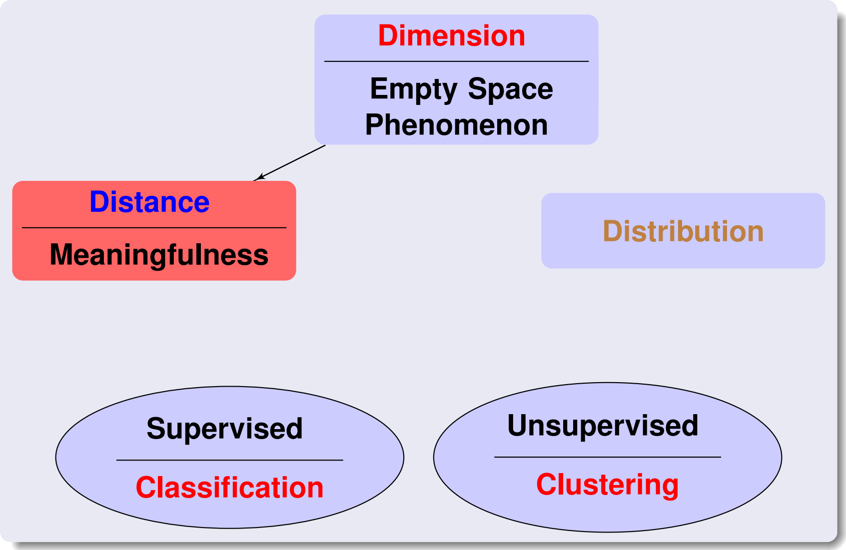
Unsupervised

Clustering

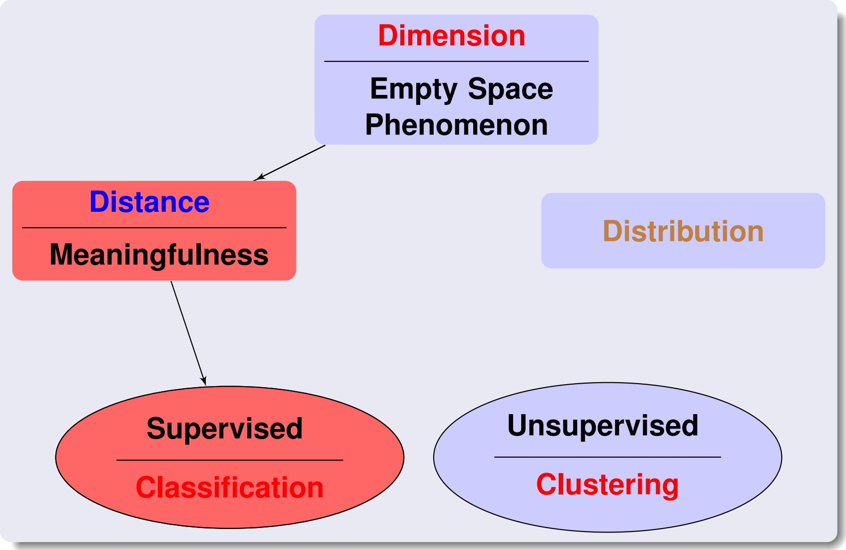
Interplay between the 3D’s - 1



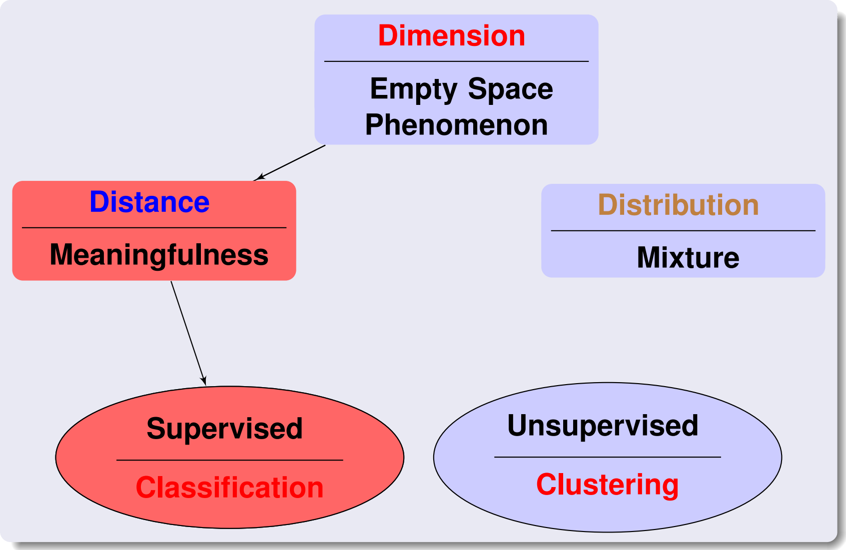
Interplay between the 3D's - 1



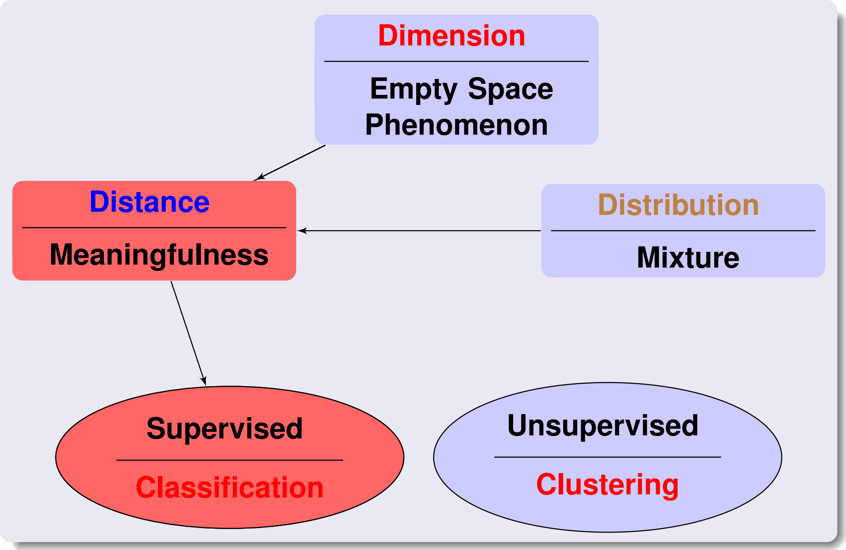
Interplay between the 3D's - 1



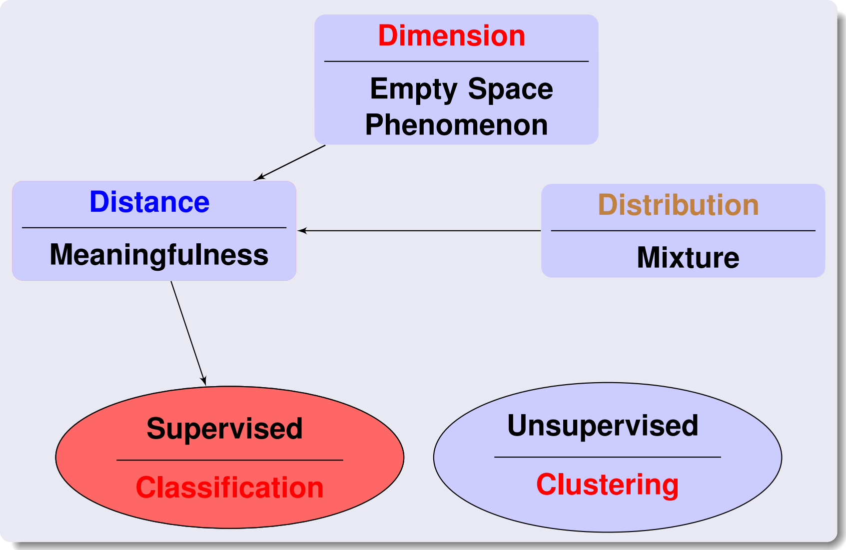
Interplay between the 3D's - 1



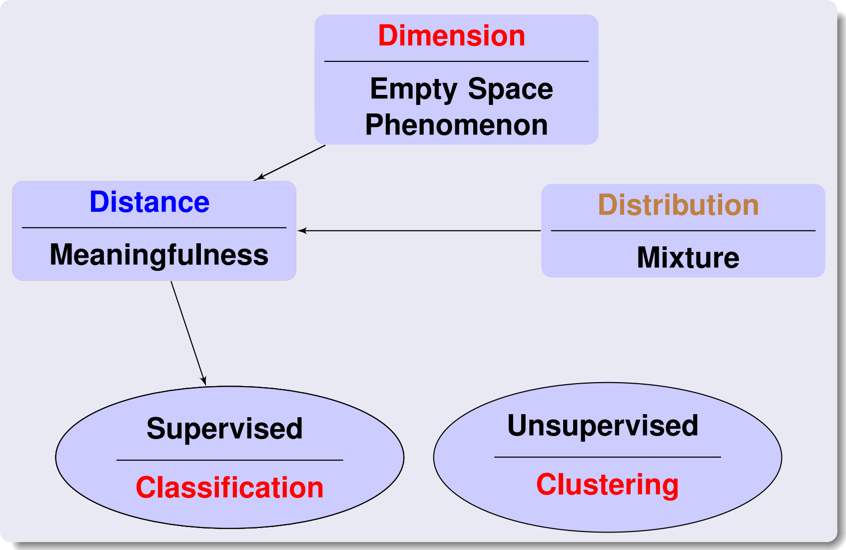
Interplay between the 3D's - 1

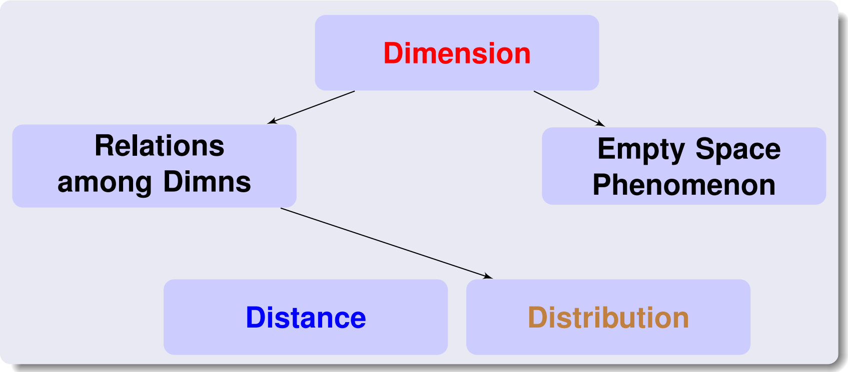


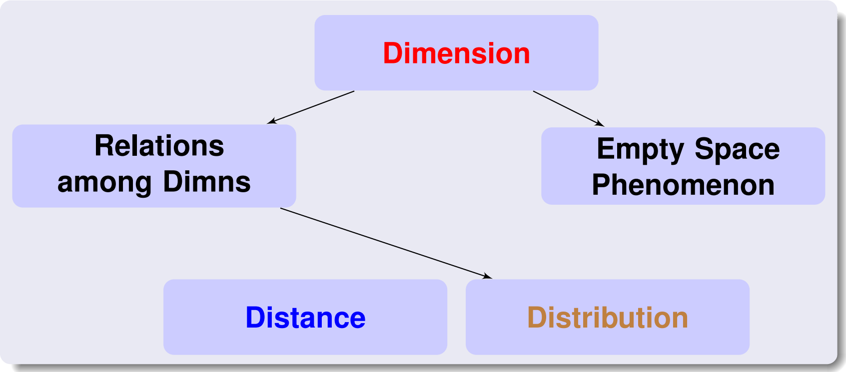
Interplay between the 3D's - 1



Interplay between the 3D's - 1




Dimension
Relations among Dimns
Empty Space Phenomenon
Distance
Distribution



Intrinsic dimn vs. Embedding dimn

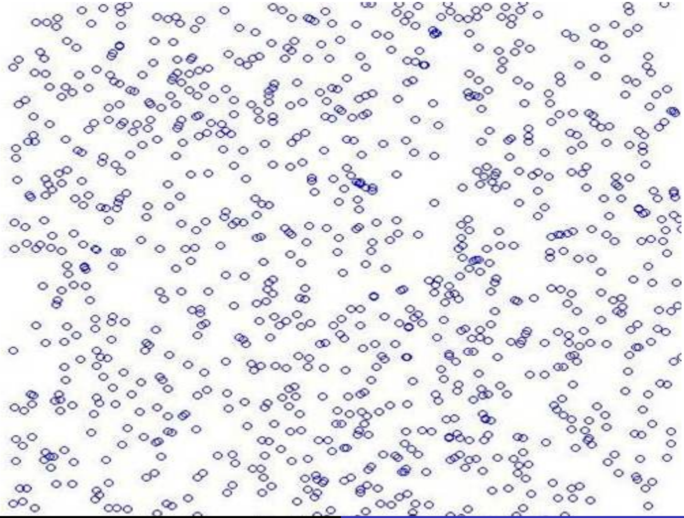
How do we get clusters of data?

How do we get clusters of data?

- Do we see clusters here ?

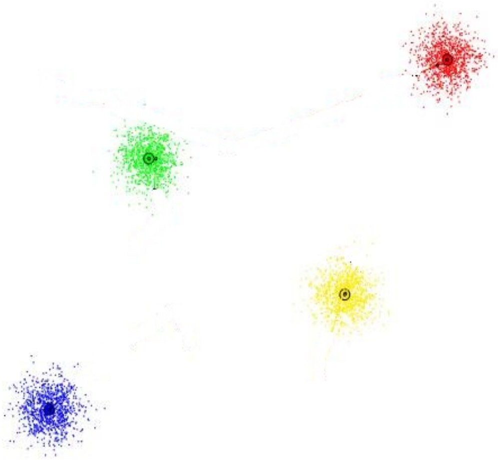
How do we get clusters of data?

- Do we see clusters here ?



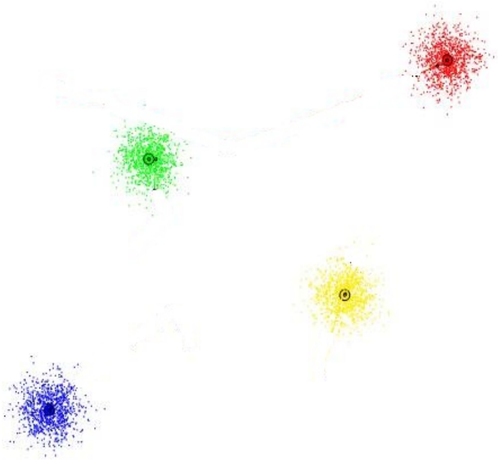
How do we get clusters of data?

- Do we see clusters here ?



How do we get clusters of data?

- Clusterable data $\sim$ Data from a mixture of distributions.

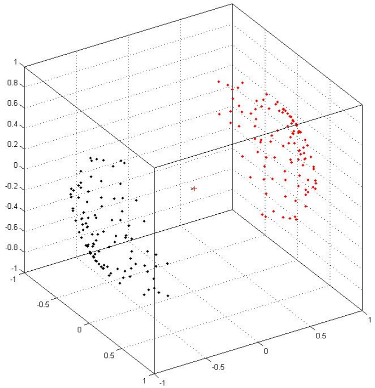


Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

What is its effect on algorithms, say clustering?

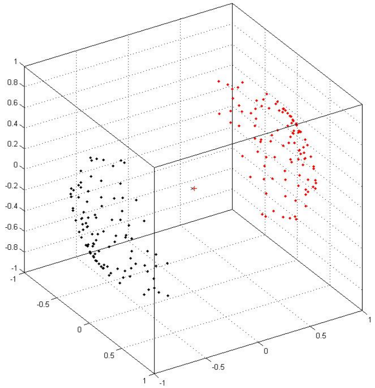
Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

What is its effect on algorithms, say clustering?



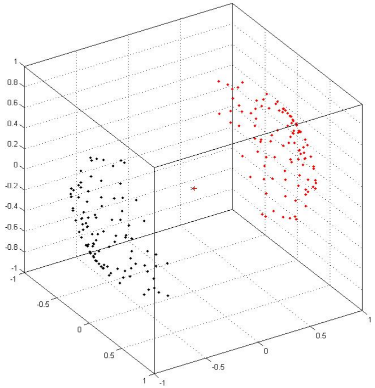
Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

Data are distributed as two well-separated clusters.

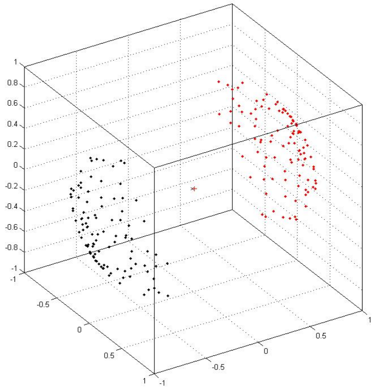


Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

Data are distributed on a 3D-sphere $\Longrightarrow \ell = \mathbf{2} < \mathbf{3} = m$

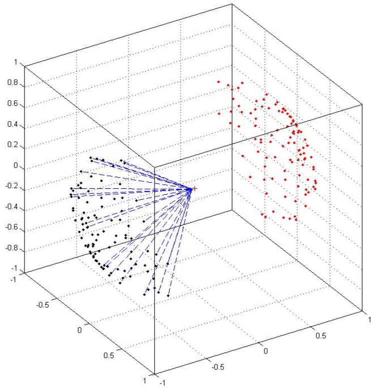


FCM proceeds by finding the centroid of the clusters.



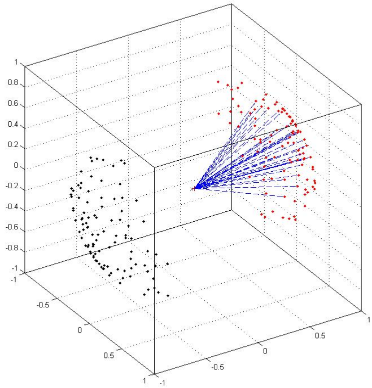
Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

FCM proceeds by finding the centroid of the clusters.



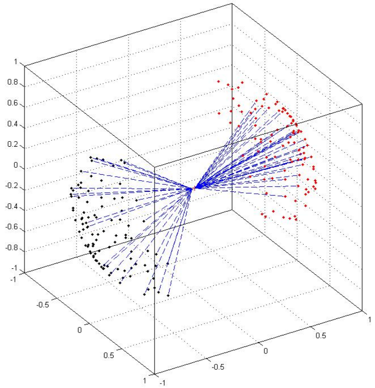
Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

FCM proceeds by finding the centroid of the clusters.



Intrinsic dimn far less than Embedding dimn - $\ell \ll m$

The centroids of both the clusters converge!!!



Interplay between the 3D's - 2

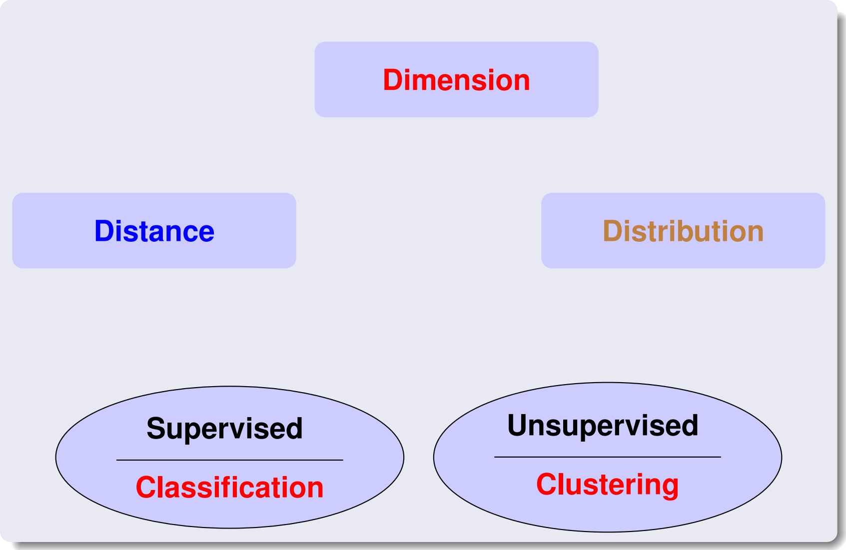
Interplay between the 3D's - 2

Dimension

Distance

Distribution

Interplay between the 3D's - 2



Interplay between the 3D's - 2

Dimension

Relation among Dimensions

Distance

Distribution

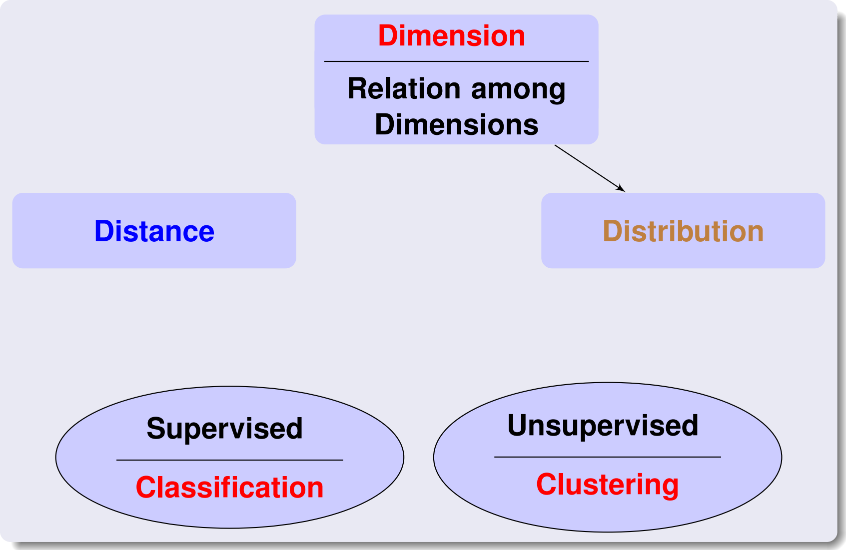
Supervised

Classification

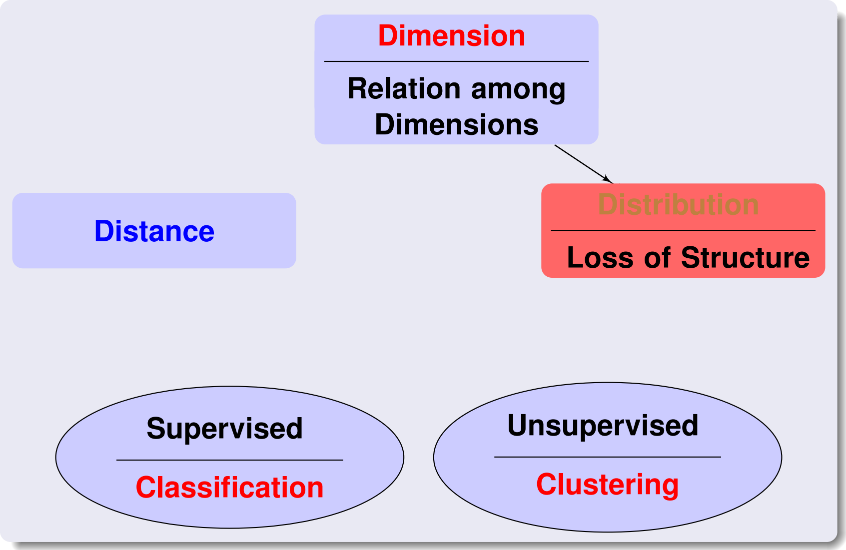
Unsupervised

Clustering

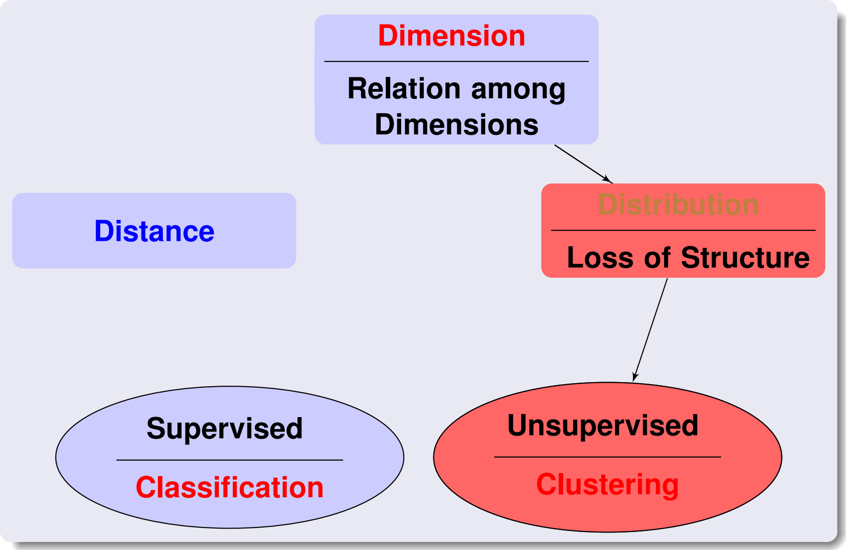
Interplay between the 3D's - 2



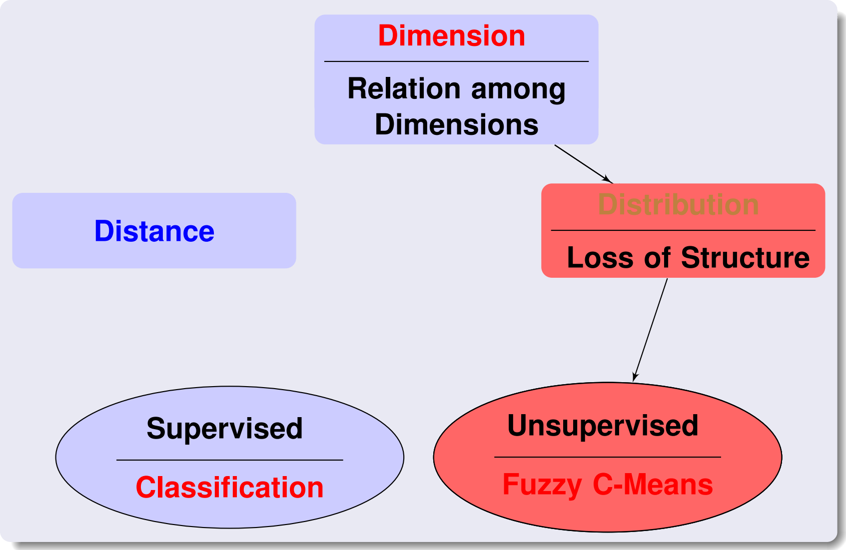
Interplay between the 3D’s - 2



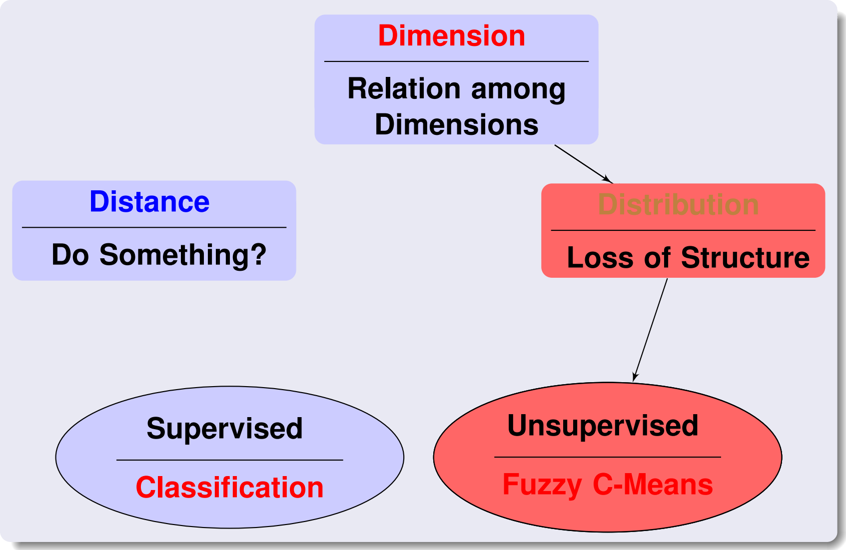
Interplay between the 3D's - 2



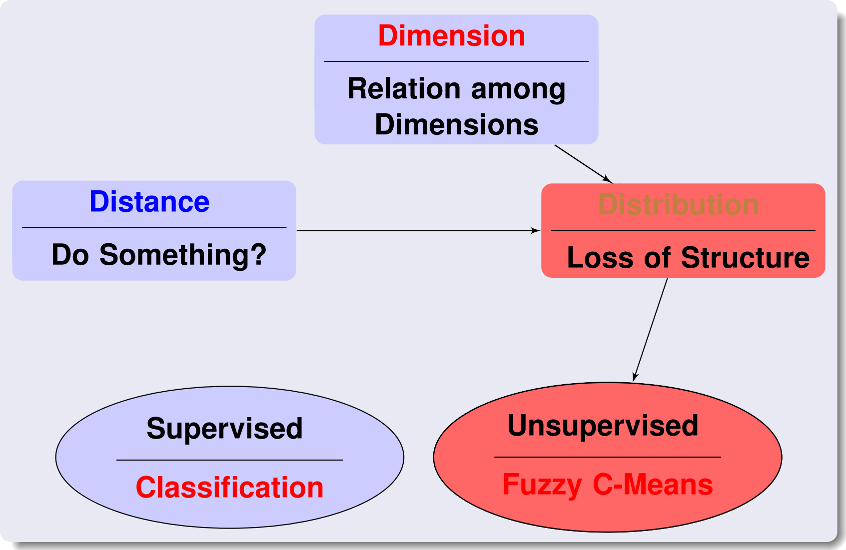
Interplay between the 3D's - 2



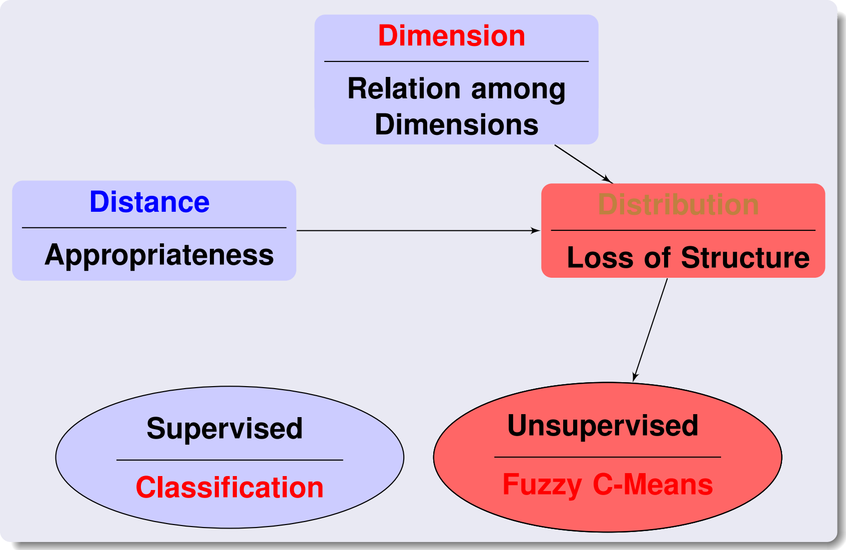
Interplay between the 3D's - 2



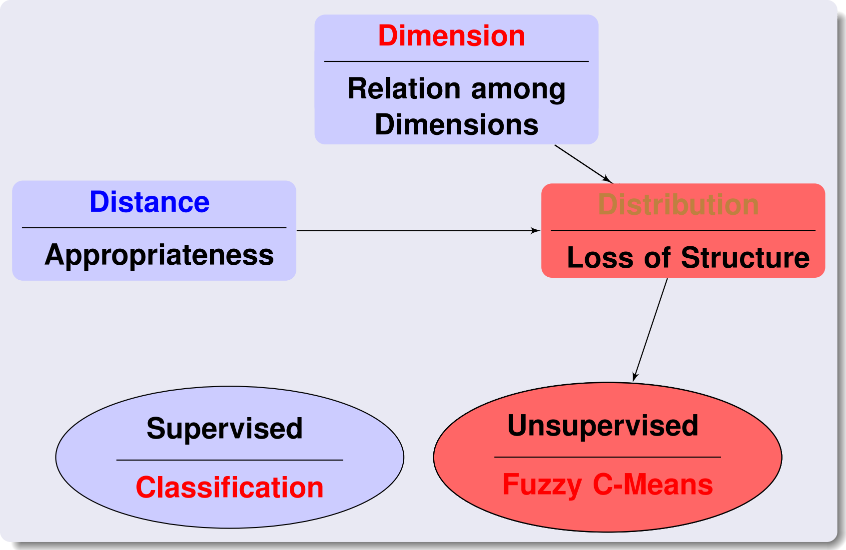
Interplay between the 3D's - 2



Interplay between the 3D's - 2



Interplay between the 3D's - 2



Clustering on the basis of correlation

Clustering on the basis of correlation

M.E. Houle et al.(**2010**)

- ”Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?”

Clustering on the basis of correlation

M.E. Houle et al.(**2010**)

- ”Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?”
- Used secondary similarity measures on the basis of the rankings ...

Clustering on the basis of correlation

M.E. Houle et al.(**2010**)

- ”Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?”
- Used secondary similarity measures on the basis of the rankings ...
- ... induced by a distance measure

Clustering on the basis of correlation

M.E. Houle et al.(**2010**)

- "Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?"
- Used secondary similarity measures on the basis of the rankings ...
- ... induced by a distance measure

Their recommendation

Clustering on the basis of correlation

M.E. Houle et al.(**2010**)

- ”Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?”
- Used secondary similarity measures on the basis of the rankings ...
- ... induced by a distance measure

Their recommendation

- Rank-based similarity measures.

Clustering on the basis of correlation

M.E. Houle et al.(**2010**)

- ”Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?”
- Used secondary similarity measures on the basis of the rankings ...
- ... induced by a distance measure

Their recommendation

- Rank-based similarity measures.
- Can result in *better* performance.

Clustering on the basis of correlation

M.E. Houle et al.(2010)

- ”Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?”
- Used secondary similarity measures on the basis of the rankings ...
- ... induced by a distance measure

Their recommendation

- Rank-based similarity measures.
- Can result in *better* performance.

Robust Rank Correlation Based Clustering (RaCoCl)

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.
- Classical measures: Spearman's ρ and Kendall's τ.

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.
- Classical measures: Spearman's ρ and Kendall's τ.

  Known to be not ideally suited in the presence of noise.

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.
- Classical measures: Spearman's ρ and Kendall's τ.

  Known to be not ideally suited in the presence of noise.
- Fuzzy/robust γ (Bodenhofer & Klawonn, **2008**).

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.
- Classical measures: Spearman's ρ and Kendall's τ.

  Known to be not ideally suited in the presence of noise.
- Fuzzy/robust γ (Bodenhofer & Klawonn, **2008**).

  Generalization of Goodman and Kruskal's γ.

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.
- Classical measures: Spearman's ρ and Kendall's τ.

  Known to be not ideally suited in the presence of noise.
- Fuzzy/robust γ (Bodenhofer & Klawonn, **2008**).

  Generalization of Goodman and Kruskal's γ.

  Shown to be more robust to noise in the data.

Robust Rank Correlation Based Clustering (RaCoCl)

Brief overview [Krone et al, 2013]

- Based on the rank correlation of every pair of points w.r.t. all other points in the dataset.
- Any rank correlation measure can be used.
- Classical measures: Spearman's ρ and Kendall's τ.

  Known to be not ideally suited in the presence of noise.
- Fuzzy/robust γ (Bodenhofer & Klawonn, **2008**).

  Generalization of Goodman and Kruskal's γ.

  Shown to be more robust to noise in the data.

Robust Rank Correlation Based Clustering (RaCoCl)

Our Findings

- Appears to work well, *especially* with HD data.

Our Findings

- Appears to work well, *especially* with HD data.

FCM was shown to break down in those cases.

Robust Rank Correlation Based Clustering (RaCoCl)

Our Findings

- Appears to work well, *especially* with HD data.

  FCM was shown to break down in those cases.

- Seems well suited for sparsely populated data.

Robust Rank Correlation Based Clustering (RaCoCl)

Our Findings

- Appears to work well, *especially* with HD data.

  FCM was shown to break down in those cases.

- Seems well suited for sparsely populated data.

Take Home Message!!

Robust Rank Correlation Based Clustering (RaCoCl)

Our Findings

- Appears to work well, *especially* with HD data.

  FCM was shown to break down in those cases.

- Seems well suited for sparsely populated data.

Take Home Message!!

Fuzzy Set Theory (still) has its role to play!!

Interplay between the 3D's - 2

Interplay between the 3D's - 2

Dimension

Distance

Distribution

Interplay between the 3D's - 2

Dimension

Distance

Distribution

Supervised

Classification

Unsupervised

Clustering

Interplay between the 3D's - 2

Dimension

Relation among Dimensions

Distance

Distribution

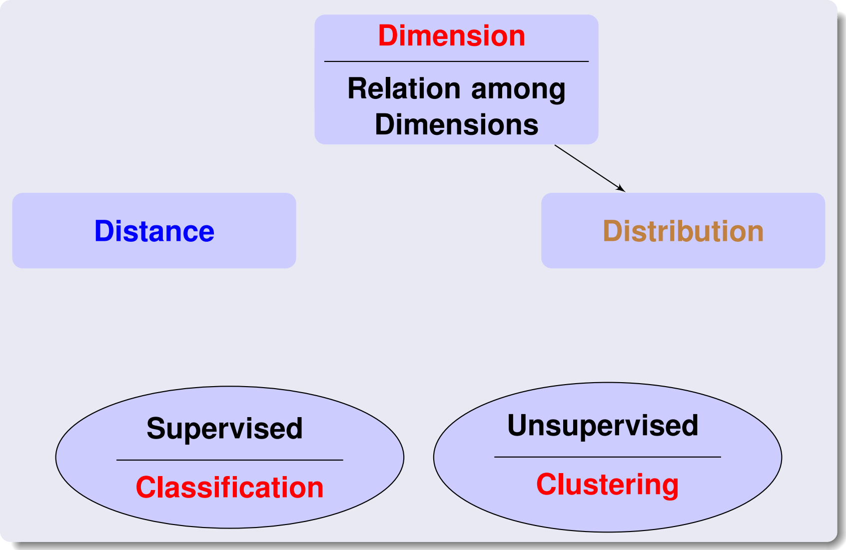
Supervised

Classification

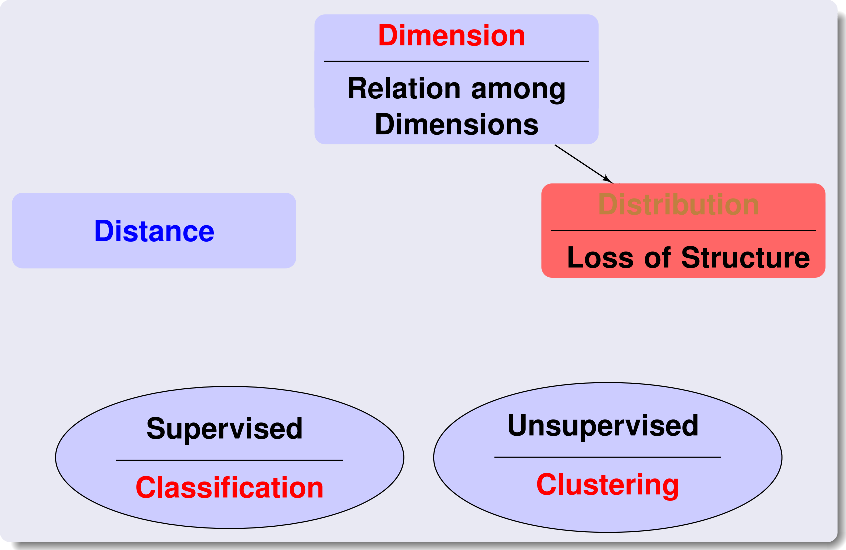
Unsupervised

Clustering

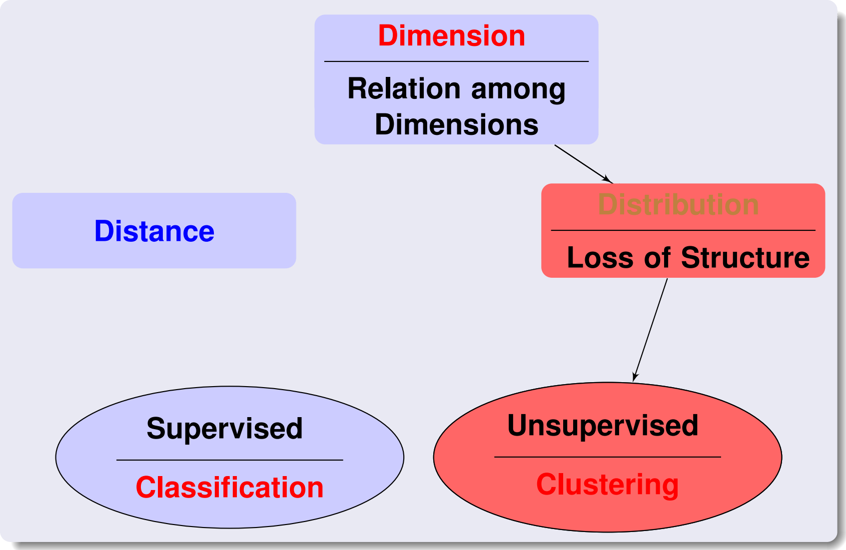
Interplay between the 3D's - 2



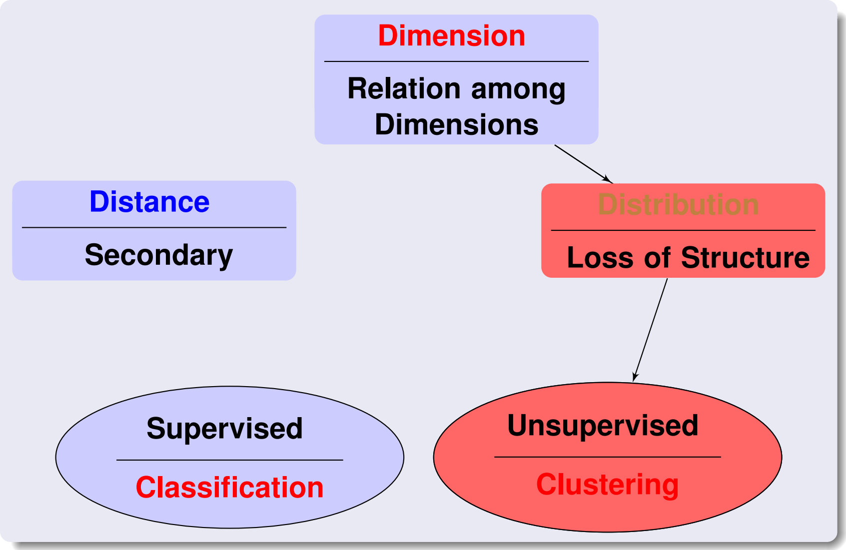
Interplay between the 3D's - 2



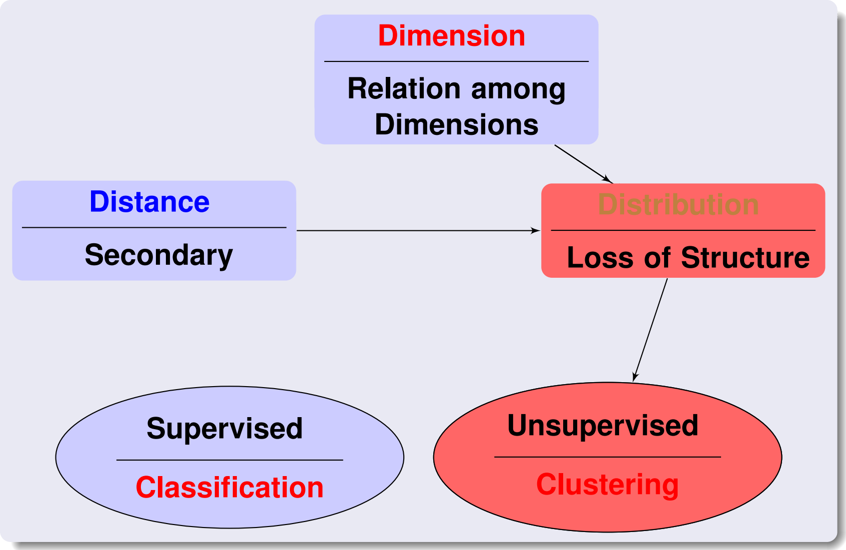
Interplay between the 3D's - 2



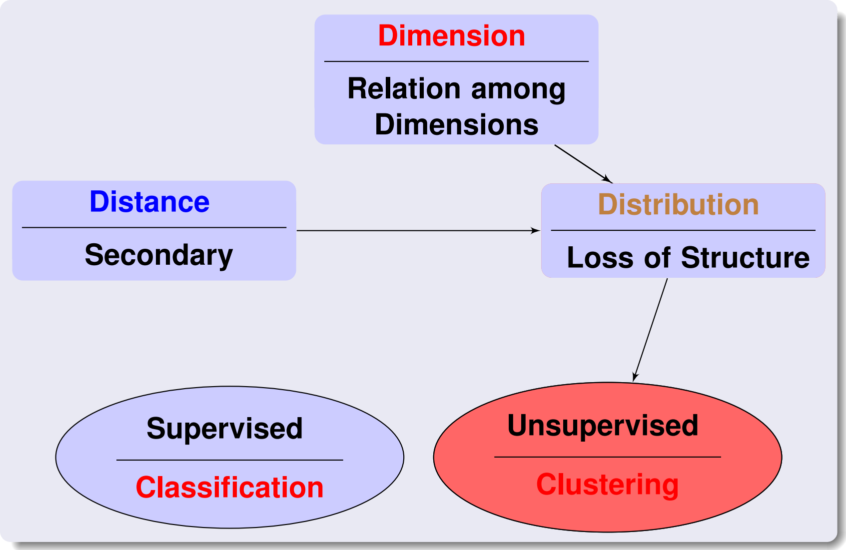
Interplay between the 3D's - 2



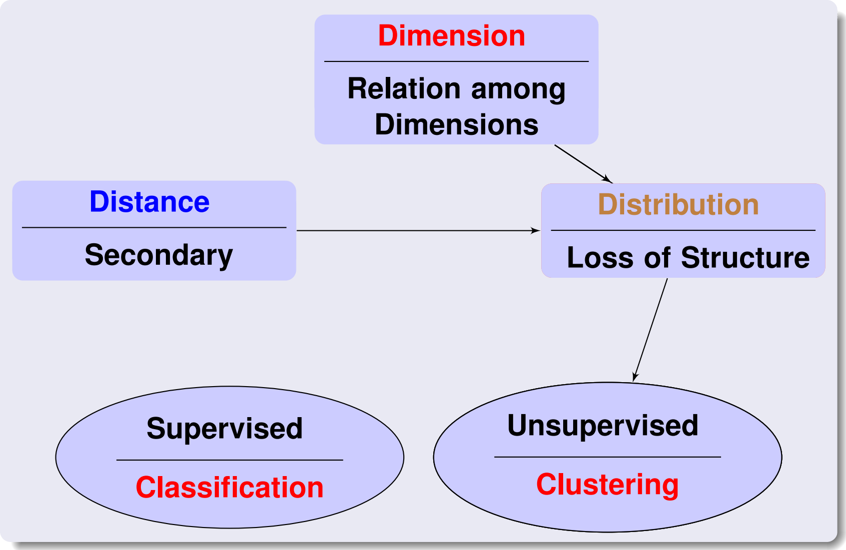
Interplay between the 3D's - 2

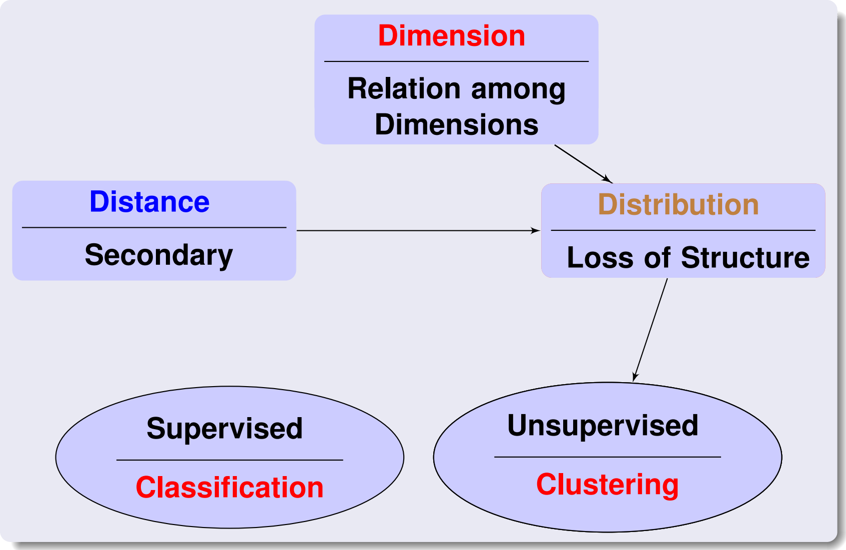


Interplay between the 3D's - 2



Interplay between the 3D's - 2



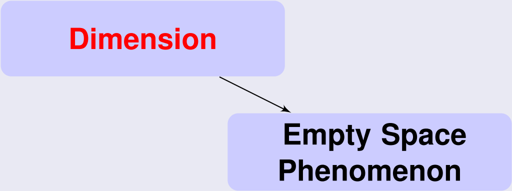

Dimension
Relation among Dimensions
Distance
Secondary
Distribution
Loss of Structure
Supervised
Classification
Unsupervised
Clustering

Summary of the talk

Summary of the talk

Dimension

Summary of the talk



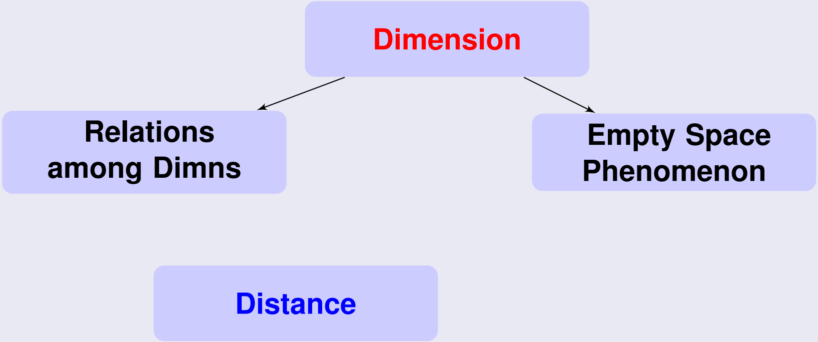
Summary of the talk

Dimension

Relations among Dimns

Empty Space Phenomenon

Summary of the talk



Summary of the talk

Dimension

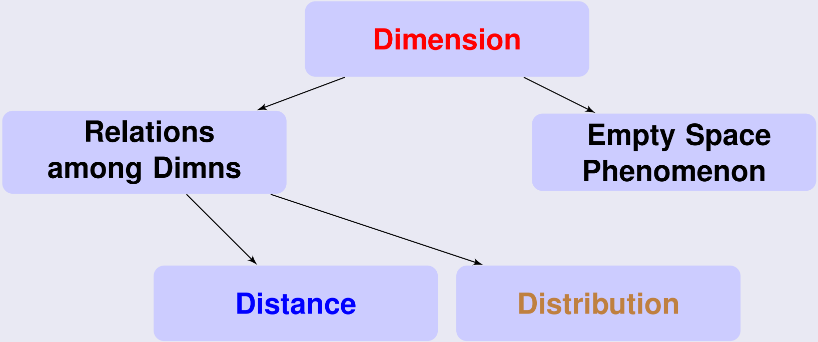
Relations among Dimns

Empty Space Phenomenon

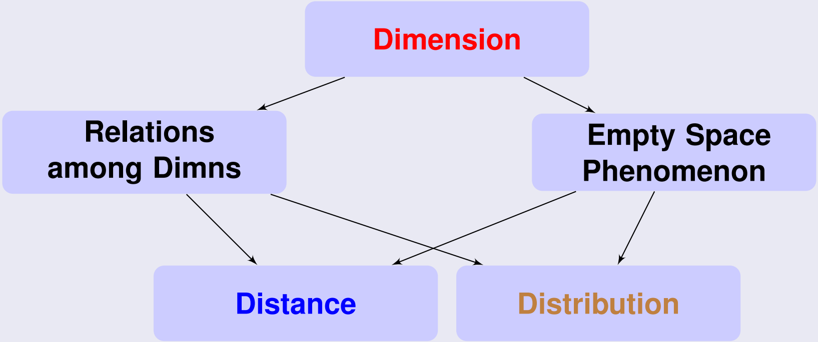
Distance

Distribution

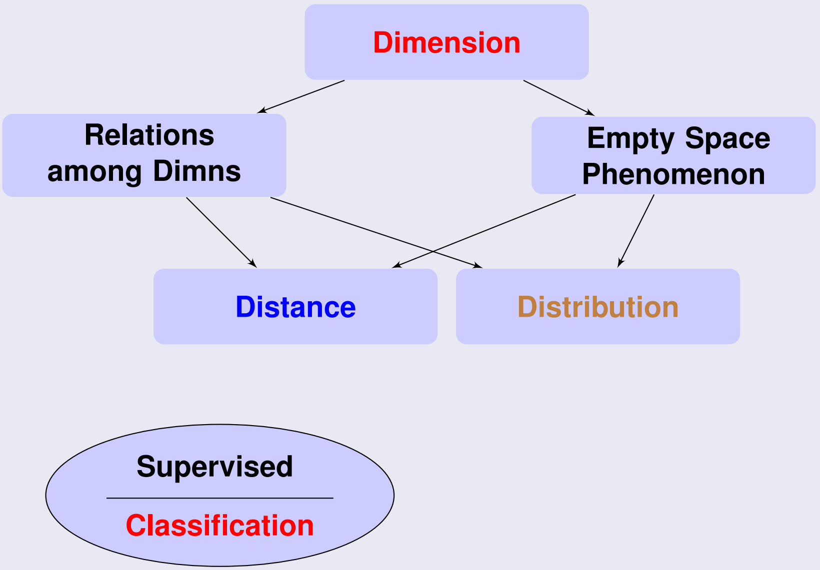
Summary of the talk



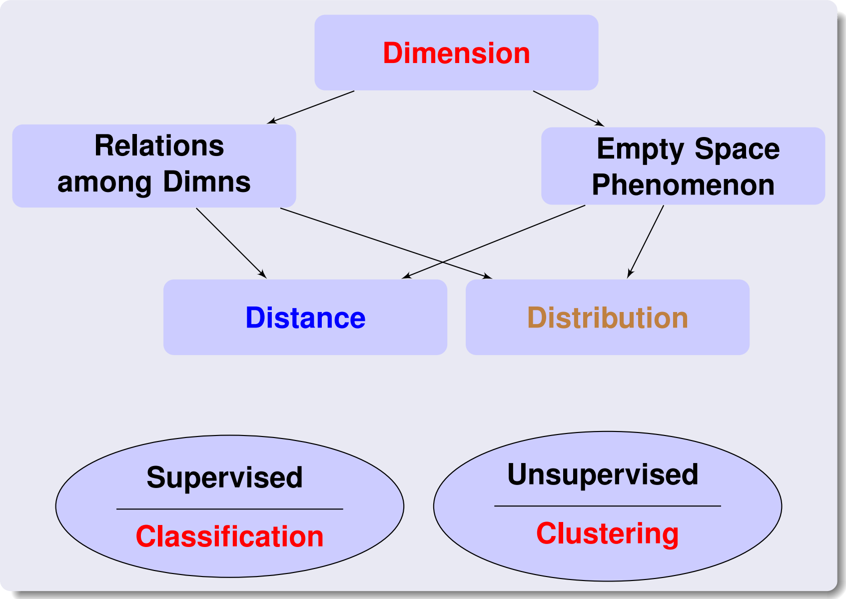
Summary of the talk



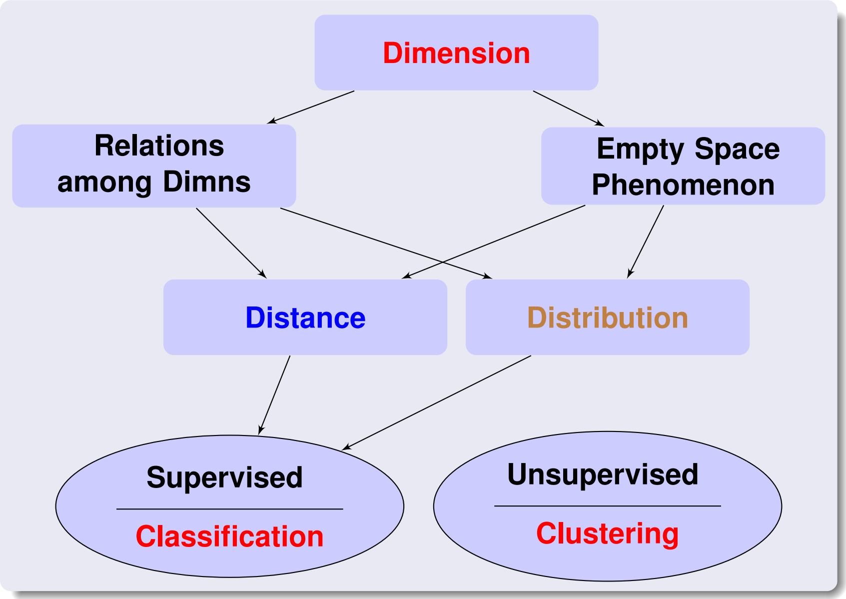
Summary of the talk



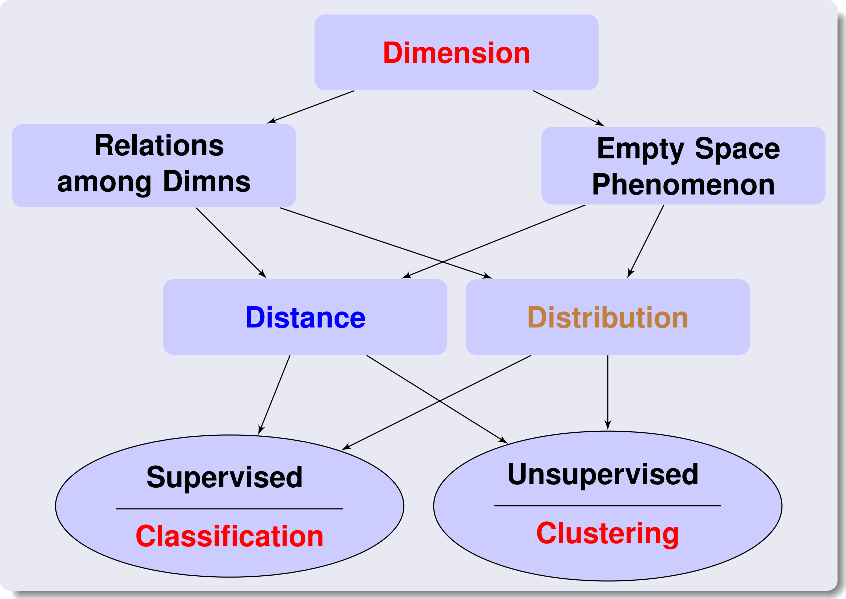
Summary of the talk



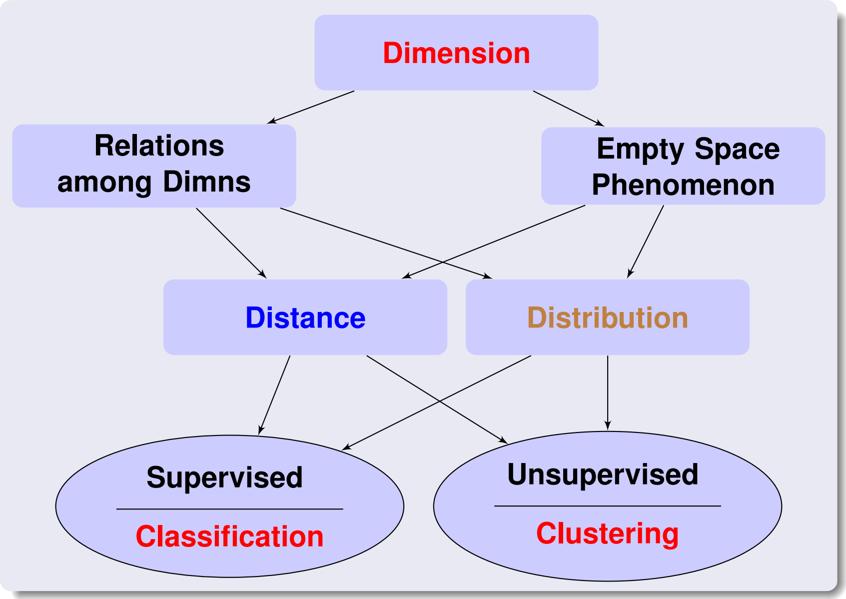
Summary of the talk



Summary of the talk



Summary of the talk



References I

C. C. Aggarwal, A. Hinneburg, D. A. Keim (2001)
On the Surprising Behavior of Distance Metrics in High Dimensional Space.
Database Theory ICDT 2001.

Bellmann, R. (1961)
Adaptive Control Processes: A Guided Tour, Princeton Univ. Press, Princeton.

K. Bennett, U. Fayyad, and D. Geiger (1999)
Density-Based Indexing for Approximate Nearest Neighbor Queries
KDD-99.

Bishop, C.M. (2006)
Pattern Recognition and Machine Learning, Springer.

References II

- Beyer, K.S., Goldstein, J., Ramakrishnan, R. and Shaft, U. (1999)
  When is nearest neighbor meaningful?, ICDT 1999.
- B. Jayaram and F. Klawonn (2012)
  Can unbounded distance metrics mitigate the curse of dimensionality?
  Int. Jl. of Data Mining, Modelling and Mgmt., 4-361 - 383.
- Radovanovic, M., Nanopoulos, A., Ivanovic, M. (2010)
  Hubs in space: Popular nearest neighbors in high-dimensional data.
  Journal of Machine Learning Research 11, 2487–2531
- R. Winkler, F. Klawonn and R. Kruse. (2011)
  Fuzzy C-Means in high dimensional spaces
  Fuzzy System Applications, 1:1-17.

Extremely grateful to ...

Extremely grateful to ...

Prof. Frank Klawonn

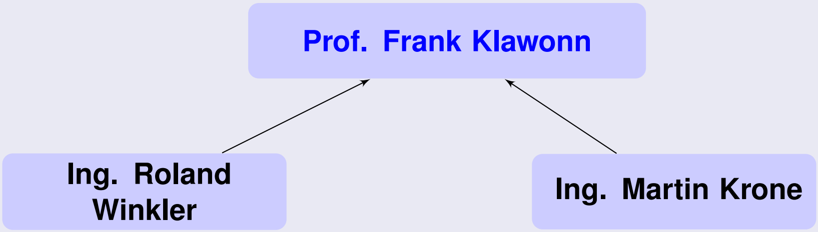
Extremely grateful to ...

Prof. Frank Klawonn

Ing. Roland Winkler

Ing. Martin Krone

Extremely grateful to ...



Extremely grateful to ...

Alexander von Humboldt
Stiftung/Foundation

Prof. Frank Klawonn

Ing. Roland Winkler

Ing. Martin Krone

Extremely grateful to ...

Prof. Radko Mesiar

Alexander von Humboldt
Stiftung / Foundation

Prof. Frank Klawonn

Ing. Roland Winkler

Ing. Martin Krone

Extremely grateful to ...

Prof. Radko Mesiar

Alexander von Humboldt
Stiftung/Foundation

Prof. Frank Klawonn

Ing. Roland Winkler

Ing. Martin Krone

Prof. Radko Mesiar

Alexander von Humboldt
Stiftung/Foundation

Prof. Frank Klawonn

Ing. Roland Winkler

Ing. Martin Krone

Thanks for your patient listening !!!

Thanks for your patient listening !!!

Any Questions ???