DS501: Graph Data and High Dimensional Data

Prof. Randy Paffenroth
rcpaffenroth@wpi.edu

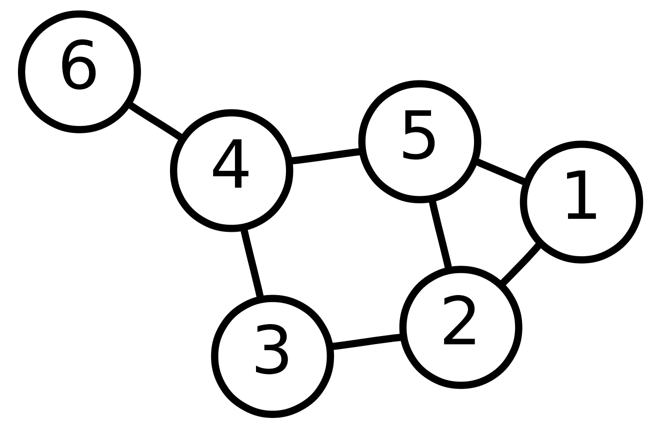
Worcester Polytechnic Institute

Announcements

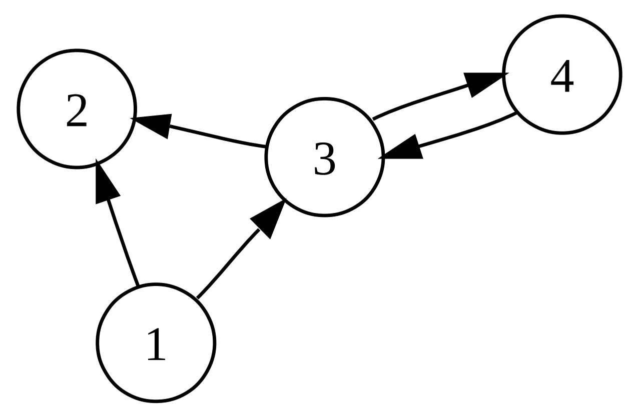
- Case Study 4 is posted!
  - The “Shark Tank”
  - Let's take a look at it.

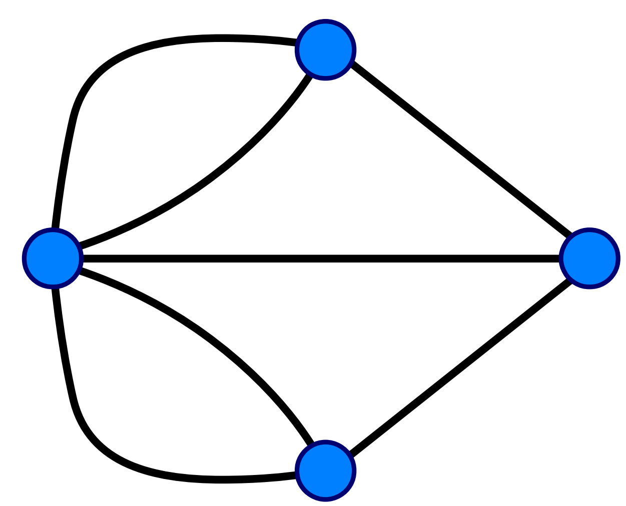
Graph Data

Types of graphs

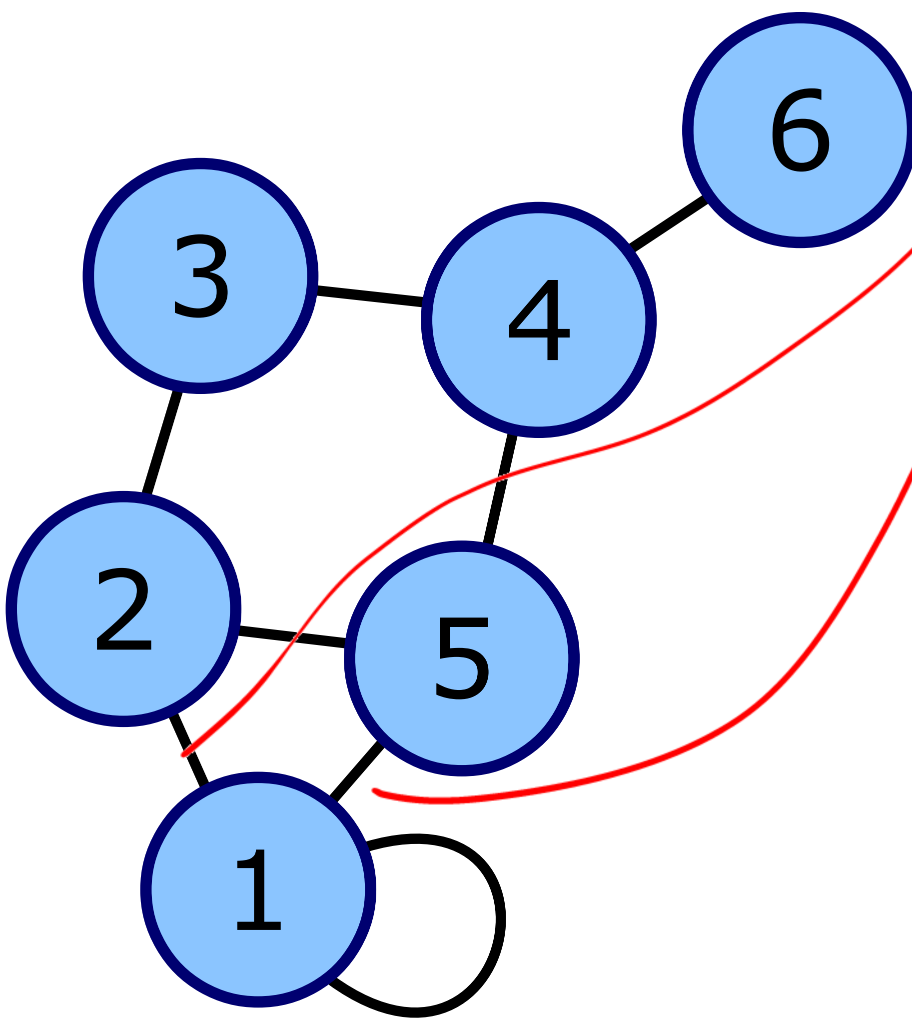


- Graph types
  - Undirected
  - Directed
  - Mixed
  - Multi-graphs





Two views of a graph

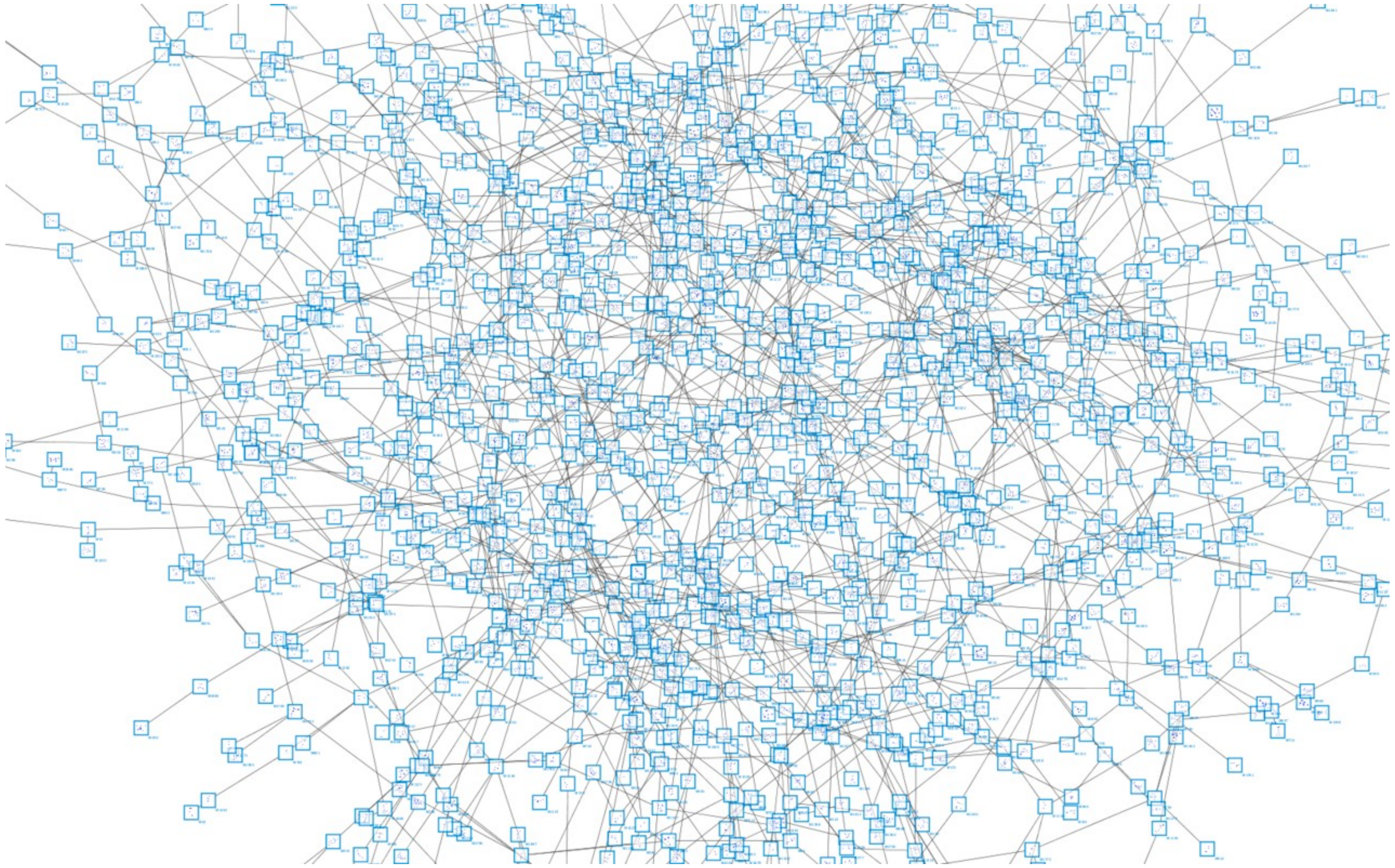


Adjacency matrix

$$
\begin{array}{c}
 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6
\end{array}
\begin{array}{c}
\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{pmatrix}
2 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}
$$

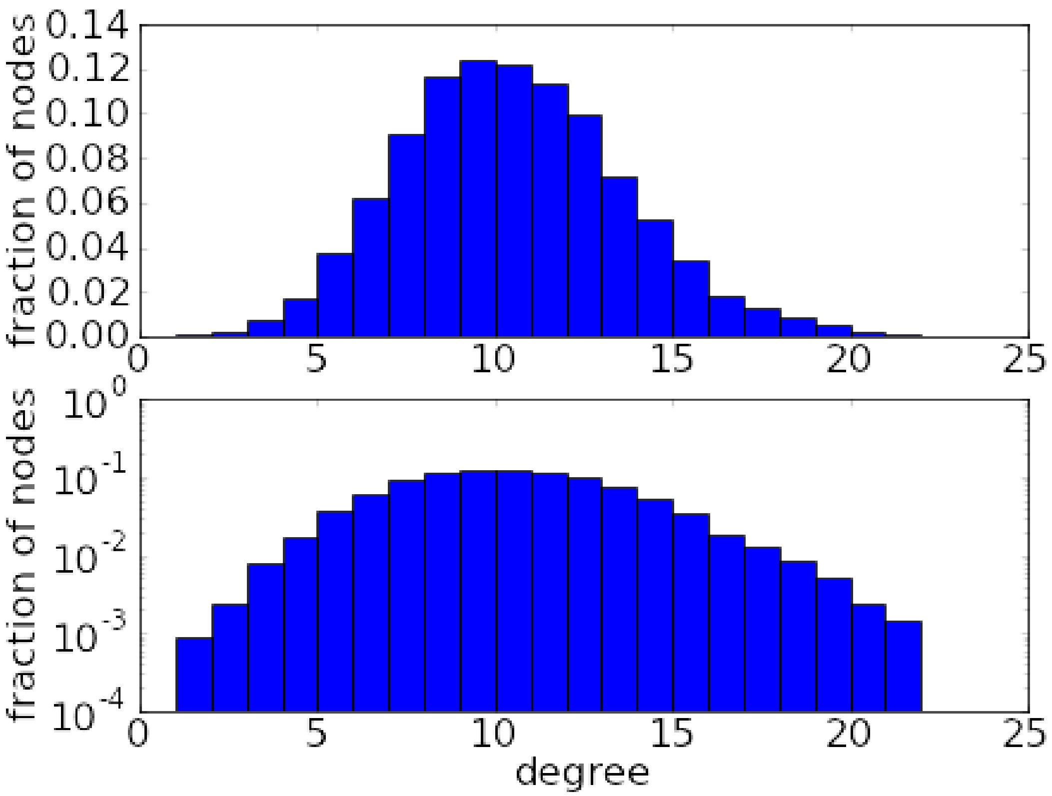
Properties of a Graph

- Degree Distribution
- Structural Properties



Degree Histogram

- Outdegree of a vertex = # outgoing edges
- For each number d, let n(d) = # vertices with outdegree d

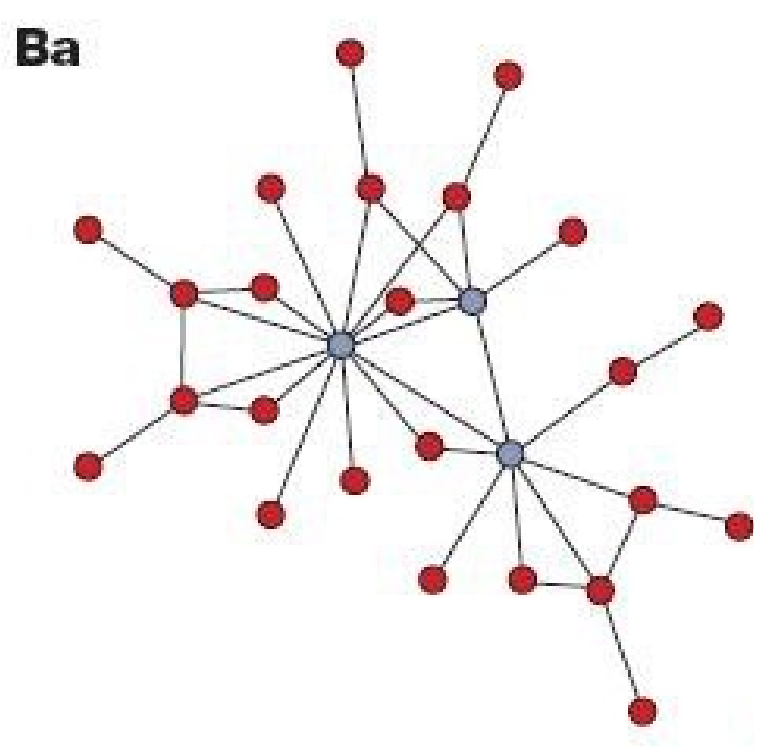


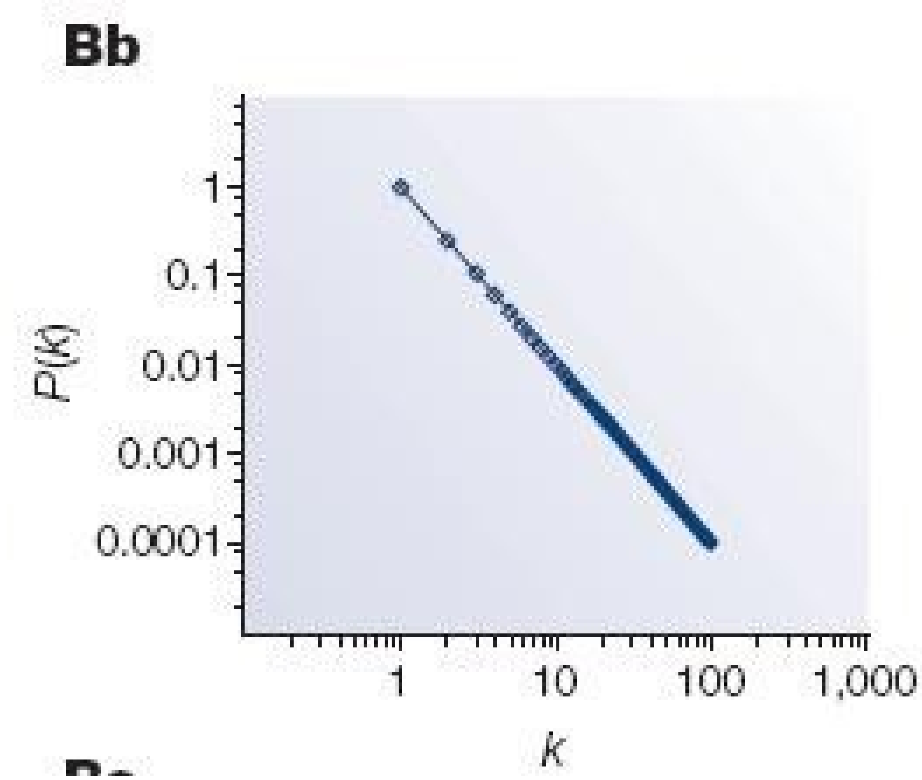
Degree Distribution

- Are there **large graphs** with “interesting” degree distributions?

Zipf Distribution (power law)

- n(d) ~ 1/ d^x for some value x > 0
- Human-generated data has Zipf distribution: letters in alphabet, words in vocabulary.
- In log-log scale





Distribution of the Web

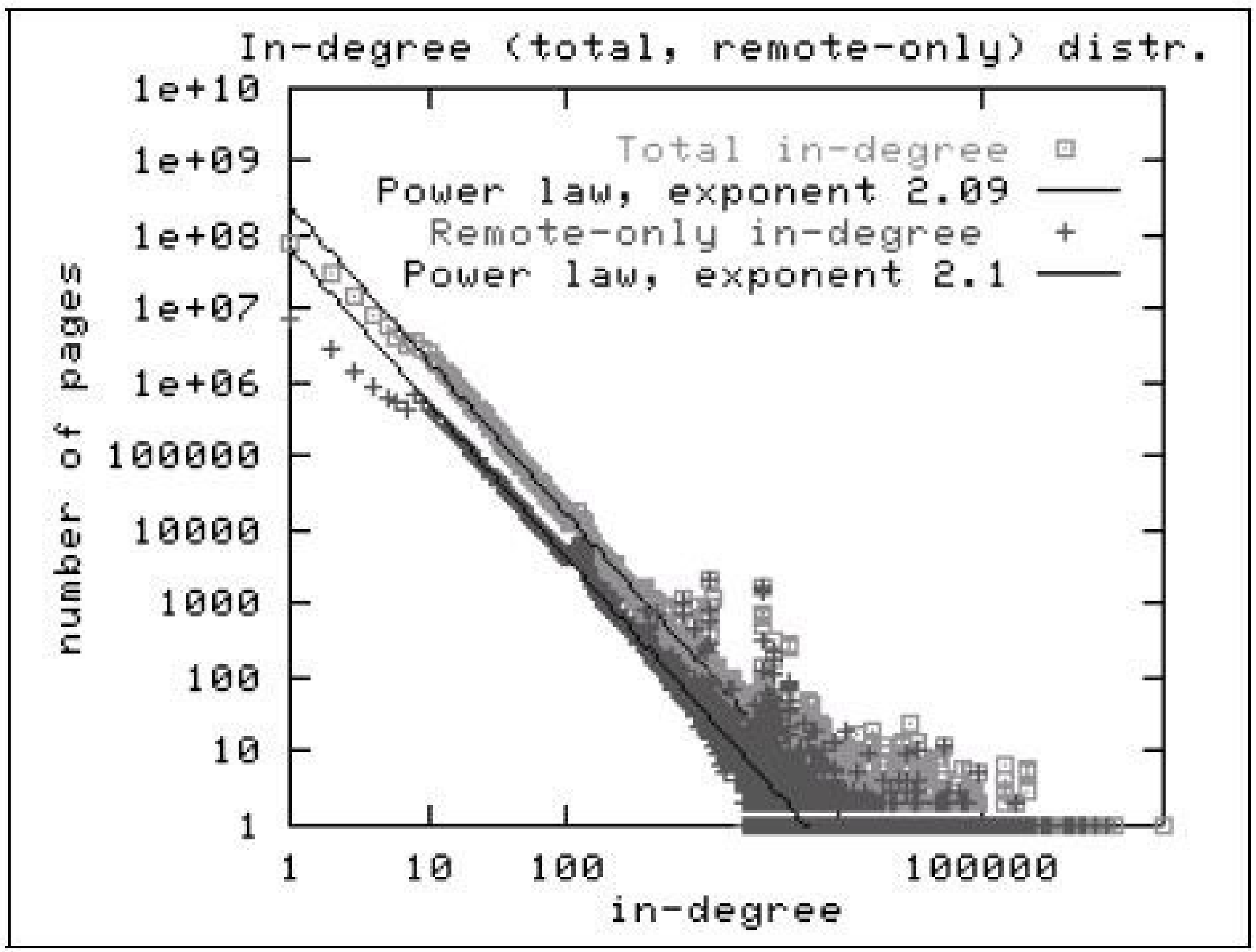


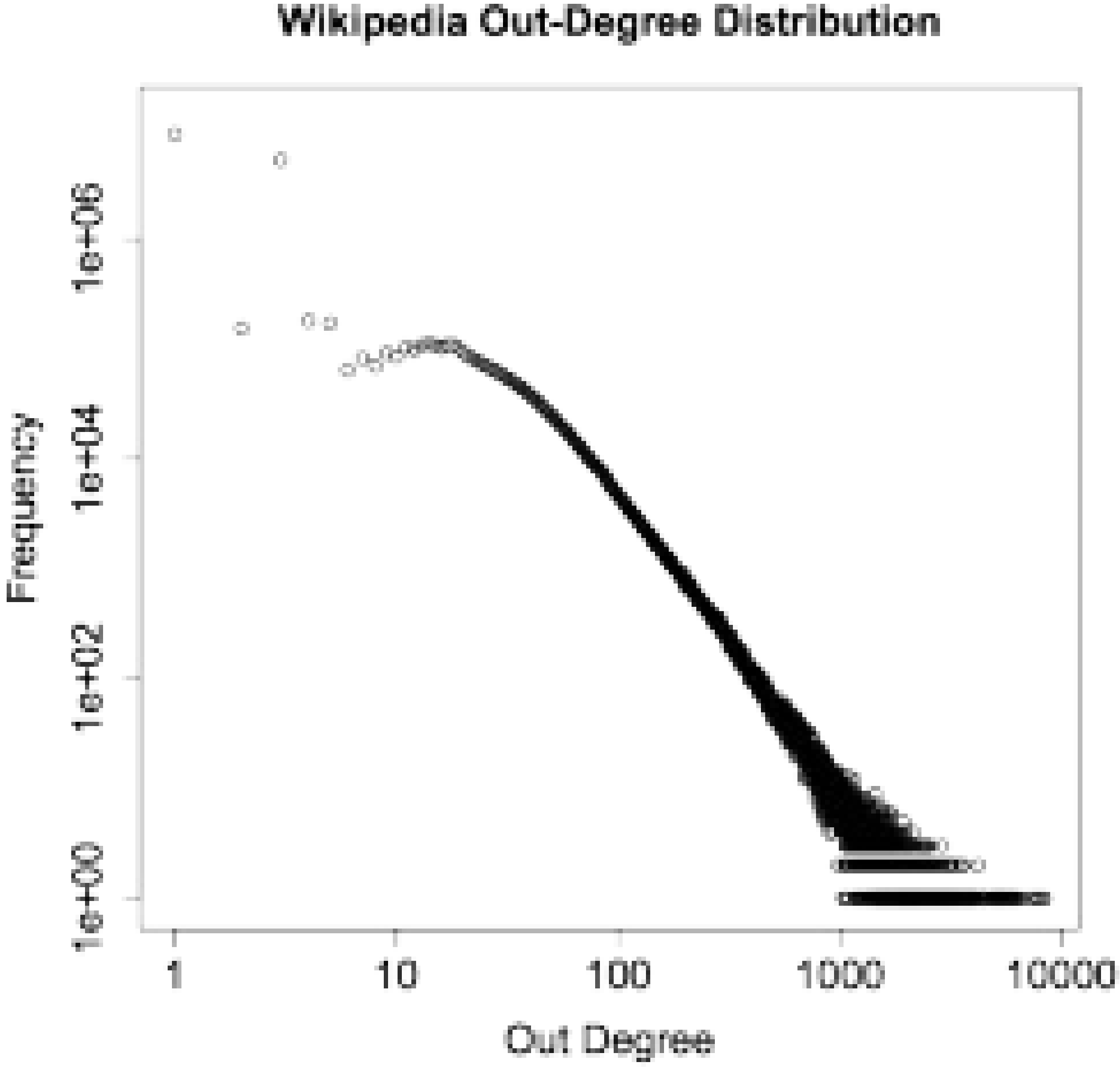
Figure 2: In-degree distribution.

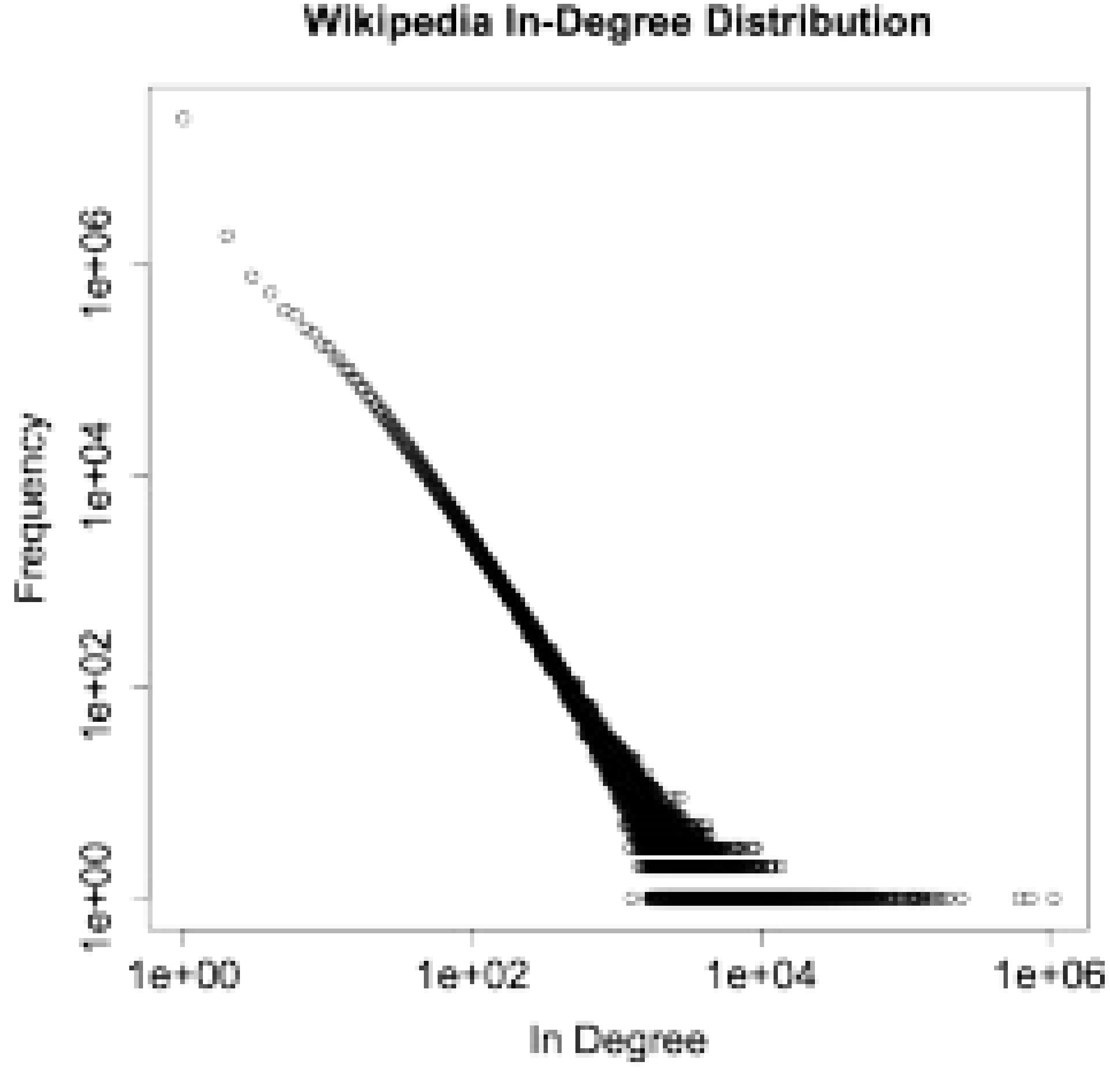
Late 1990's
200M Webpages

Exponential ?

Power Law?

Distribution of Wikipedia



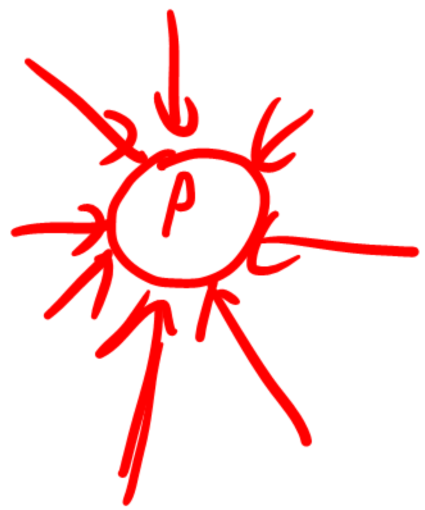


Evaluating the Nodes

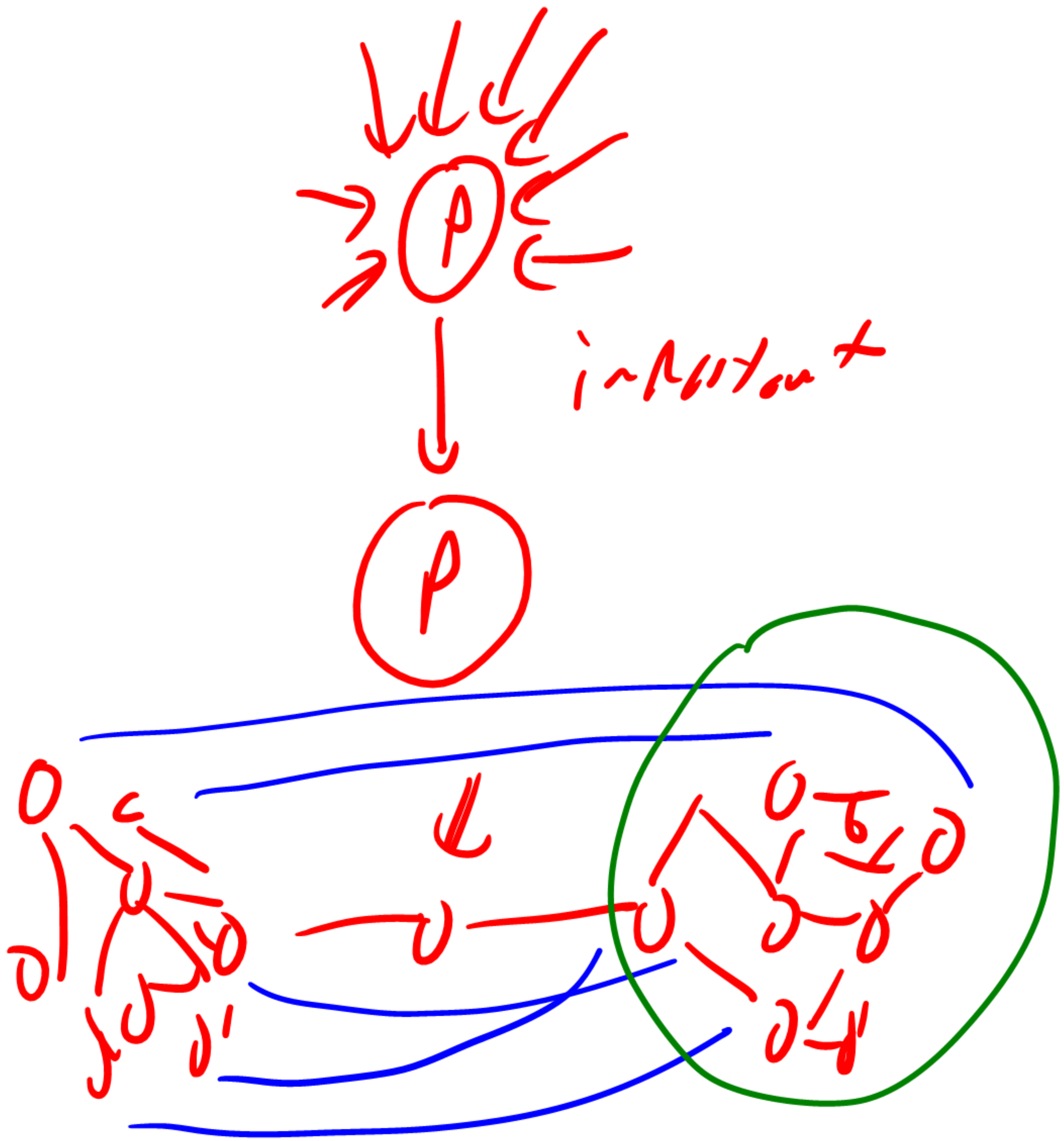
- How would you define the important nodes in a graph?

Following – edges

people – vertices



v.s.

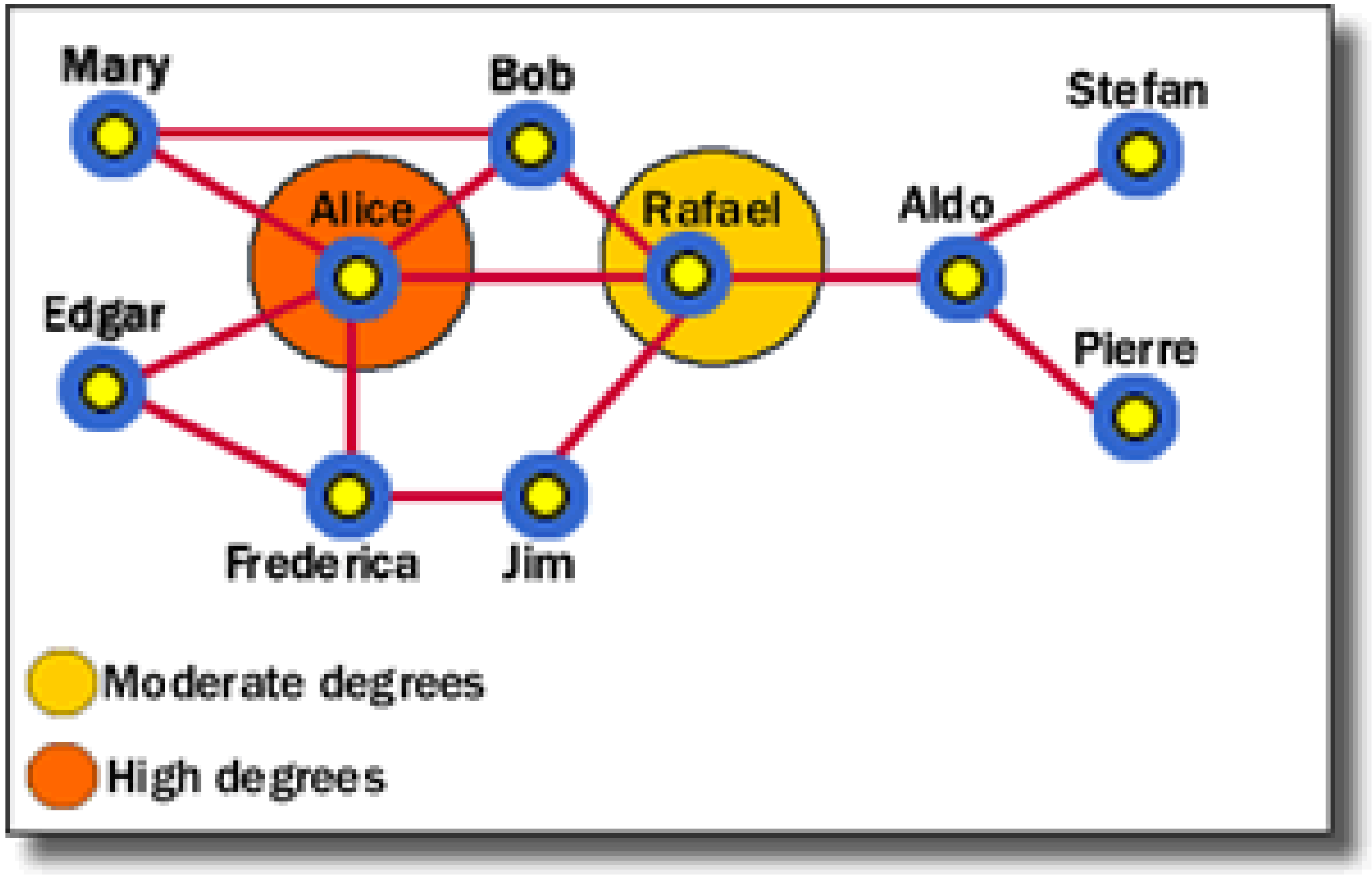


Evaluating the Nodes

- Degree Centrality
- Betweenness : Number paths
- Eigenvector Centrality (i.e, PageRank)

Node: **Degree** Centrality

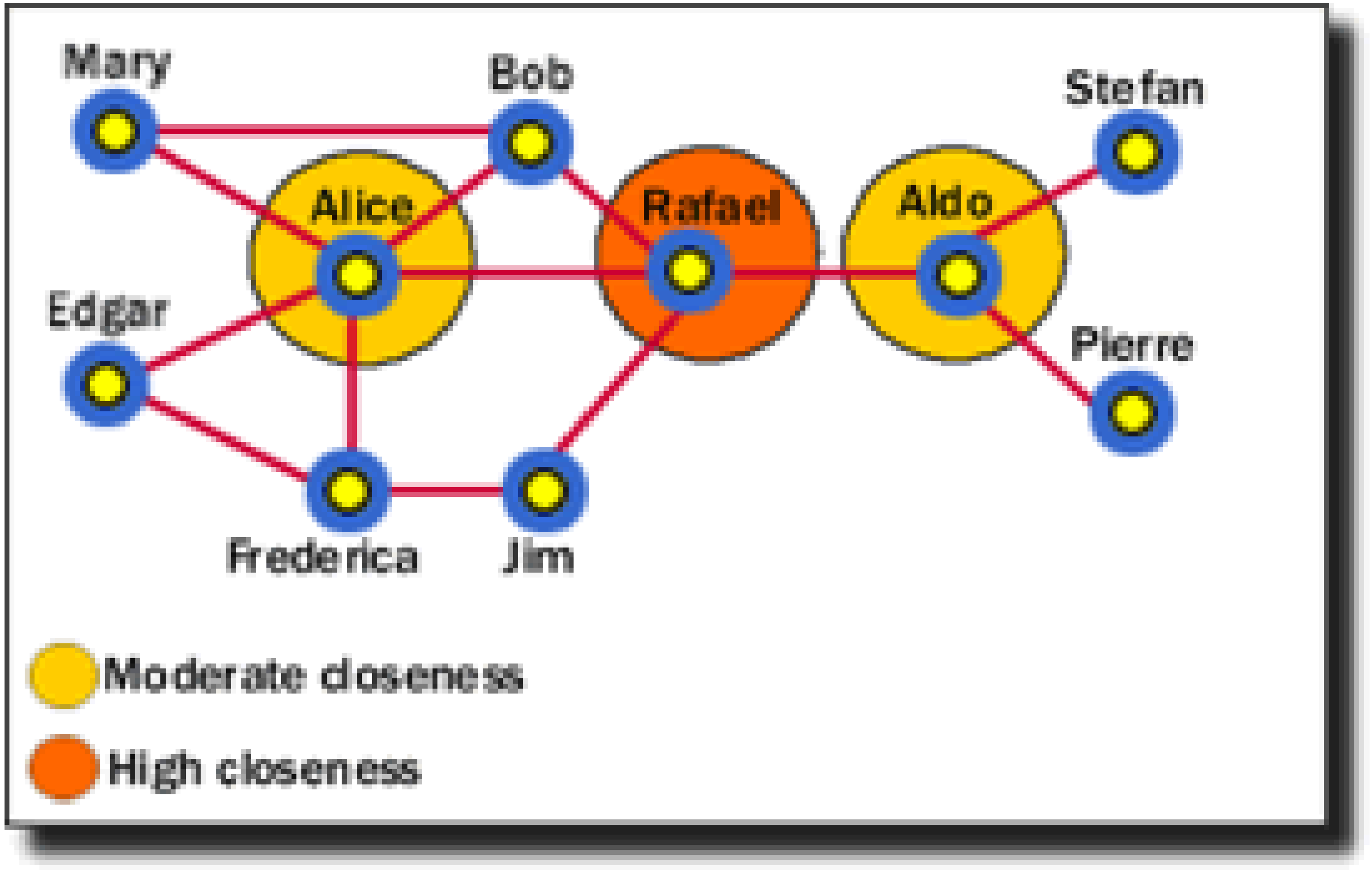
- **Degree** Centrality: number of edges



Pros: Is generally an active player in the network
Cons: ignore the whole graph

Node: **Closeness** Centrality

- **Closeness** of a vertex: the inverse of the average length of all its shortest paths



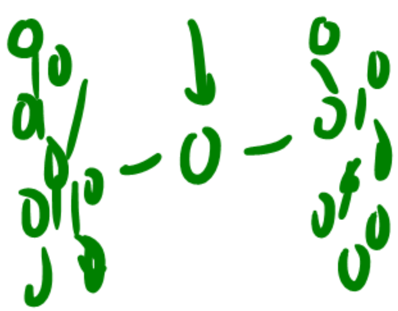
Pros: Is close to other entities

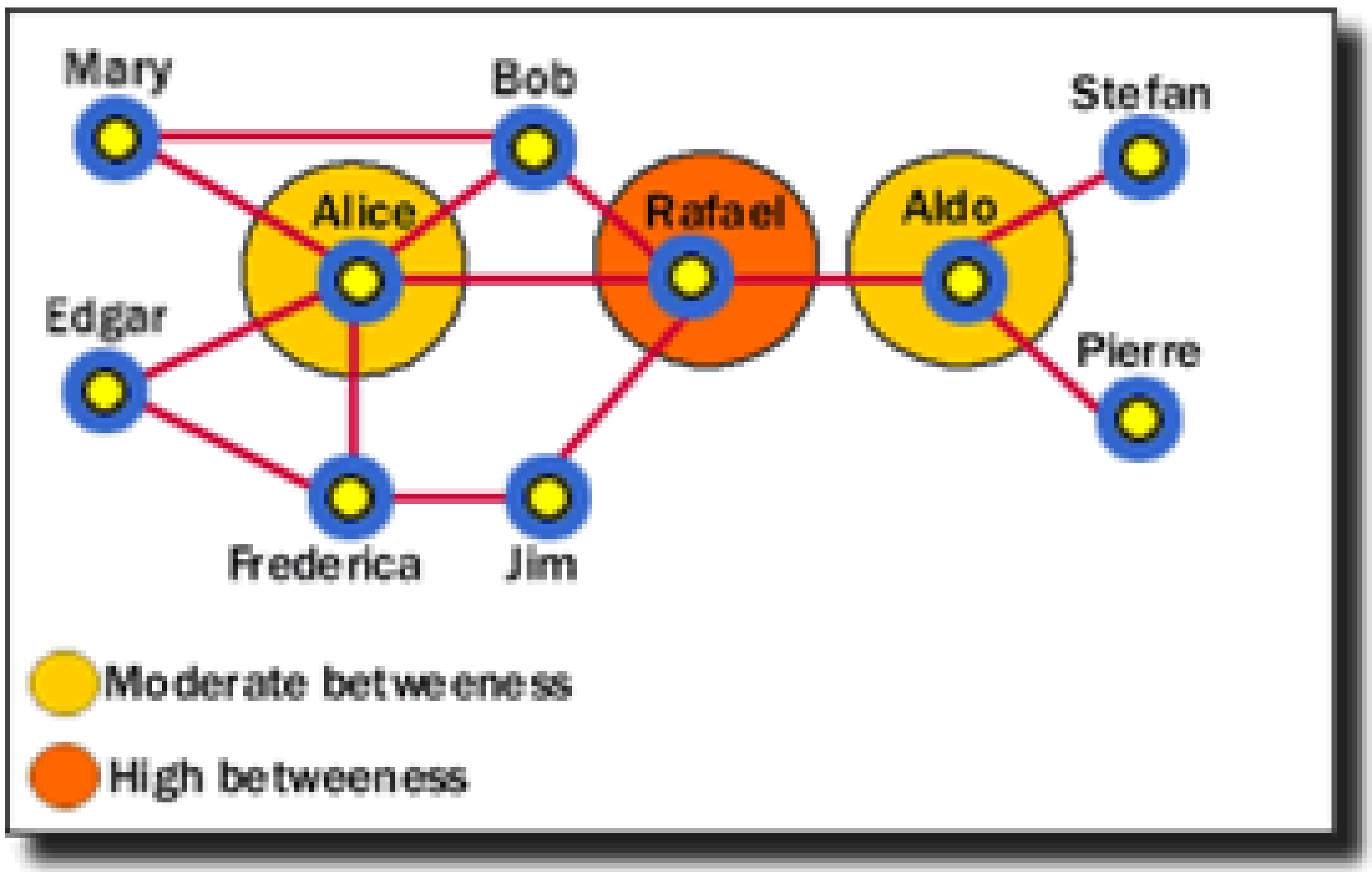
Cons: ignore the importance of different nodes

Node: **Betweenness** Centrality

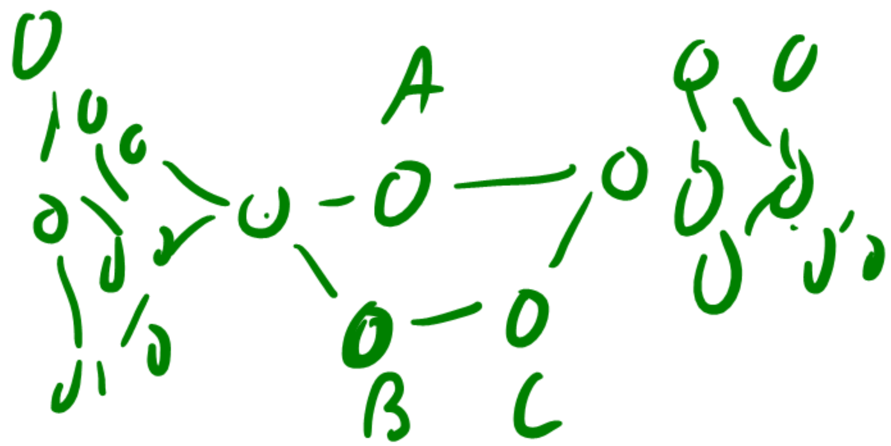
- **Betweeness** of a vertex *v*: the fraction of all shortest paths that pass through *v*.

High





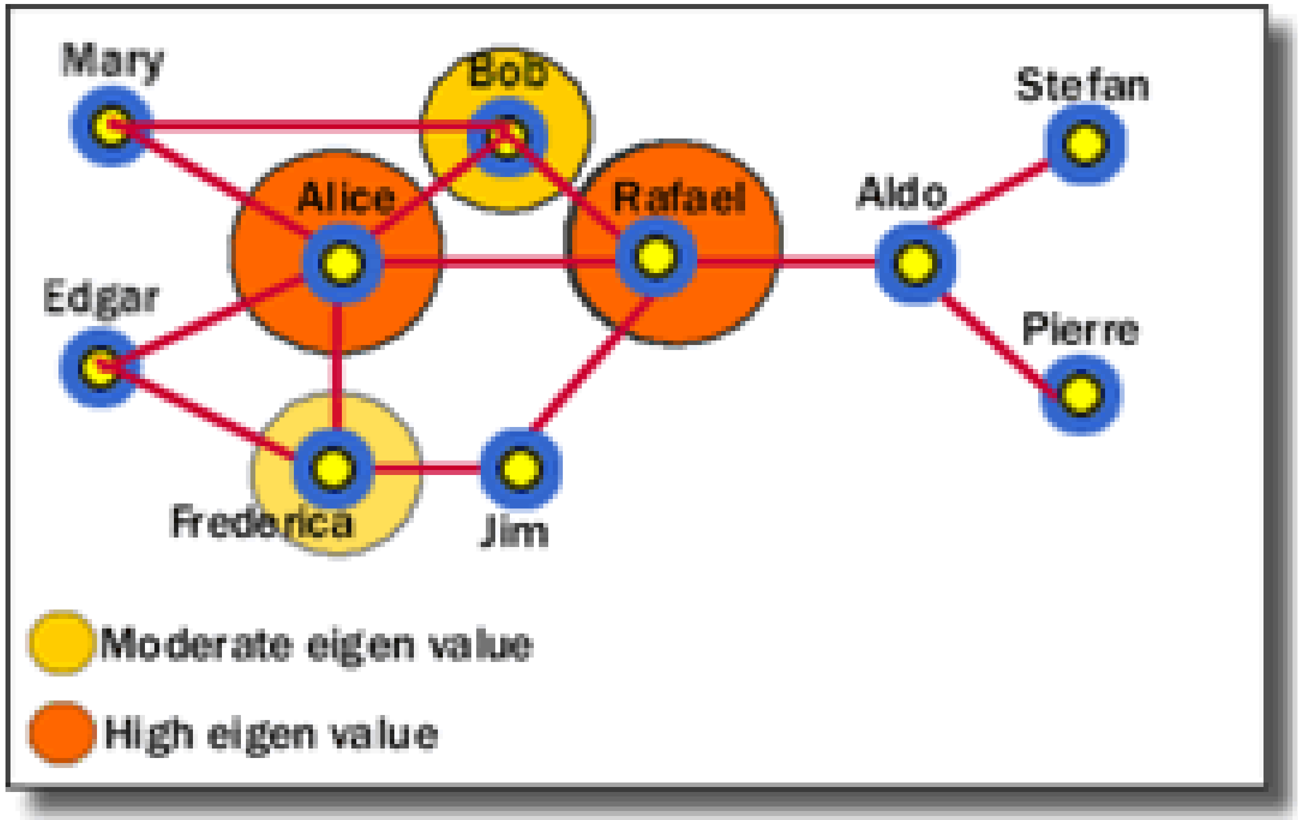
Pros: represents a single point of failure, has a greater amount of influence over what happens in a network.
Cons: ignore the importance of different nodes



which has highest betweenness centrality?

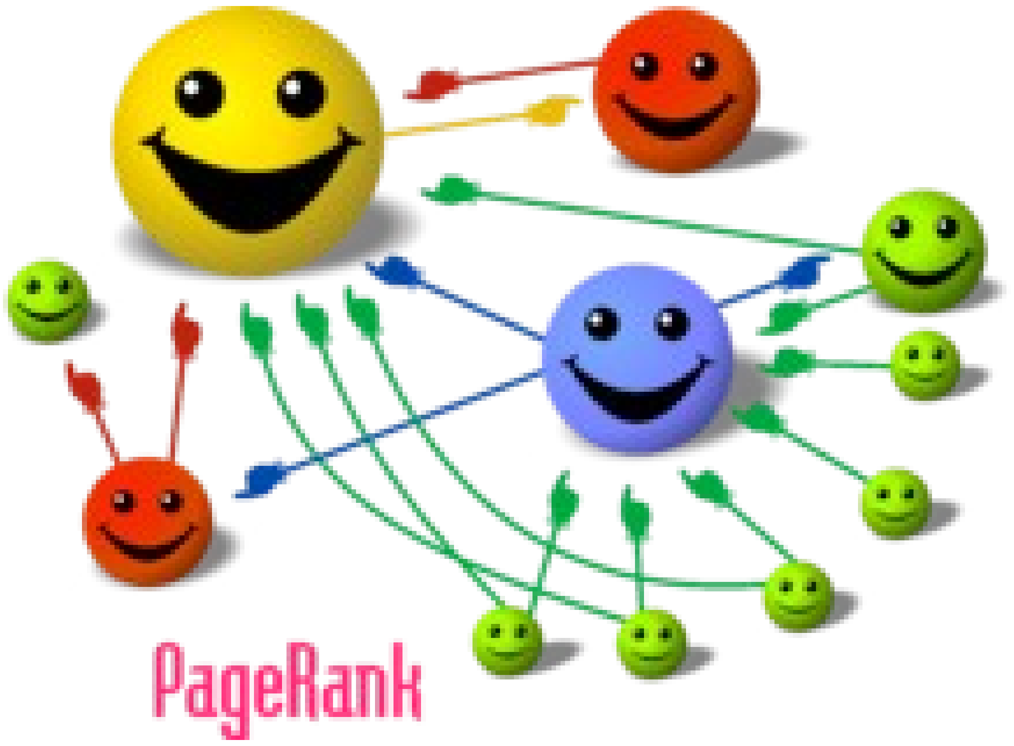
Node: **Eigenvalue** (PageRank)

- Eigenvalue measures how ***close*** an entity is to other highly ***close*** entities within a network



PageRank

- Web**Page**: ranking pages
- Larry **Page**:author of the paper



The PageRank Citation Ranking: Bringing Order to the Web

January 29, 1998

Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.

We compare PageRank to an idealized random Web surfer. We show how to efficiently compute PageRank for large numbers of pages. And, we show how to apply PageRank to search and to user navigation.

1 Introduction and Motivation

The World Wide Web creates many new challenges for information retrieval. It is very large and heterogeneous. Current estimates are that there are over 150 million web pages with a doubling life of less than one year. More importantly, the web pages are extremely diverse, ranging from "What is Joe having for lunch today?" to journals about information retrieval. In addition to these major challenges, search engines on the Web must also contend with inexperienced users and pages engineered to manipulate search engine ranking functions.

However, unlike "flat" document collections, the World Wide Web is hypertext and provides considerable auxiliary information on top of the text of the web pages, such as link structure and

Idea of PageRank

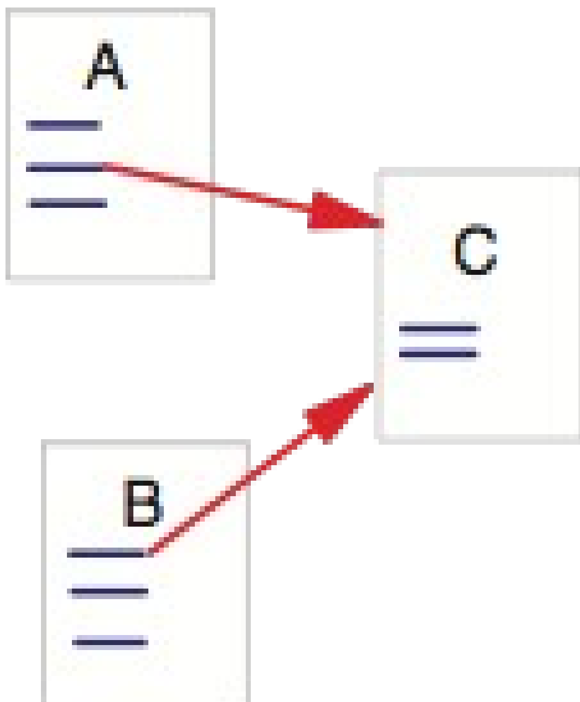


Figure 1: A and B are Backlinks of C

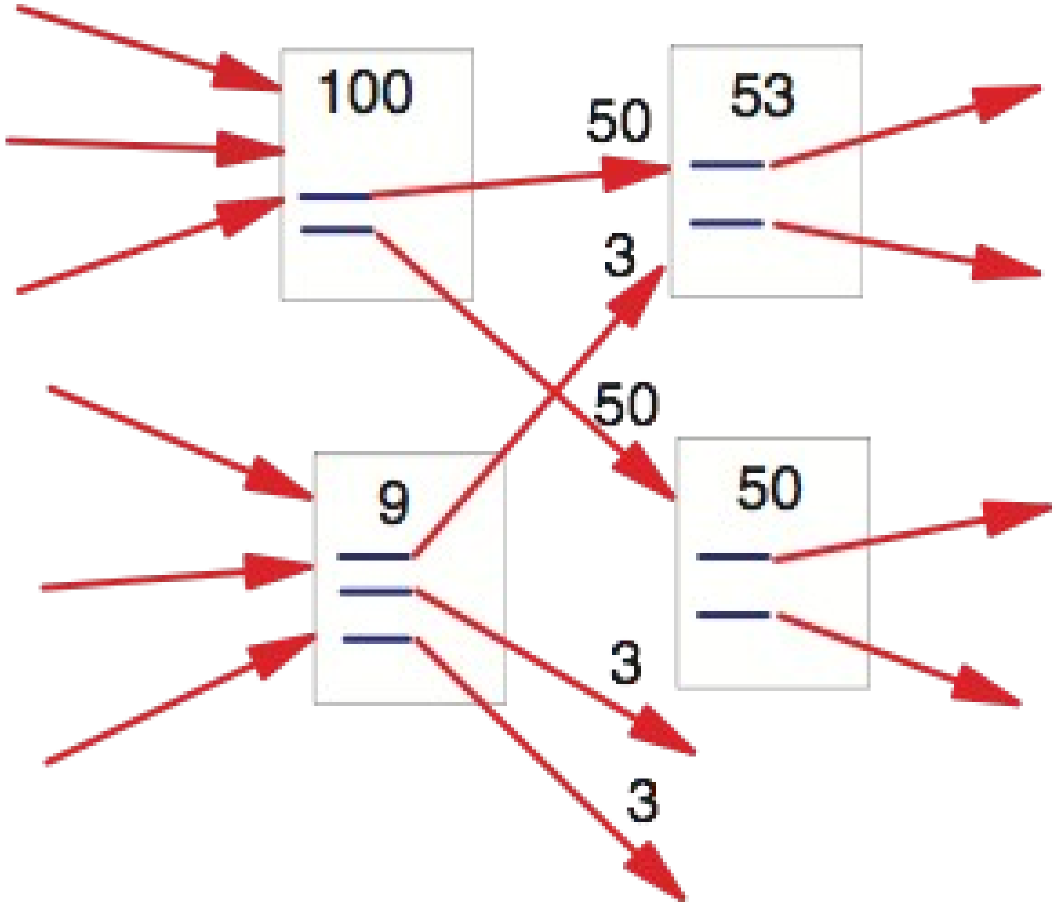
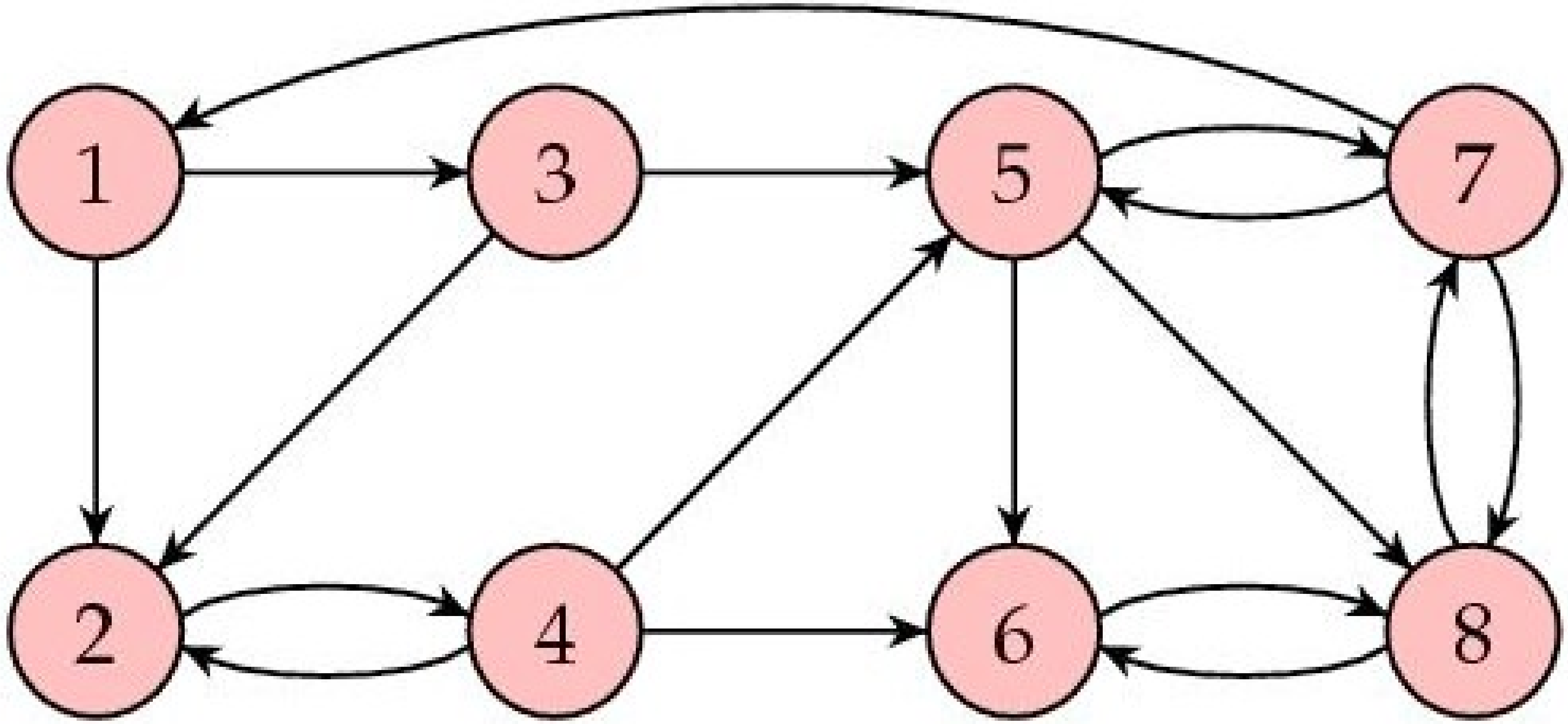


Figure 2: Simplified PageRank Calculation

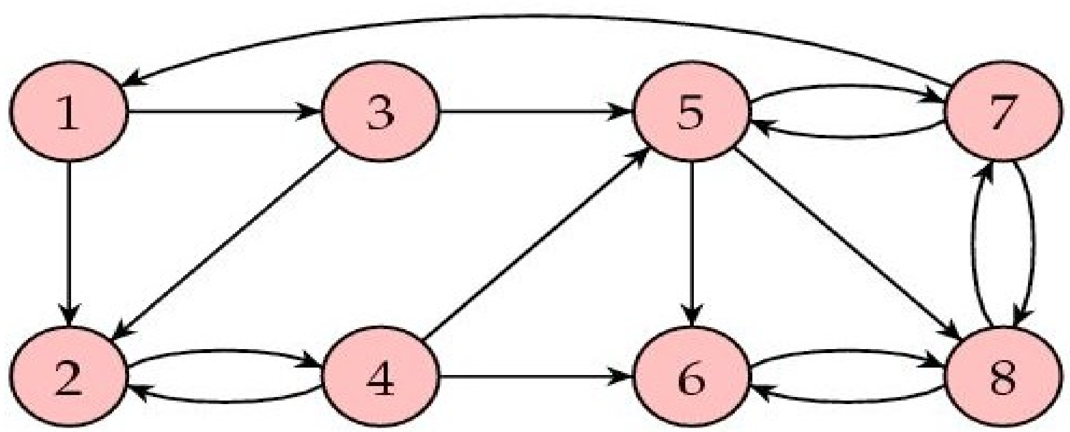


The corresponding matrix is

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

Probabilistic Interpretation of H

Random Surfing



e corresponding matrix is

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Random walk

$H(HI)$

$$HI = \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

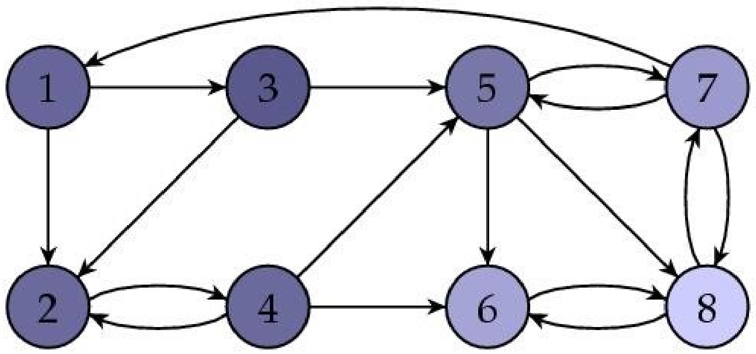
How to Compute!? Power method

$$I^{\cancel{k+1}}_{k+1} = H I^{k}_{k}$$

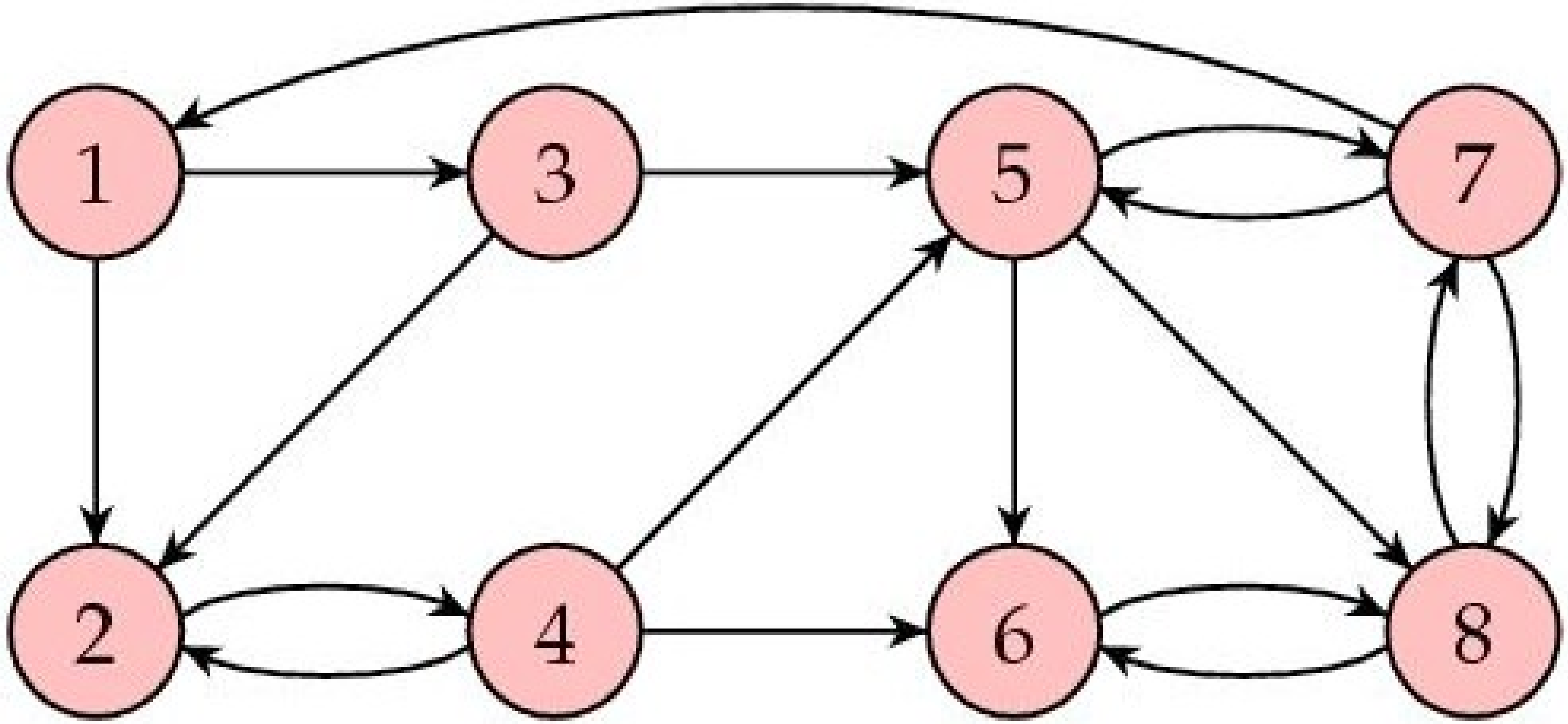
$$I_k = H^k I_0$$

How to Compute!? Power method

$$I^{k+1} = HI^k$$



I^0	I^1	I^2	I^3	I^4	...	I^{60}	I^{61}
1	0	0	0	0.0278	...	0.06	0.06
0	0.5	0.25	0.1667	0.0833	...	0.0675	0.0675
0	0.5	0	0	0	...	0.03	0.03
0	0	0.5	0.25	0.1667	...	0.0675	0.0675
0	0	0.25	0.1667	0.1111	...	0.0975	0.0975
0	0	0	0.25	0.1806	...	0.2025	0.2025
0	0	0	0.0833	0.0972	...	0.18	0.18
0	0	0	0.0833	0.3333	...	0.295	0.295



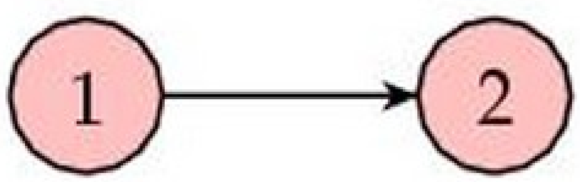
The corresponding matrix is

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix} \quad \text{with stationary vector} \quad I = \begin{bmatrix} 0.0600 \\ 0.0675 \\ 0.0300 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.1800 \\ 0.2950 \end{bmatrix}$$

Does it always converge!?

Problem with this solution

- Does the sequence I^k always converge?
- Is the vector to which it converges independent of the initial vector I^0?
- Do the importance rankings contain the information that we want?

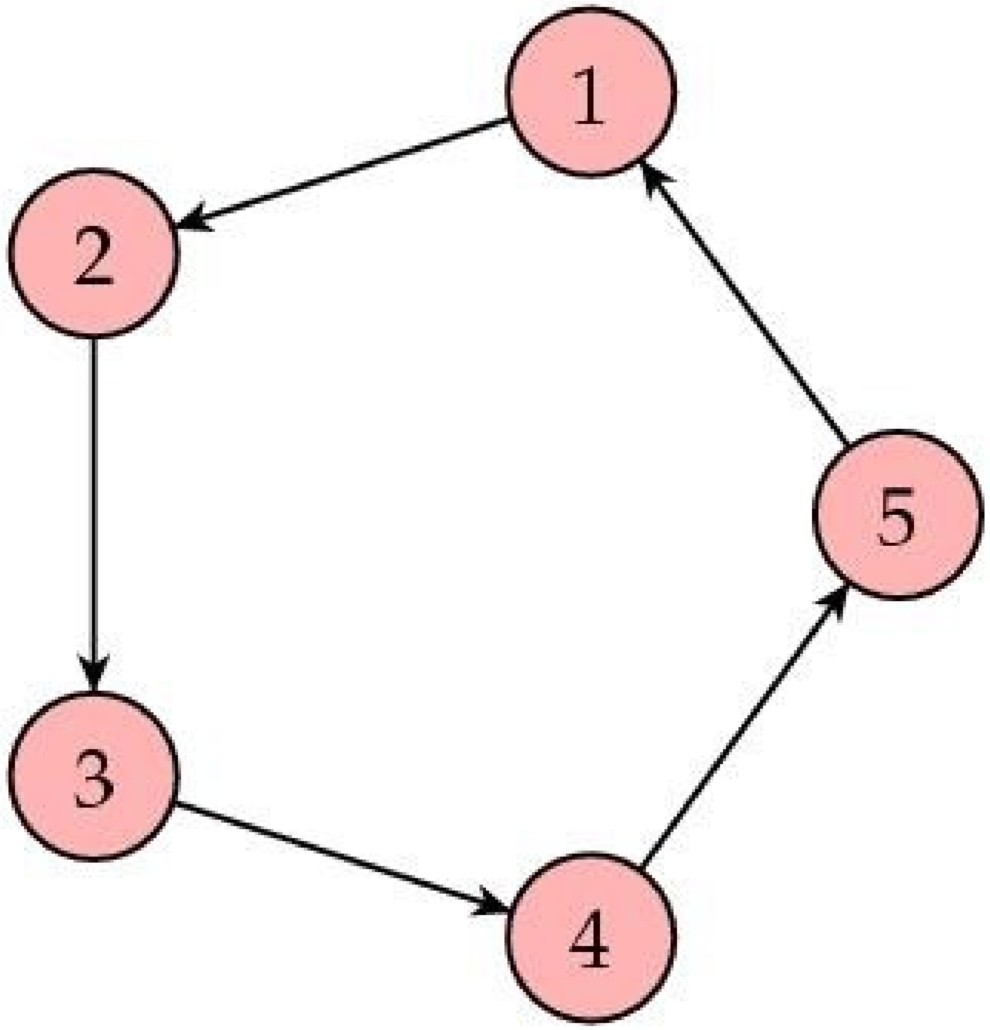


with matrix $$H = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Here is one way in which our algorithm could proceed:

I^0	I^1	I^2	$I^3=I$
1	0	0	0
0	1	0	0

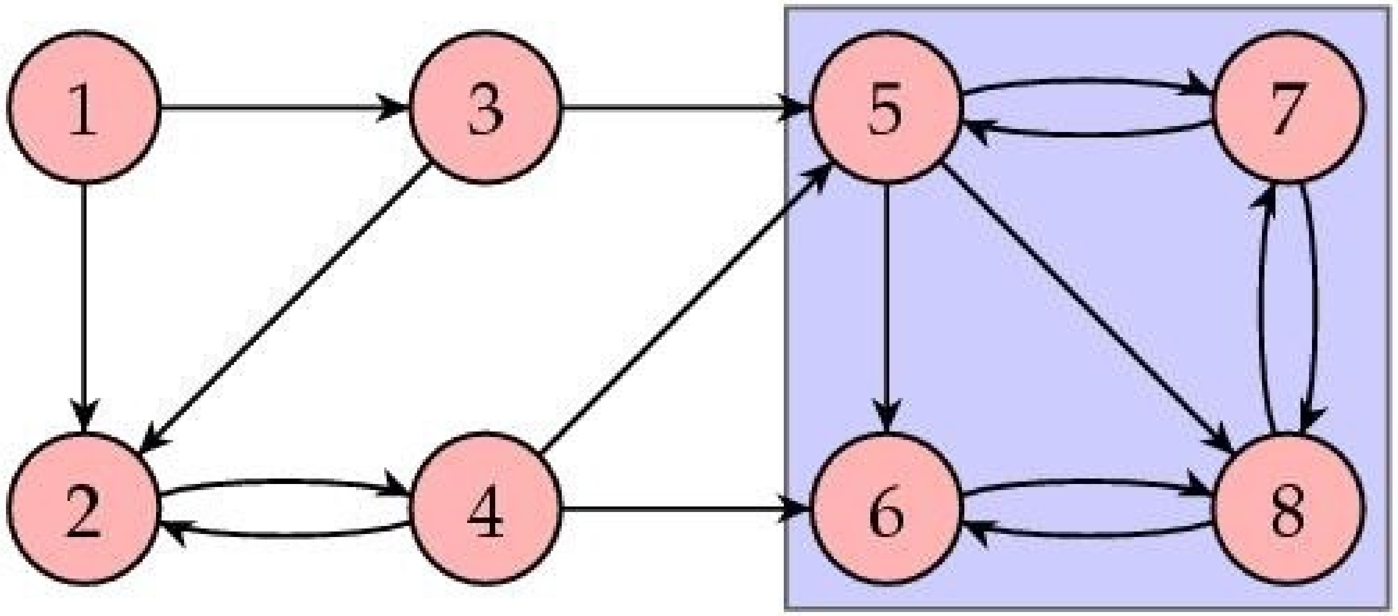
Problem 2



$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

I^0	I^1	I^2	I^3	I^4	I^5
1	0	0	0	0	1
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0

Problem 3: No Outlinks



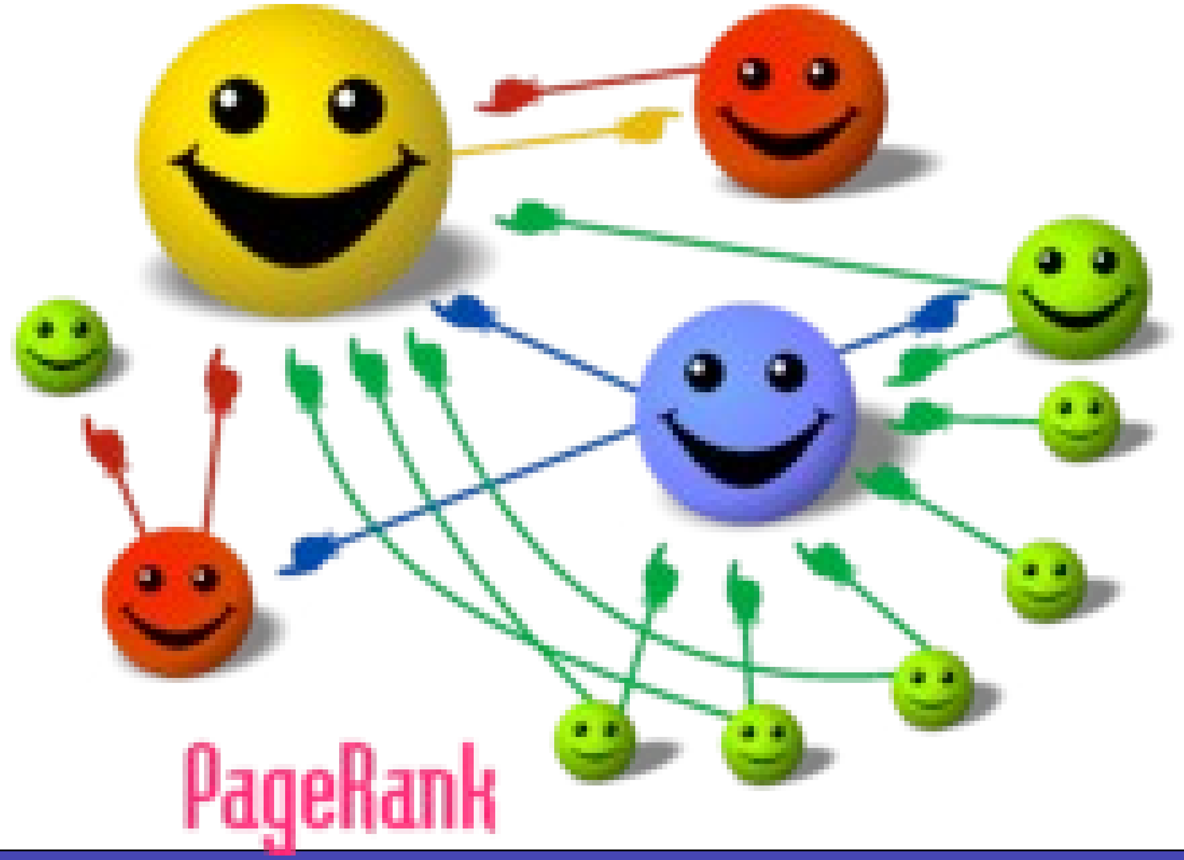
$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/2 & 0 \end{bmatrix}$$

Better Solution

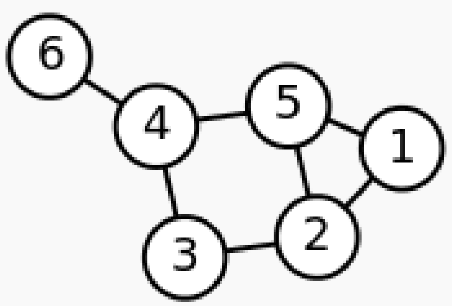
Teleportation!

$$G = \alpha H + (1-\alpha)O \qquad O = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

Adding (very unlikely) outlinks for every node



Graph Laplacian

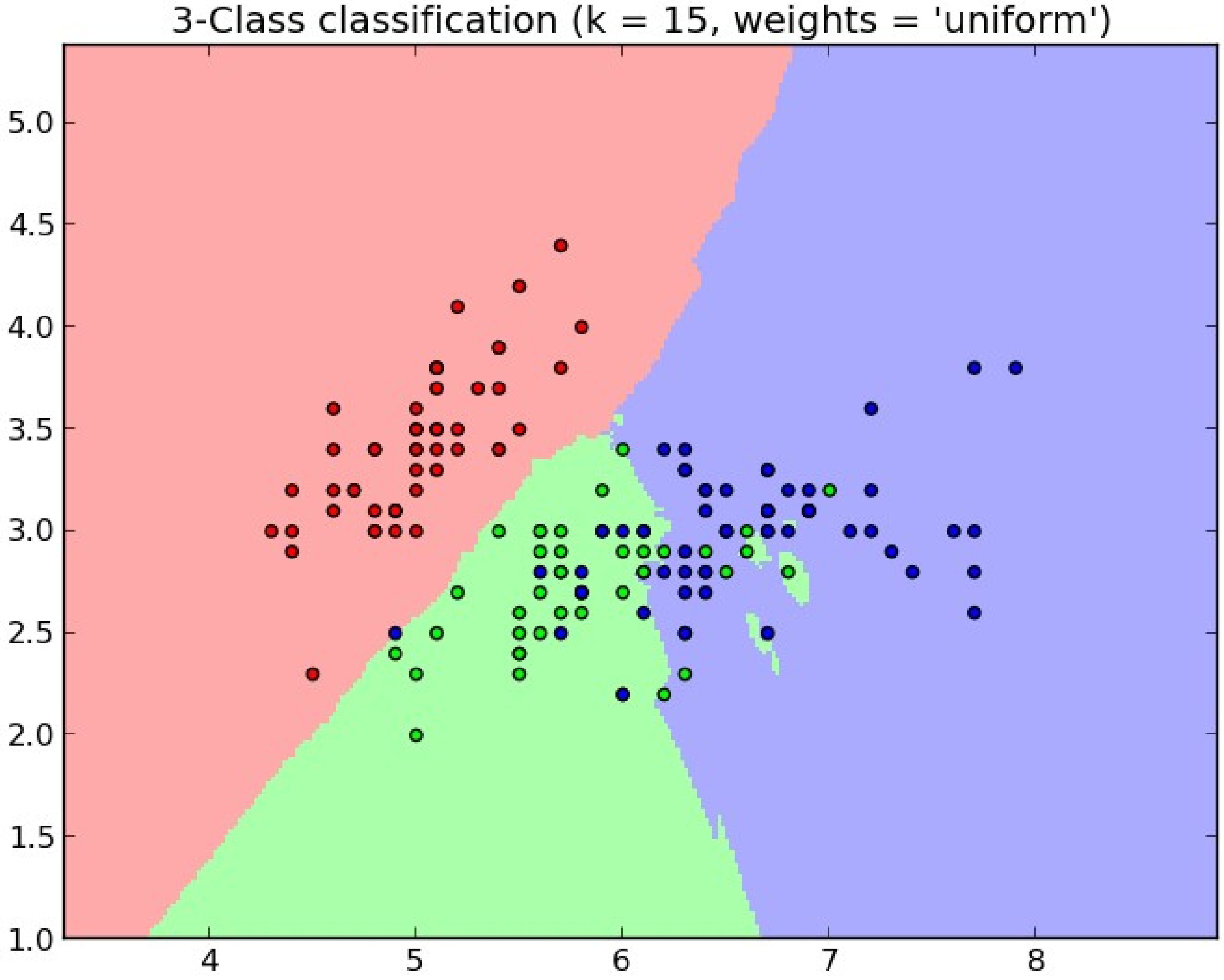
Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

https://www.youtube.com/watch?v=yDltiyH1Hxw

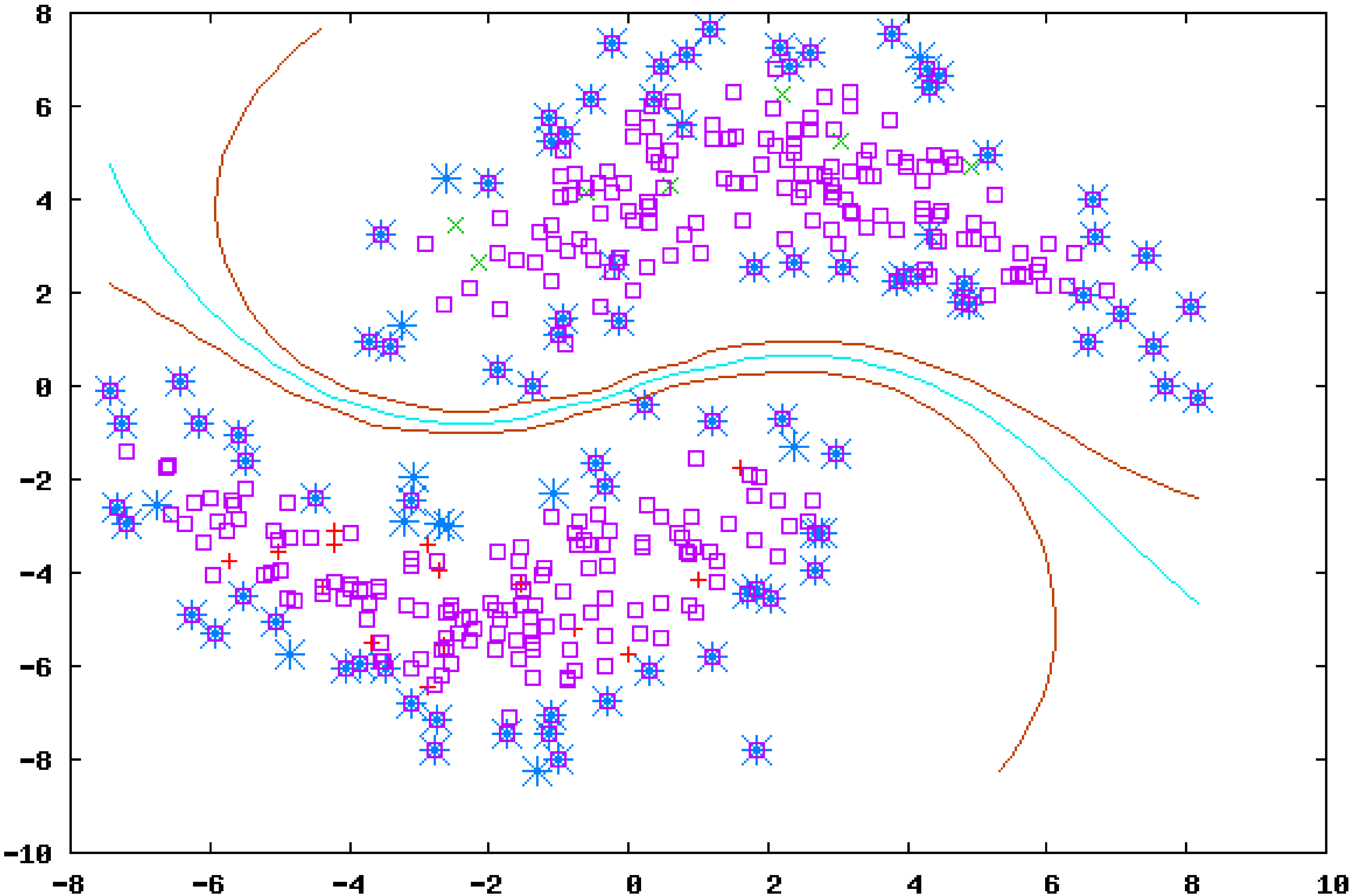
10:45

High Dimensional Data

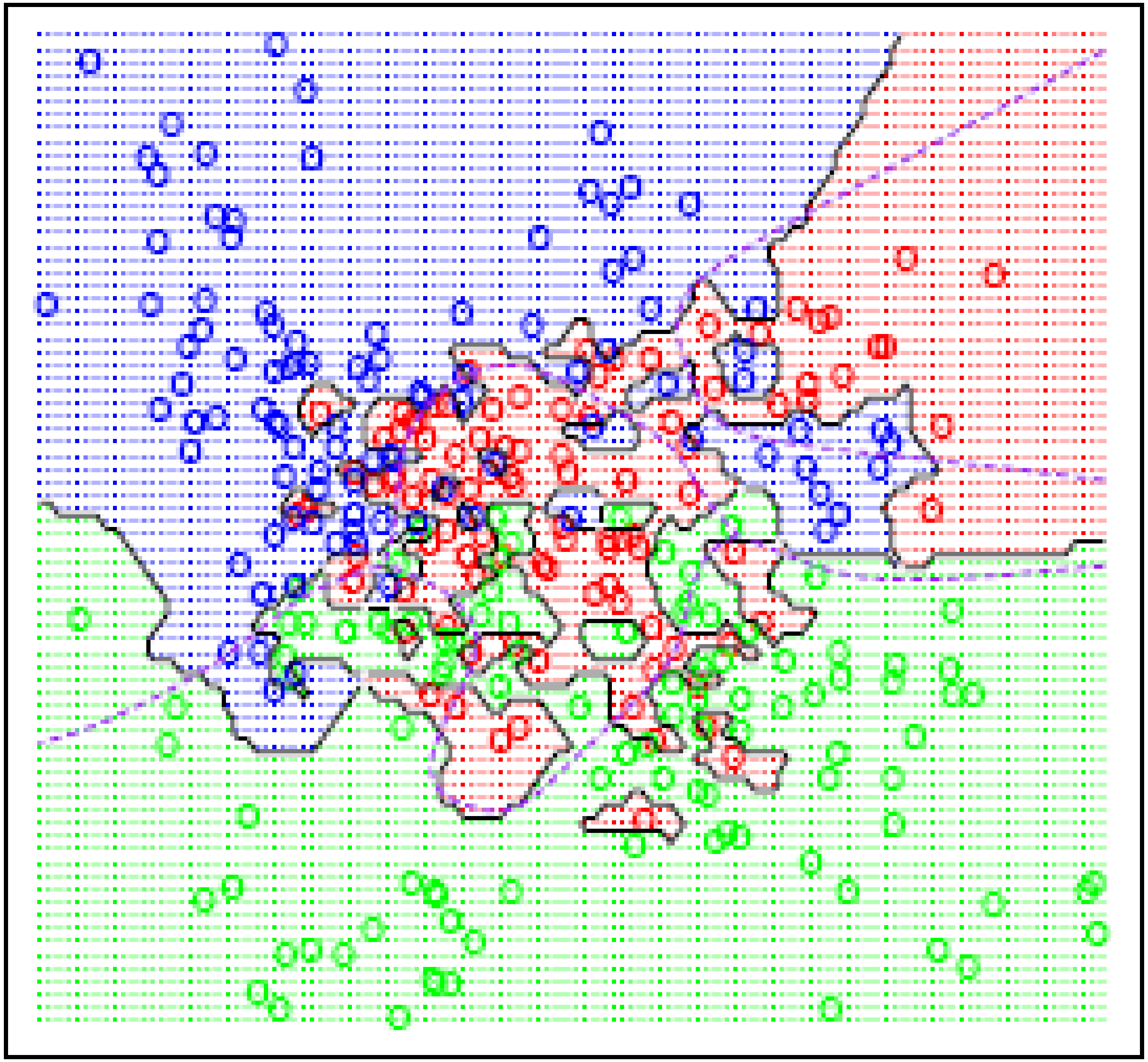
“Perfect World” for Data Analysis



“Perfect World” for Data Analysis



“Perfect World” for Data Analysis



Real World Data Analysis

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

High Dimension?

P ⇐ Big

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

N

⇑

Big

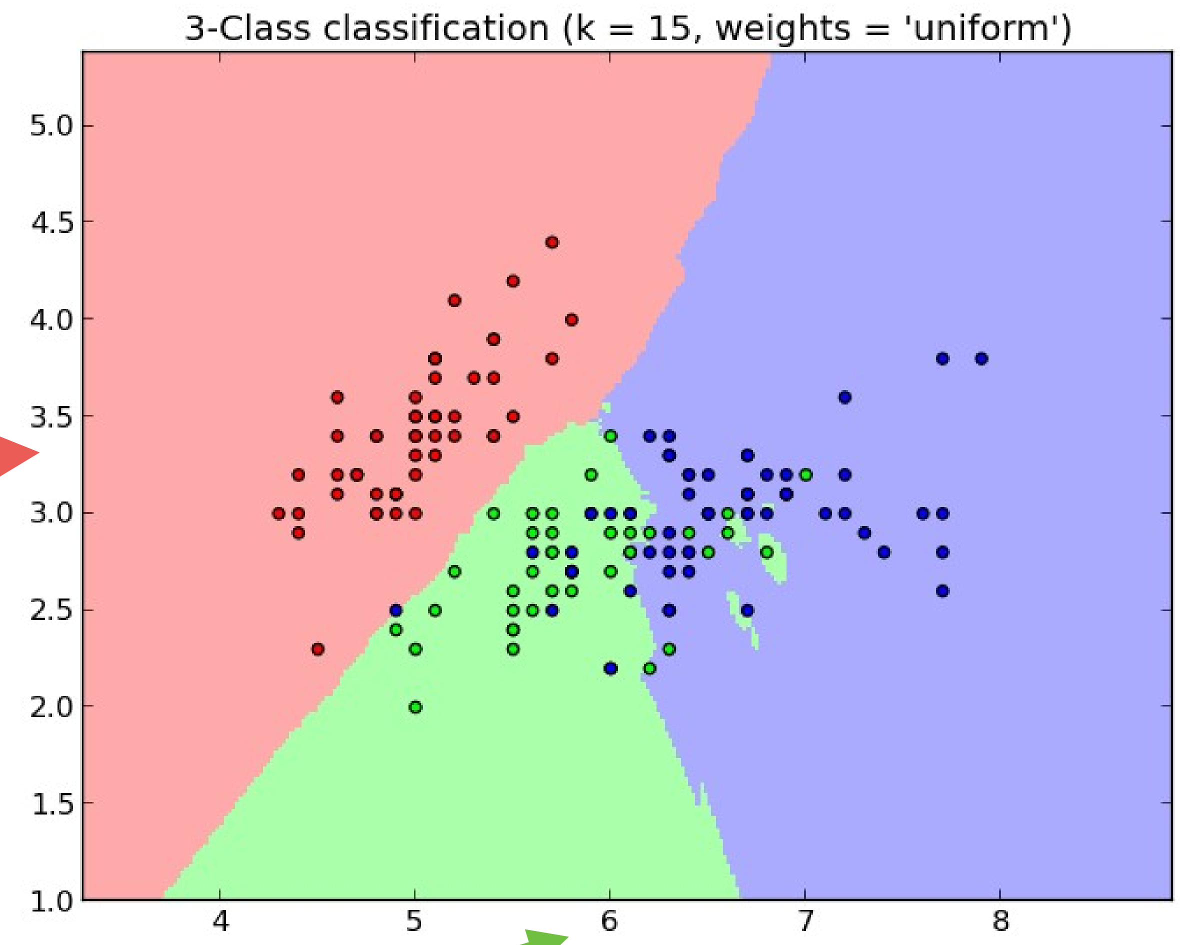
Issue is not p "big"!

Issue is "p is big relative to n"

For a given p How much n do you need?

“Perfect World” for Data Analysis

Volume (cm3)	Pressure (bar)
3.701502	-0.366814
7.858678	0.132824
11.6975	0.907633
15.46395	1.790605
19.18672	2.770453
23.00292	3.734302
26.83077	4.753389
30.75468	5.746156
34.73695	6.742996

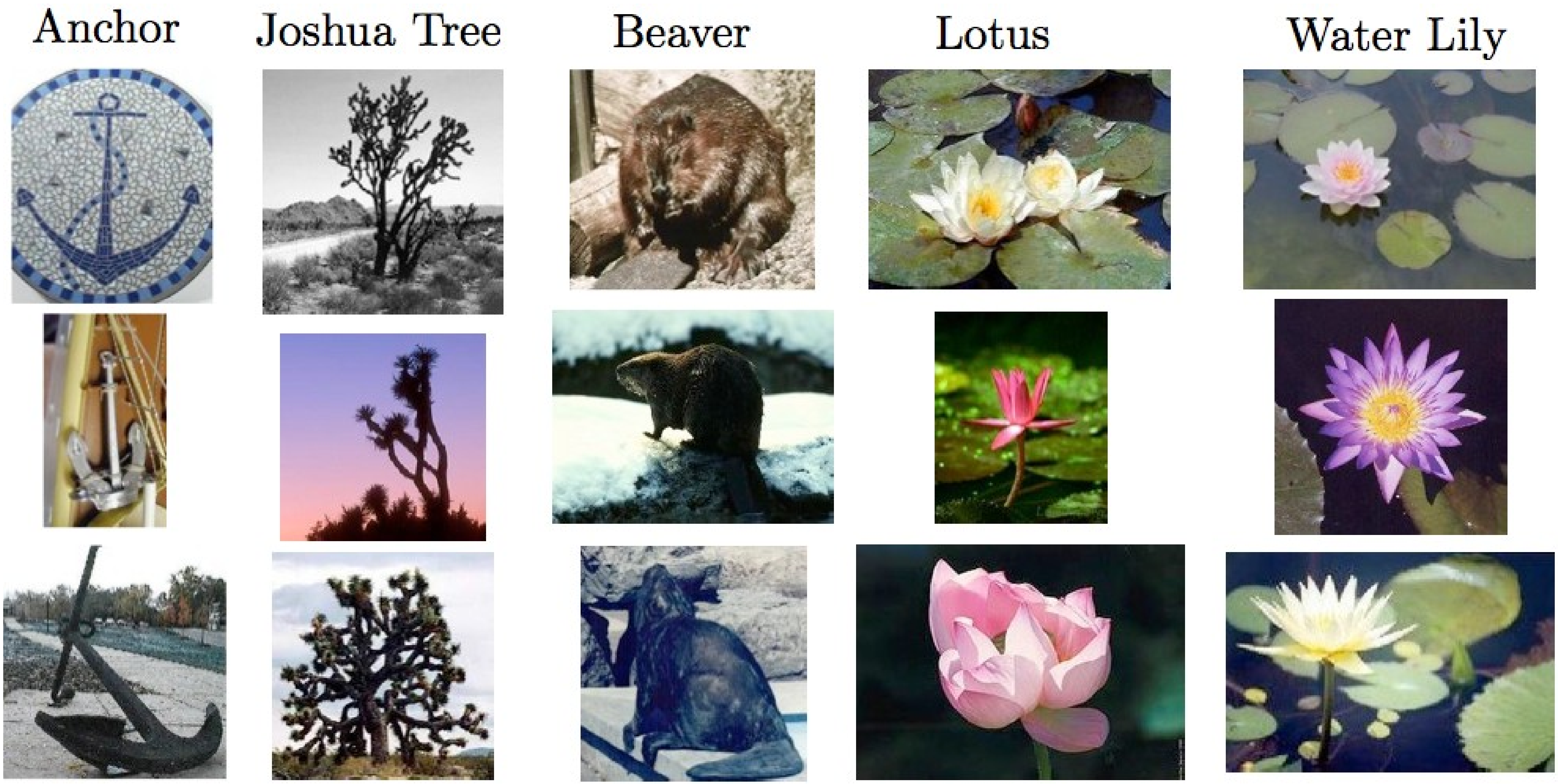


Less than Perfect

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

5 dimensional

High Dimensional Data Classification



predictors are pixels!

High Dimensional Data
Other Examples?

Genes

Voice Recognition

text processing

Handwriting

Brain mapping

ekg

weather

sports

stocks

What is really happening in

High Dimensional Space?

https://www.youtube.com/watch?v=C6kn6nXMWF0

Outline

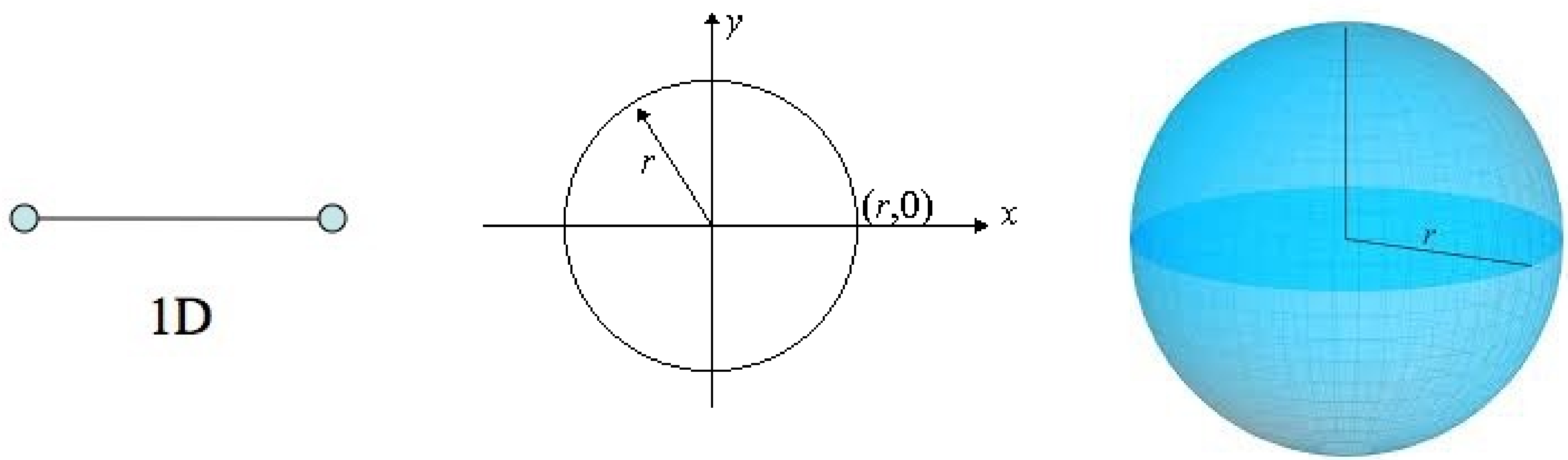
- What are spheres and cubes in high dimensions?
- What does distance mean in high dimensions?
- What does random mean in high dimensions?
- What does data mean in high dimensions?

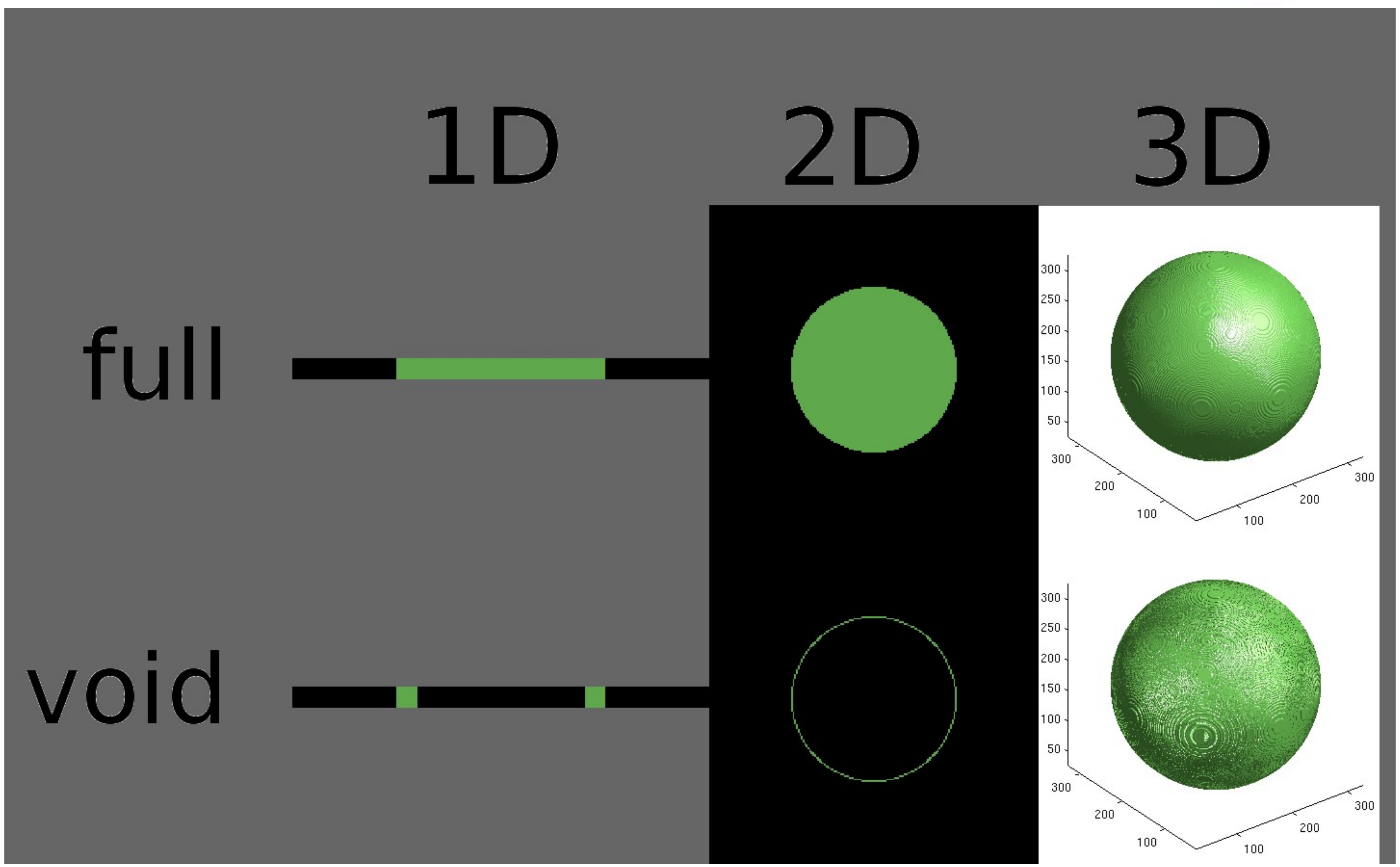
Who is this guy?

Eigenface #16

Let's start simple: What are spheres?

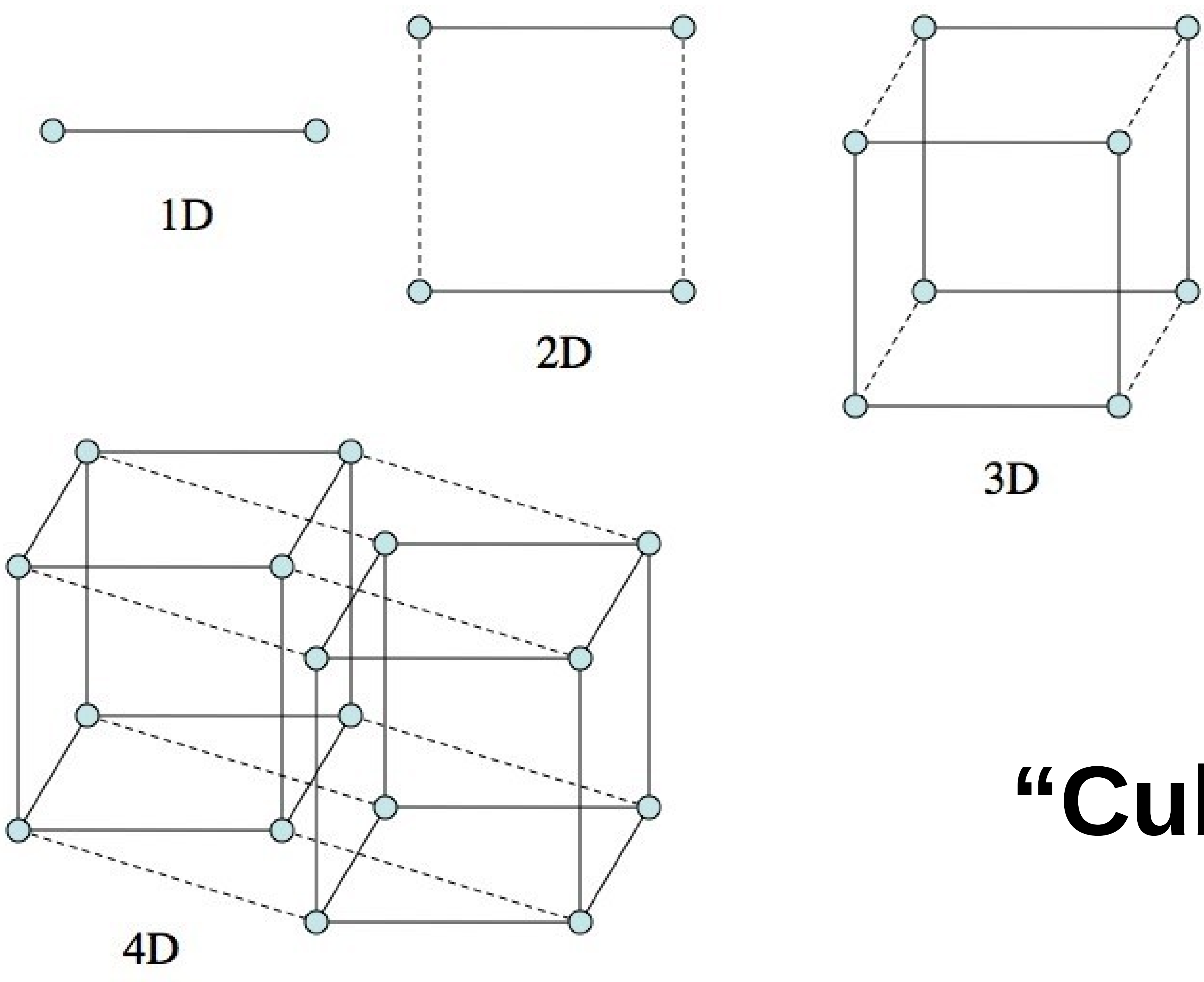
Sphere





Let's start simple: What are cubes?

Line - Square - Cube - HyperCube

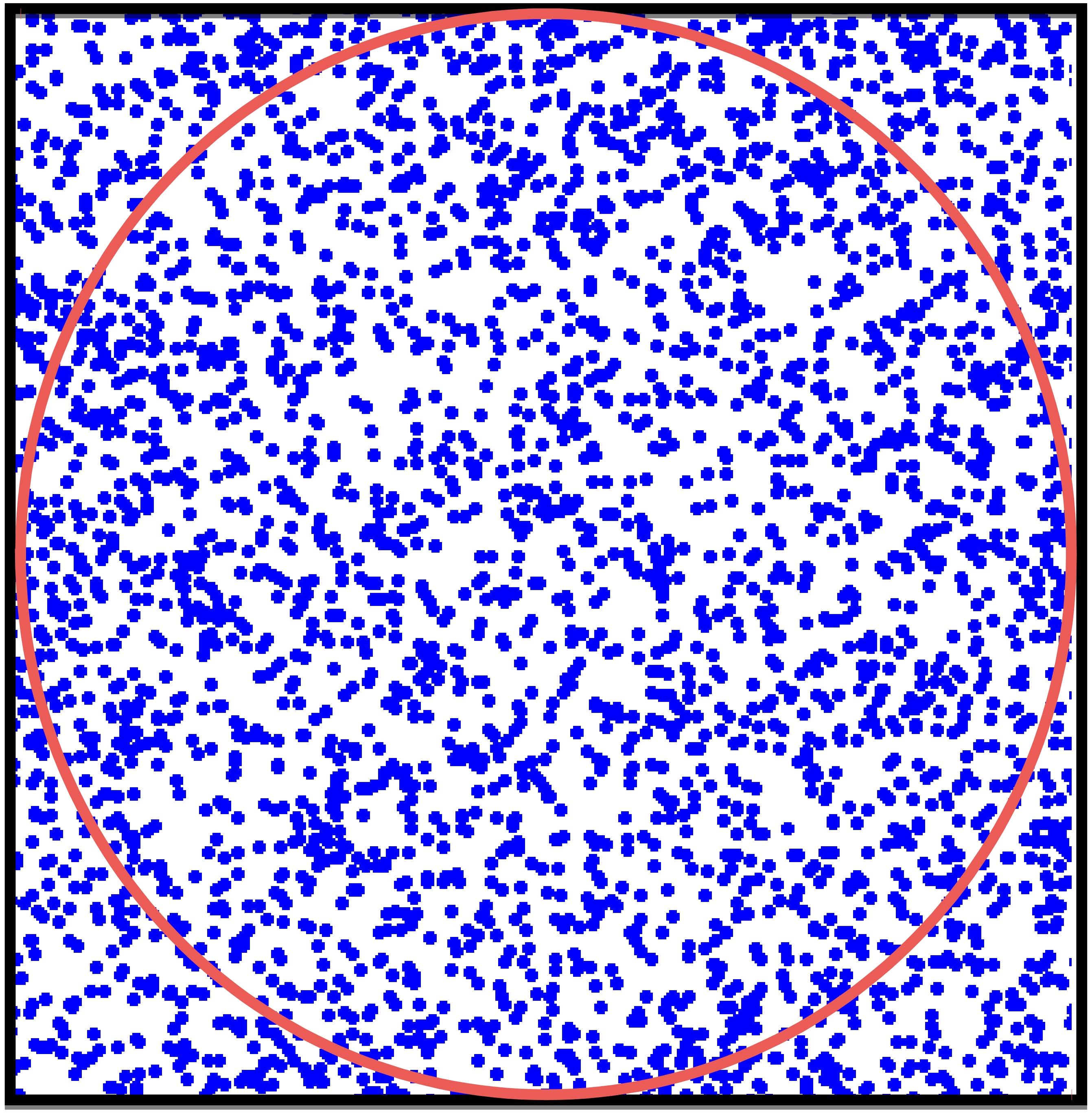


“Cube”

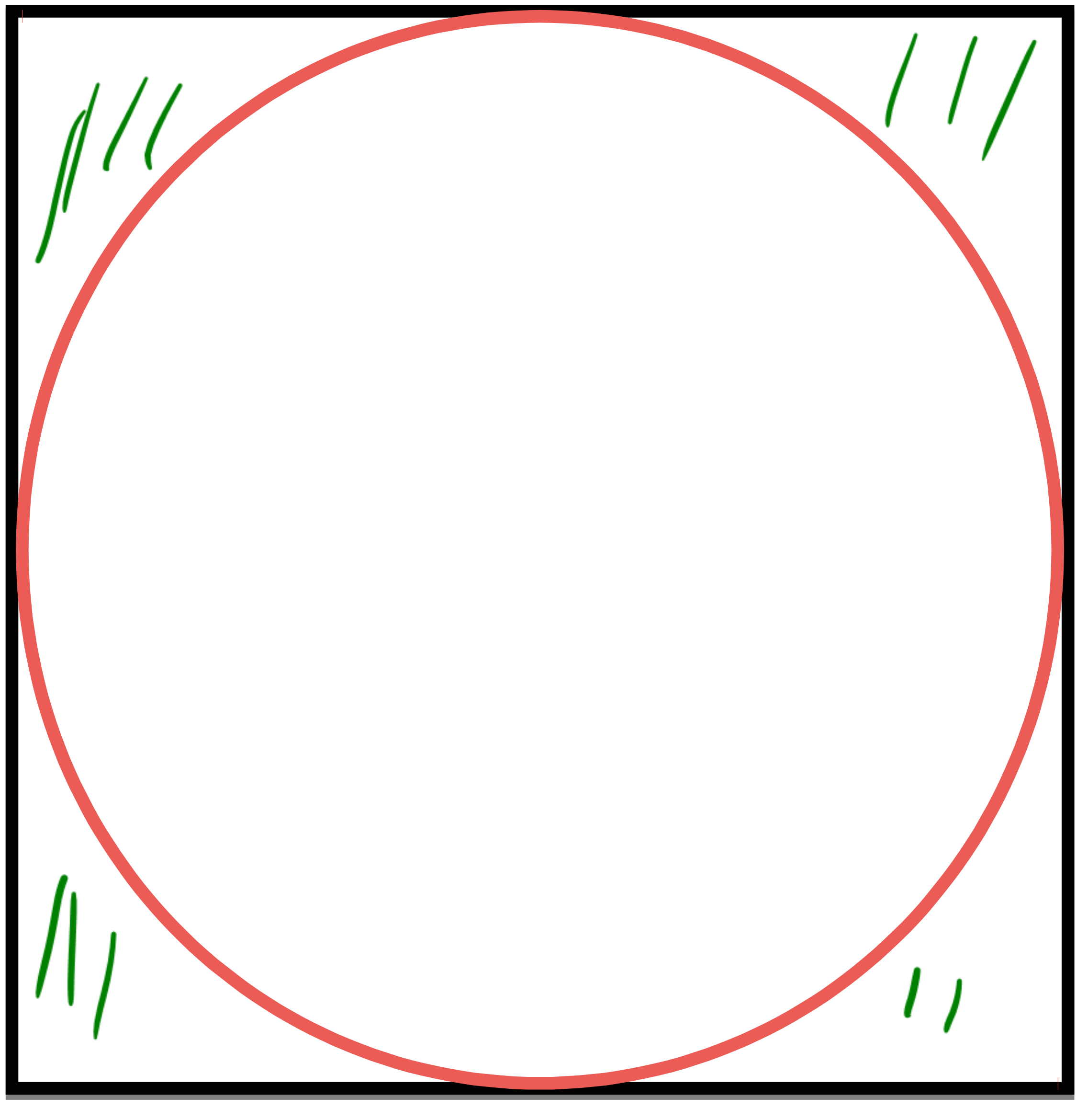
How do spheres and cubes relate in high-dimensions?

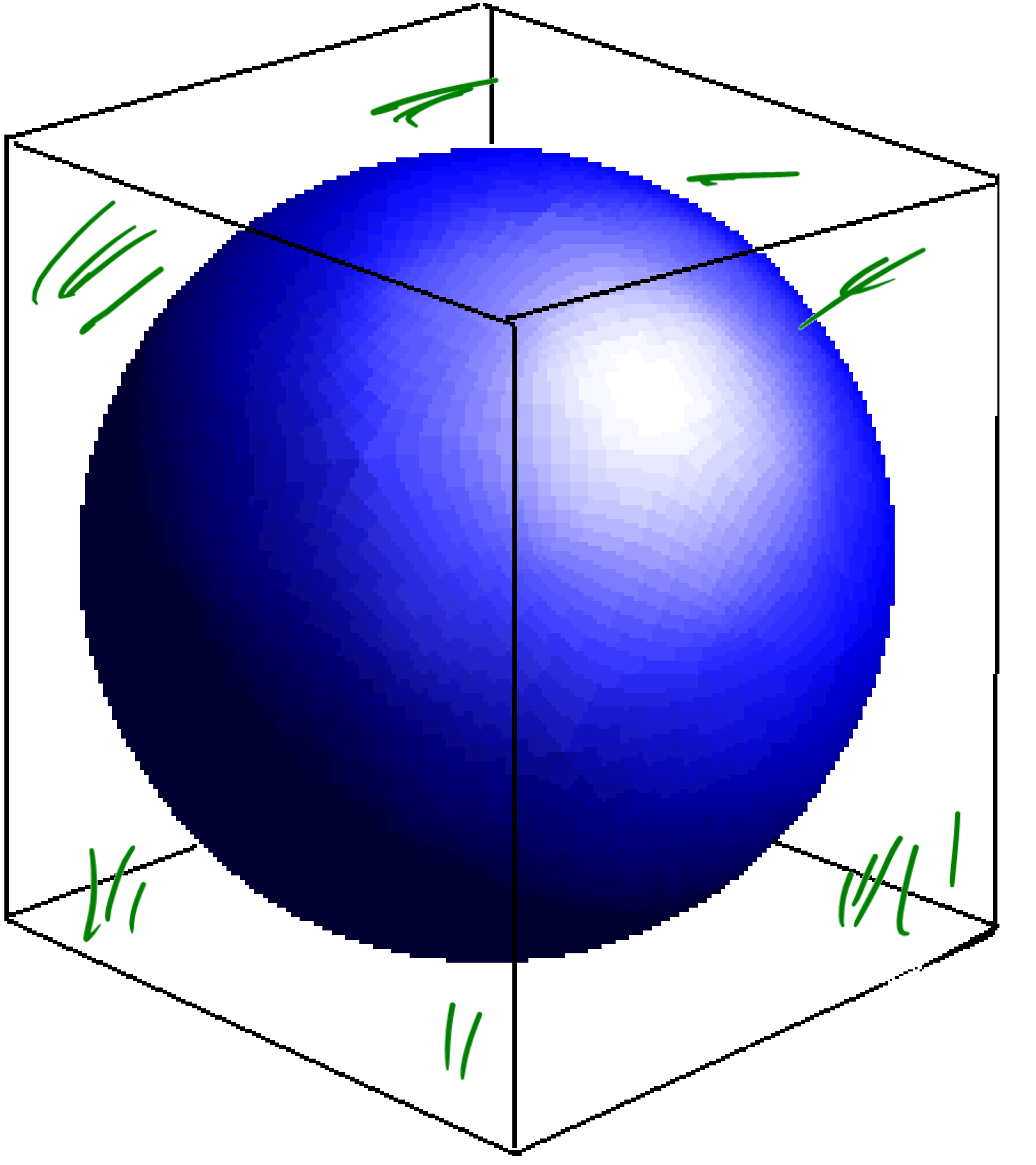
Question #1:

Uniform Distribution (Center vs Corner)

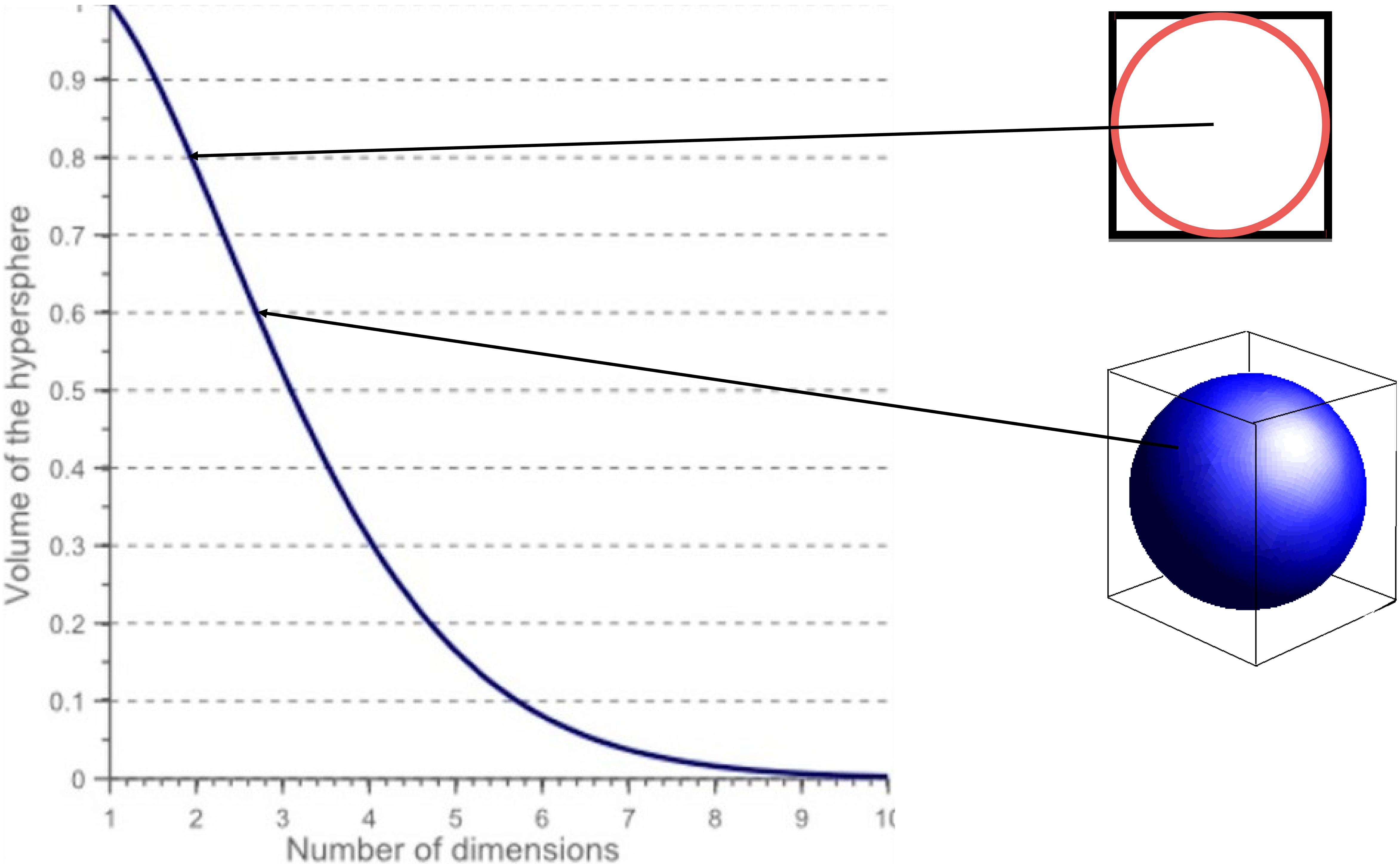


Center vs Corner





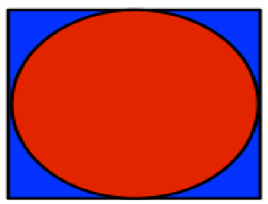
Volume of the hypersphere

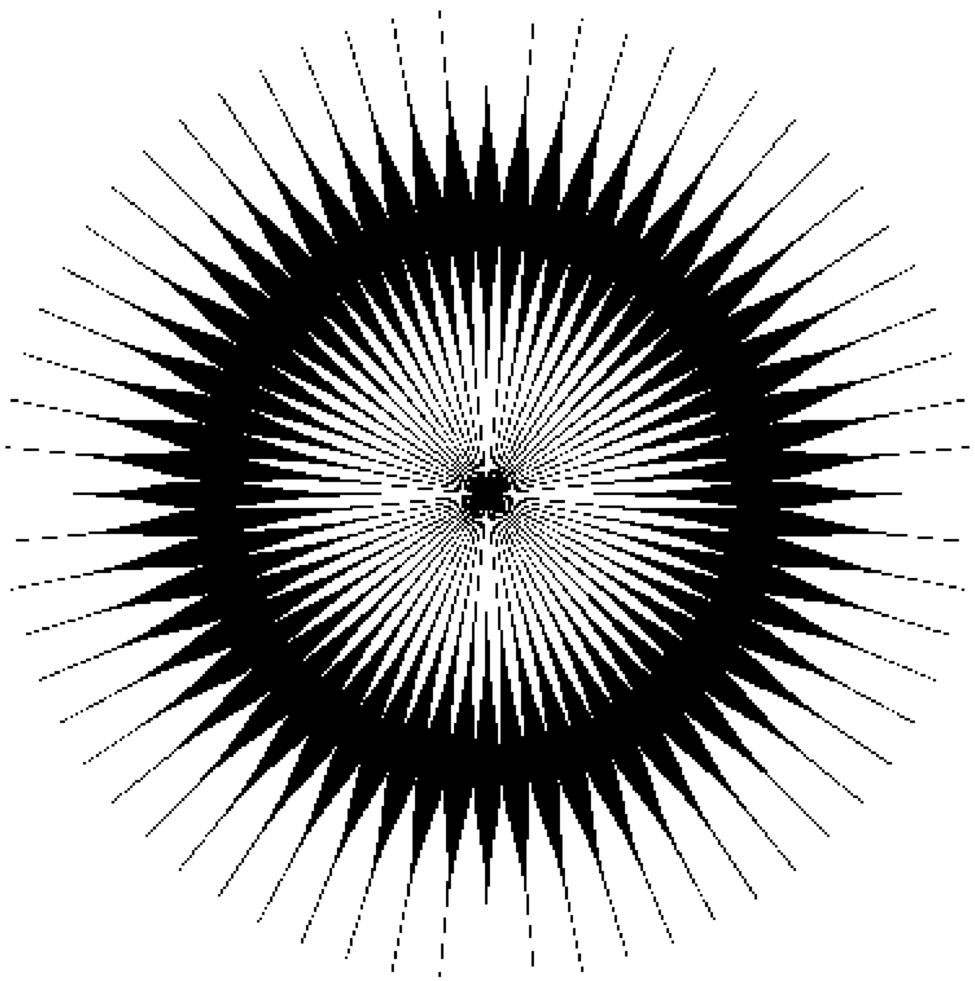


High Dimensional Cube

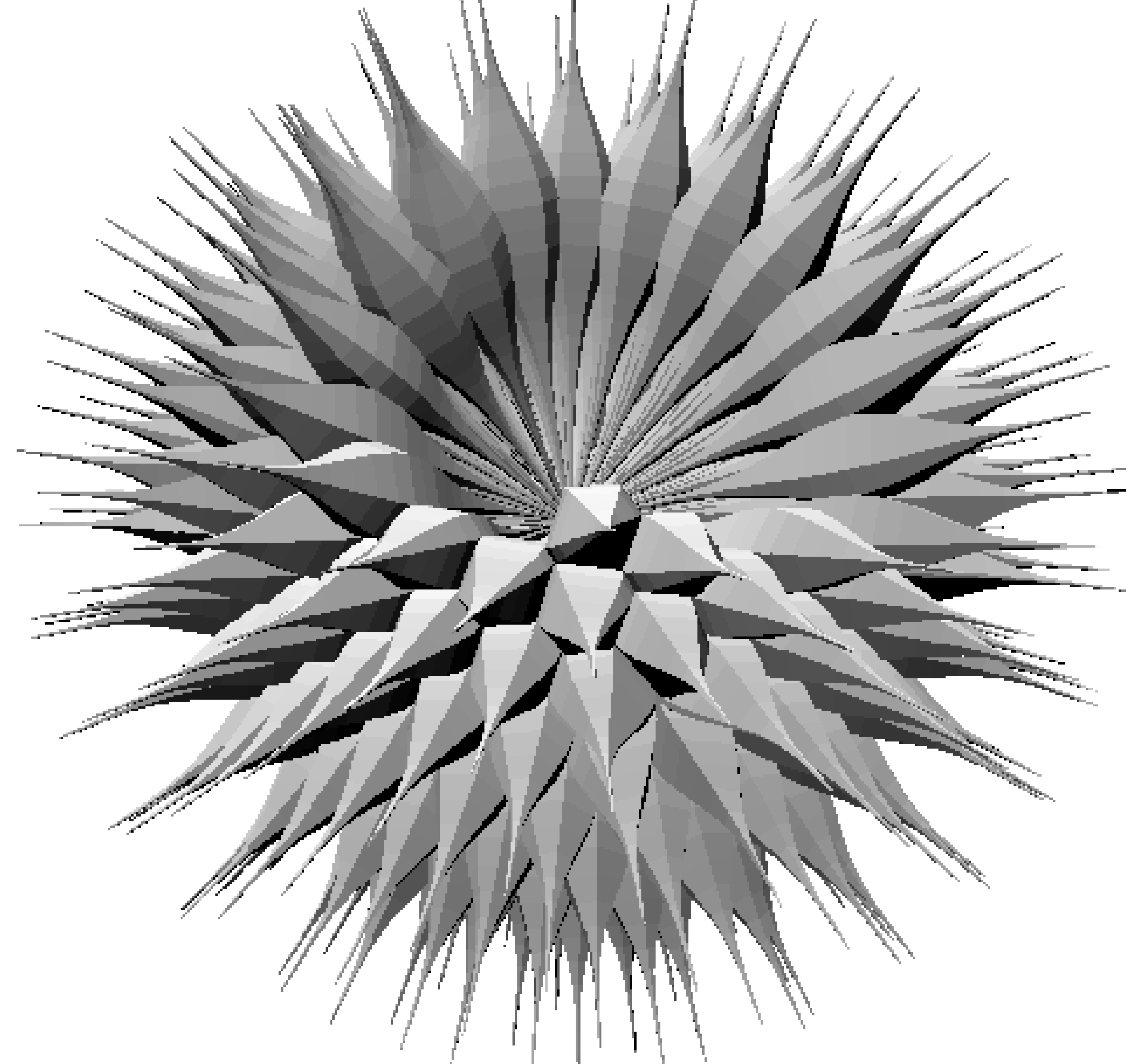
- 10^d points cover $[0,1]^d$ at a distance 10^{-1}.

$$\lim_{d\to\infty} \frac{\text{volume sphere of radius r}}{\text{volume } [0,r]^d} = 0$$

 : nearly all points are in the 2^d corners!

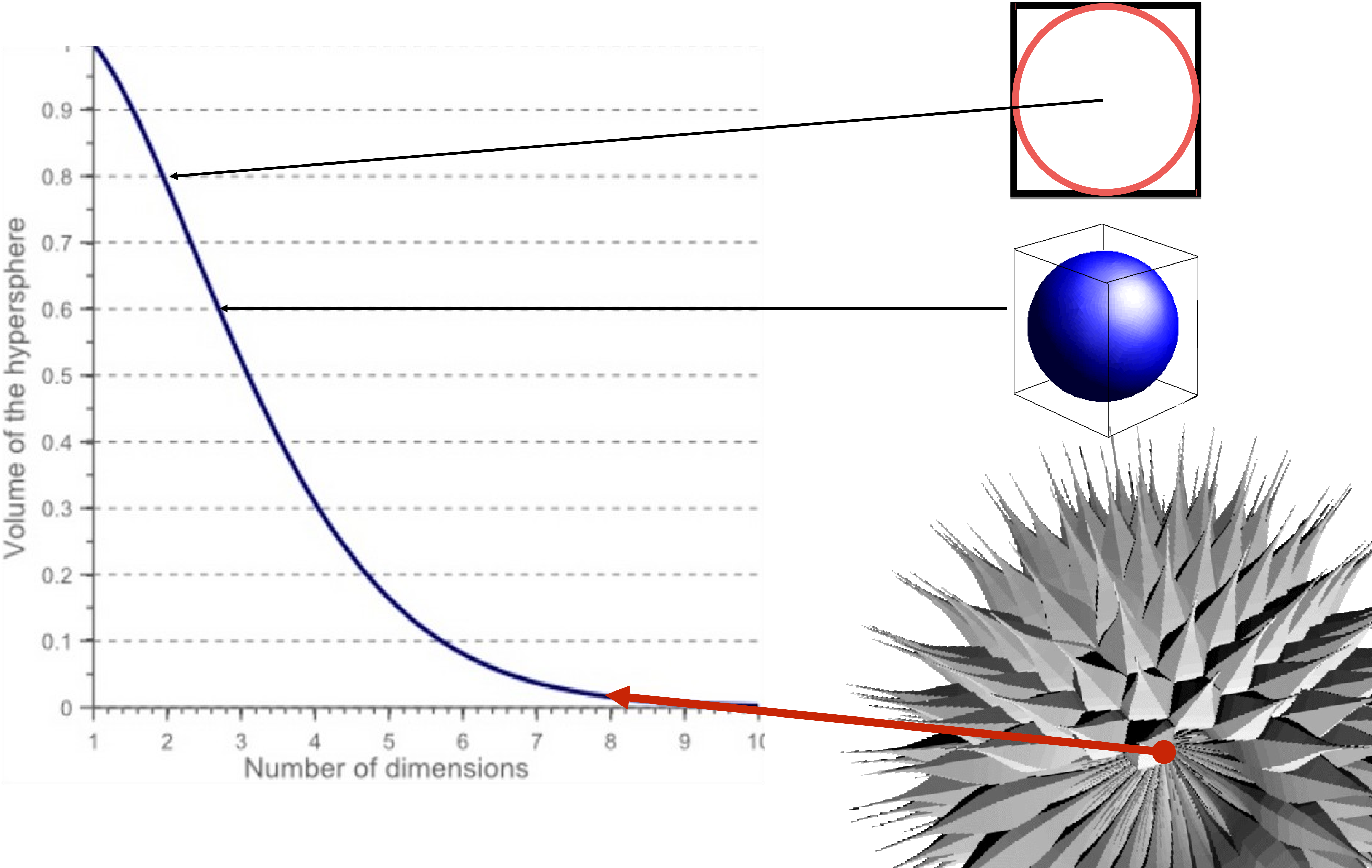


6 dimensional cube projection



8 dimensional cube projection

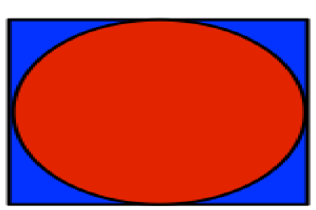
Volume of the hypersphere



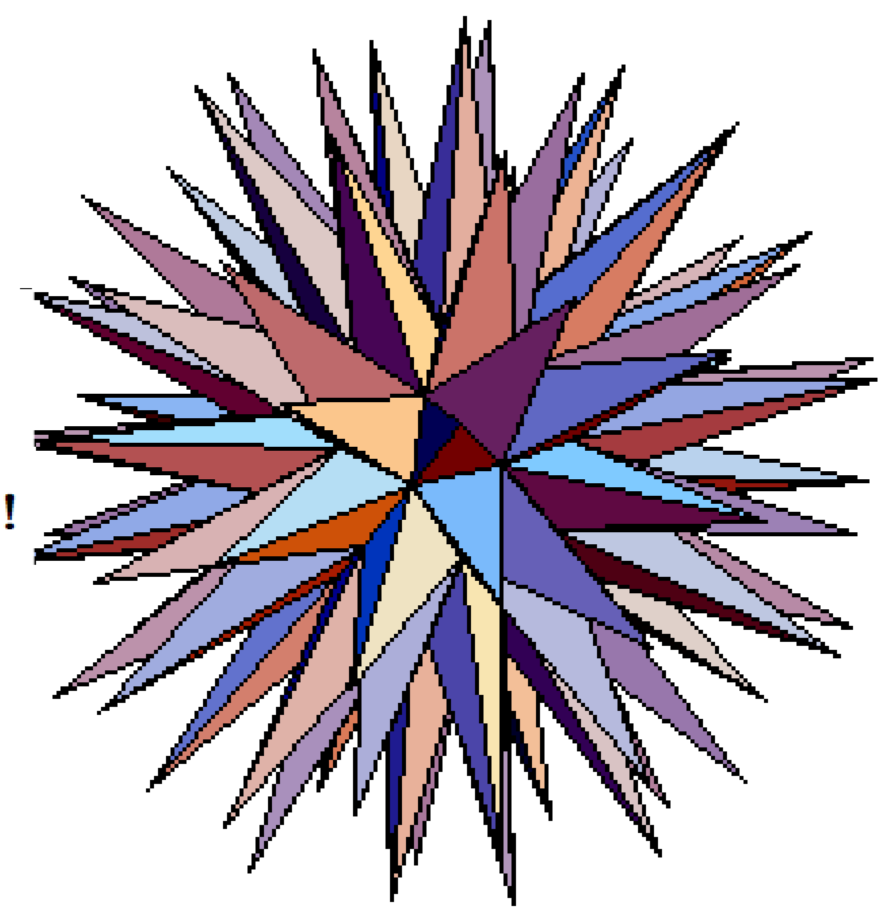
In High Dimensional Space: **Rule # 1**

- **Center v.s. Corners :**
  - **Corners win!**

$$\underset{n \to \infty}{} \frac{\text{volume sphere of radius r}}{\text{volume } [0, r]^d} = 0$$

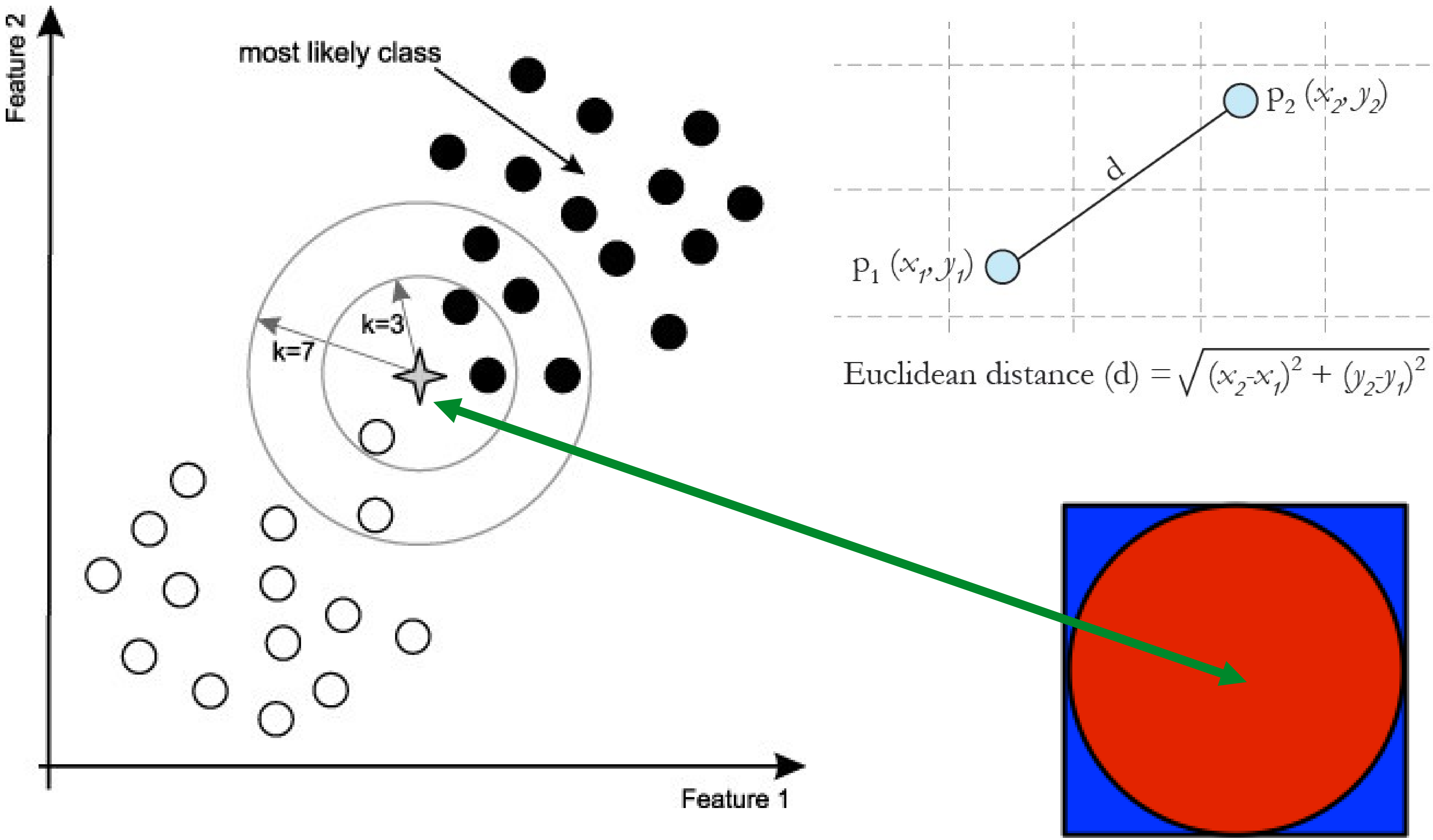


: nearly all points are in the 2^d corners!



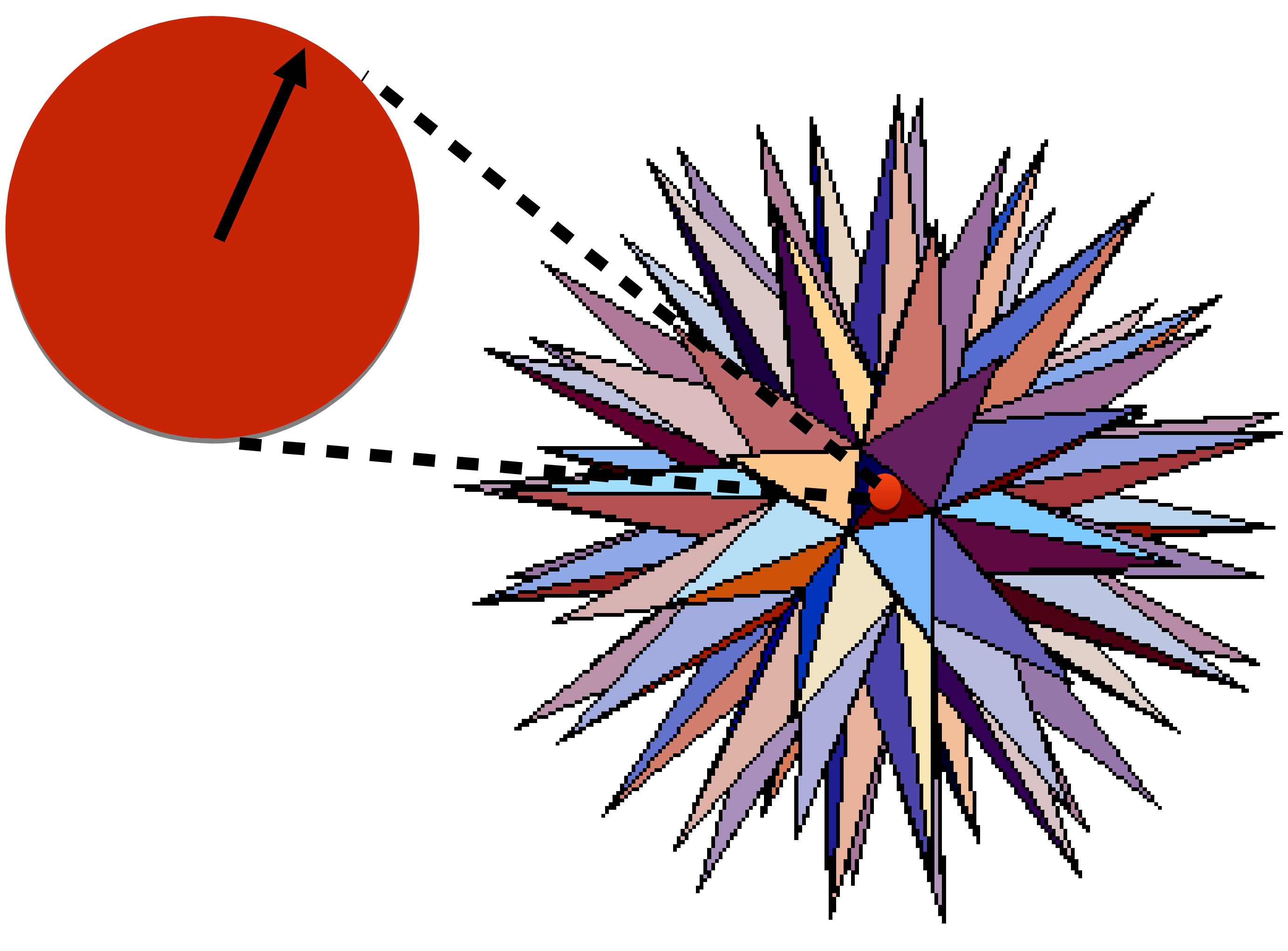
What does distance mean in high-dimensions?

Question #2: Euclidean Distance



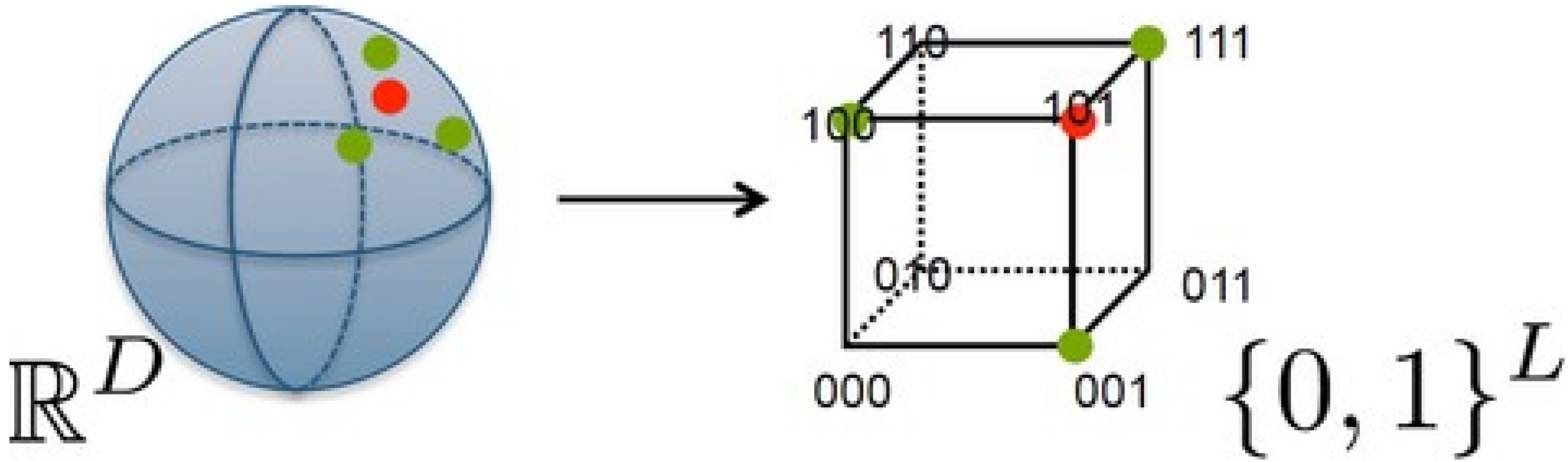
In High Dimensional Space: **Rule # 2**

- **Don’t Use Euclidean Distance**

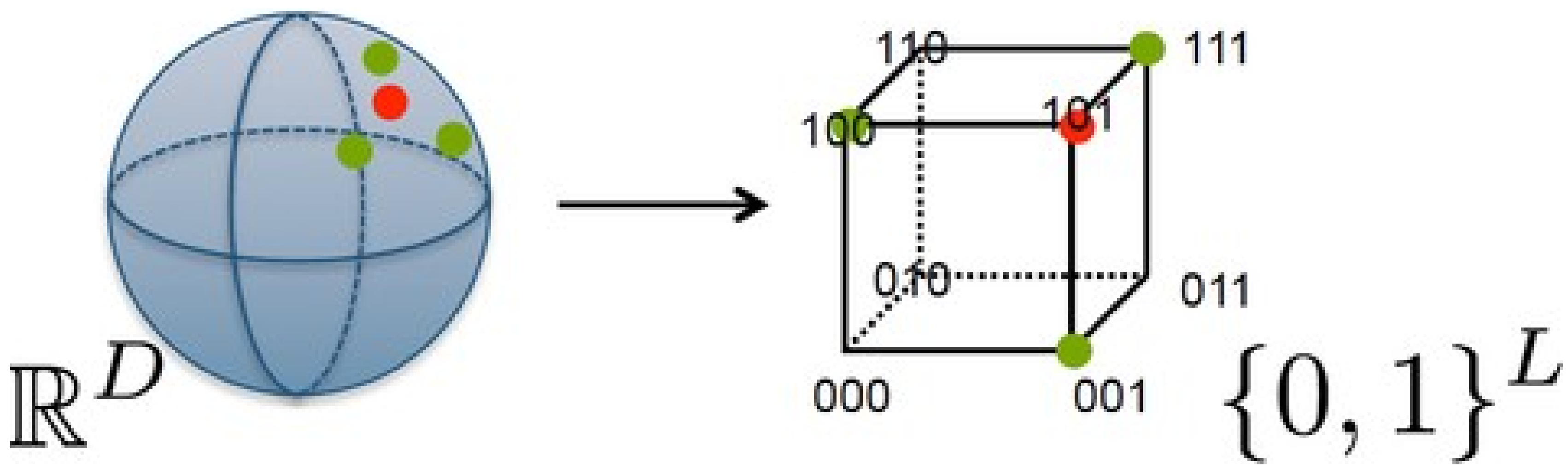


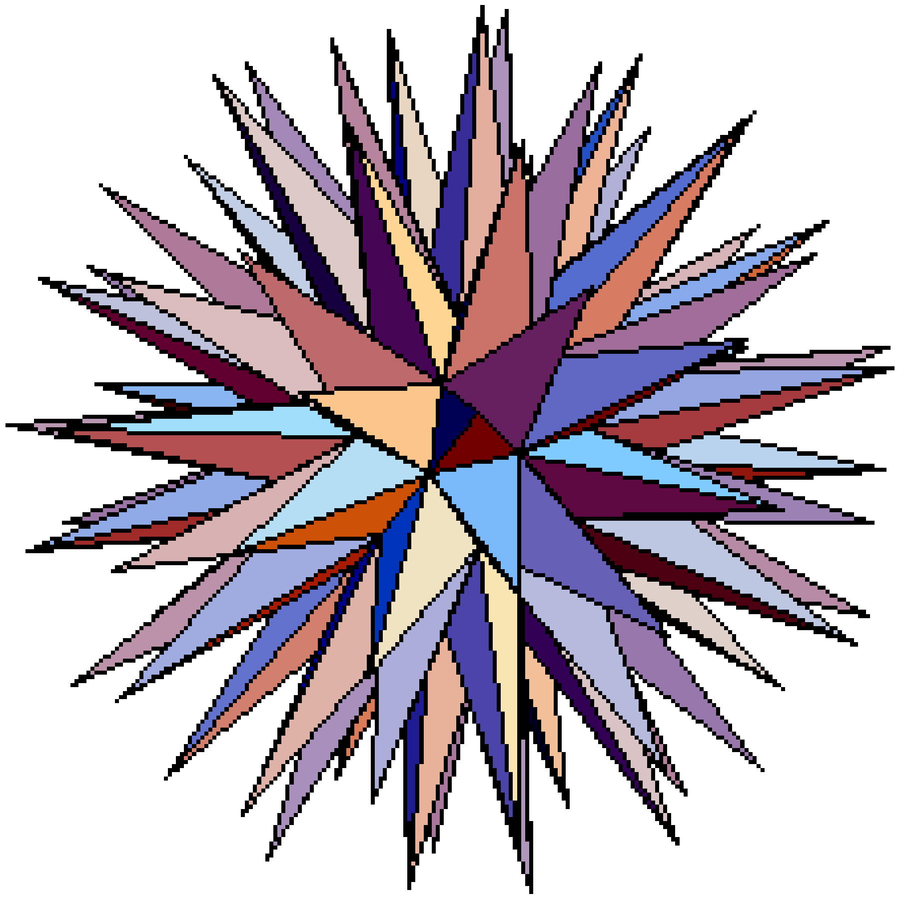
all your “neighbors” are far away

All “neighbors” are “equally” far away

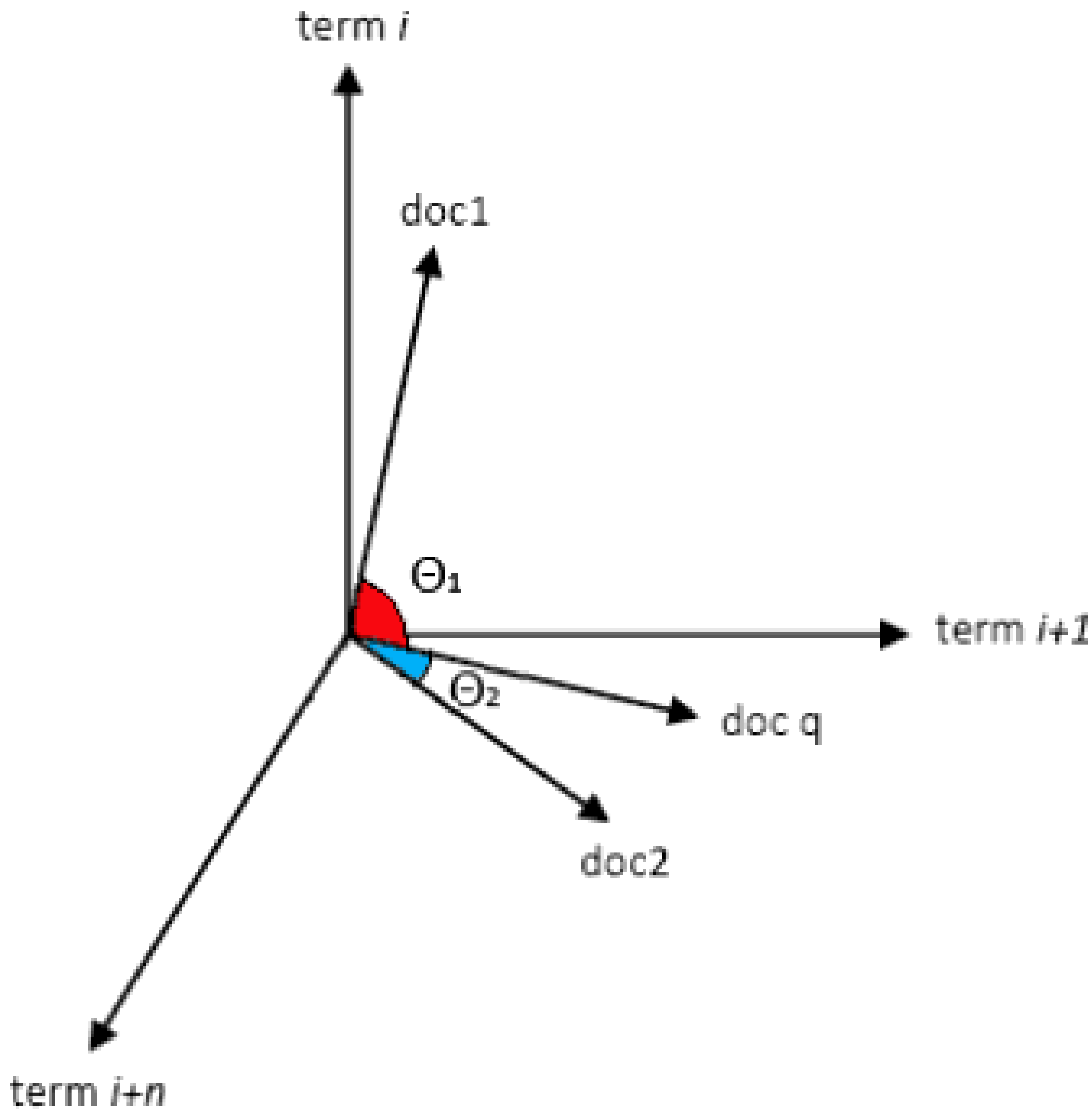


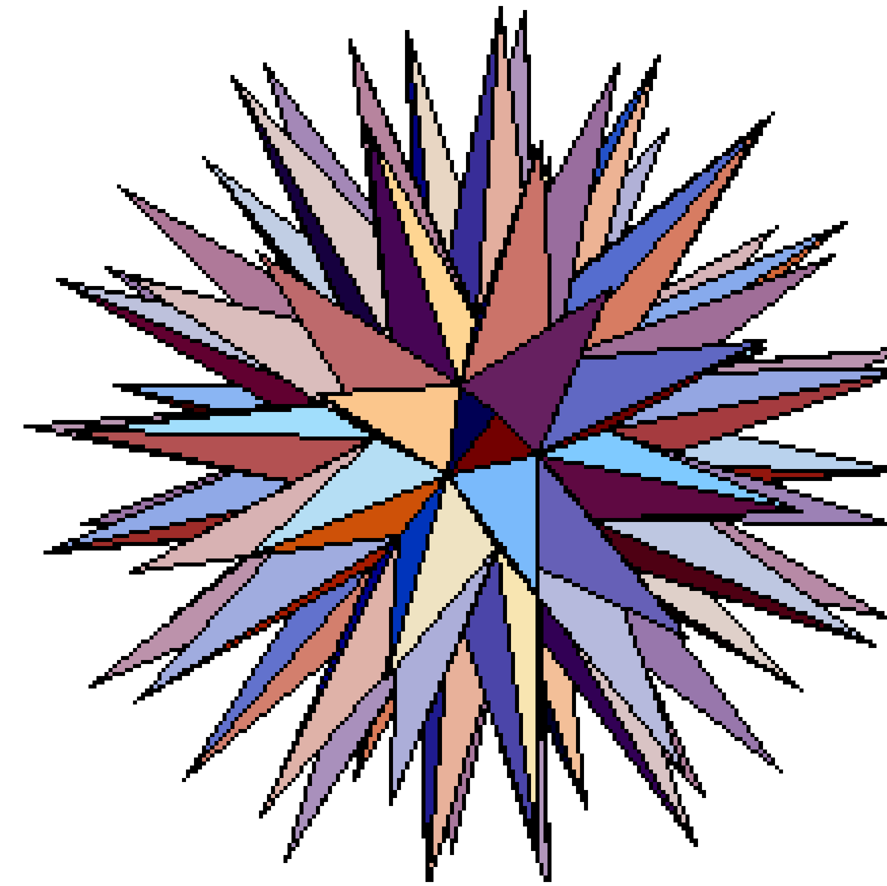
All “neighbors” are many **very far** away

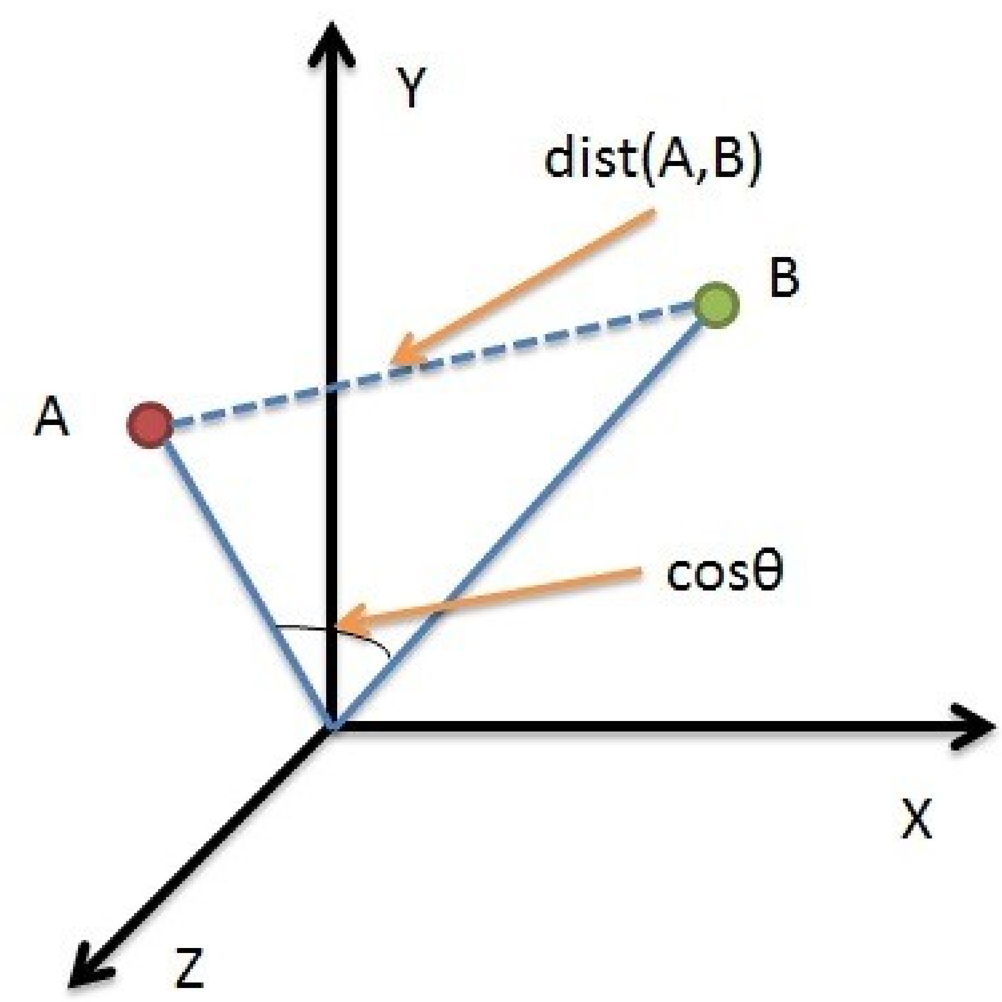


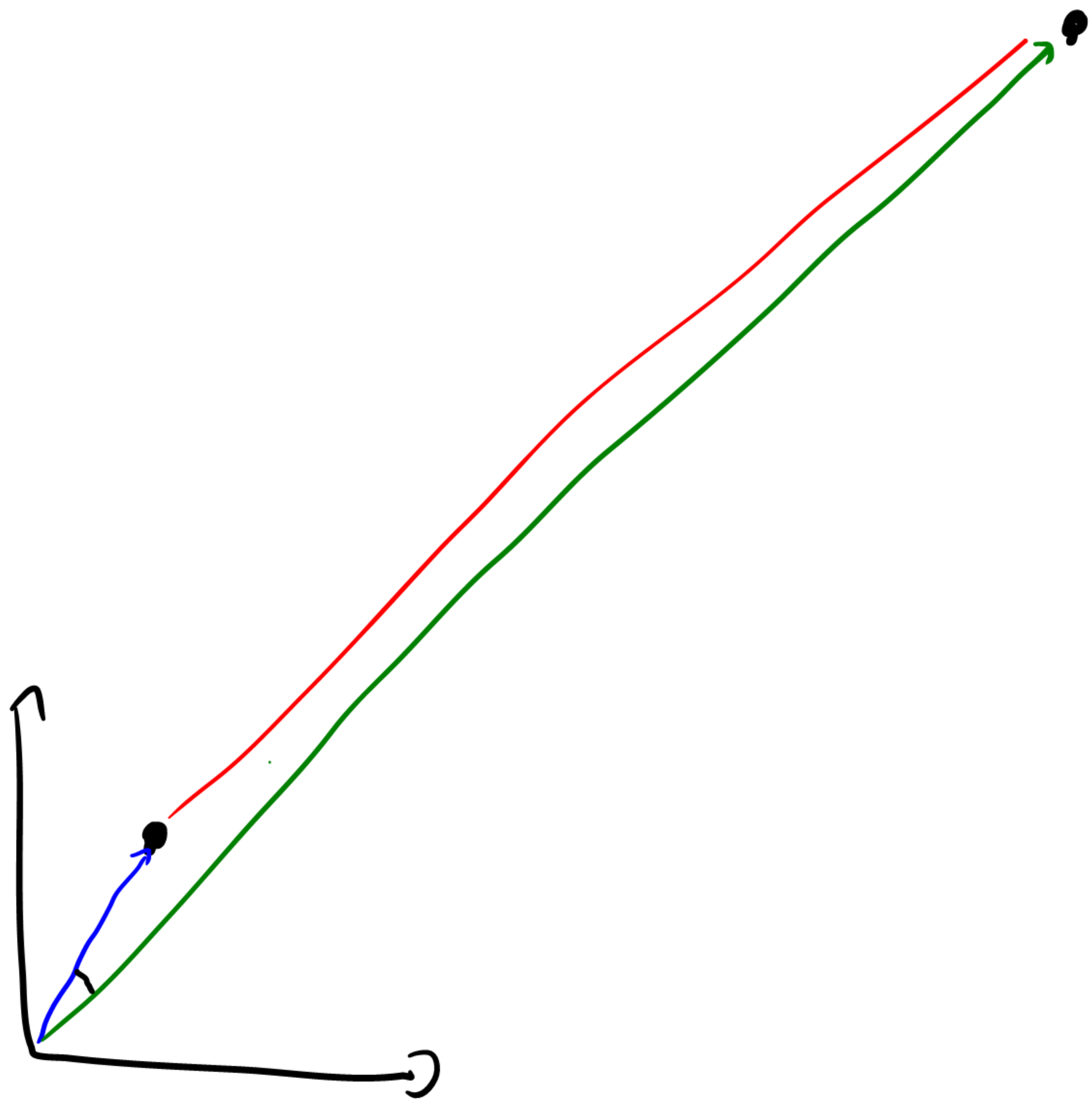


Cosine Distance



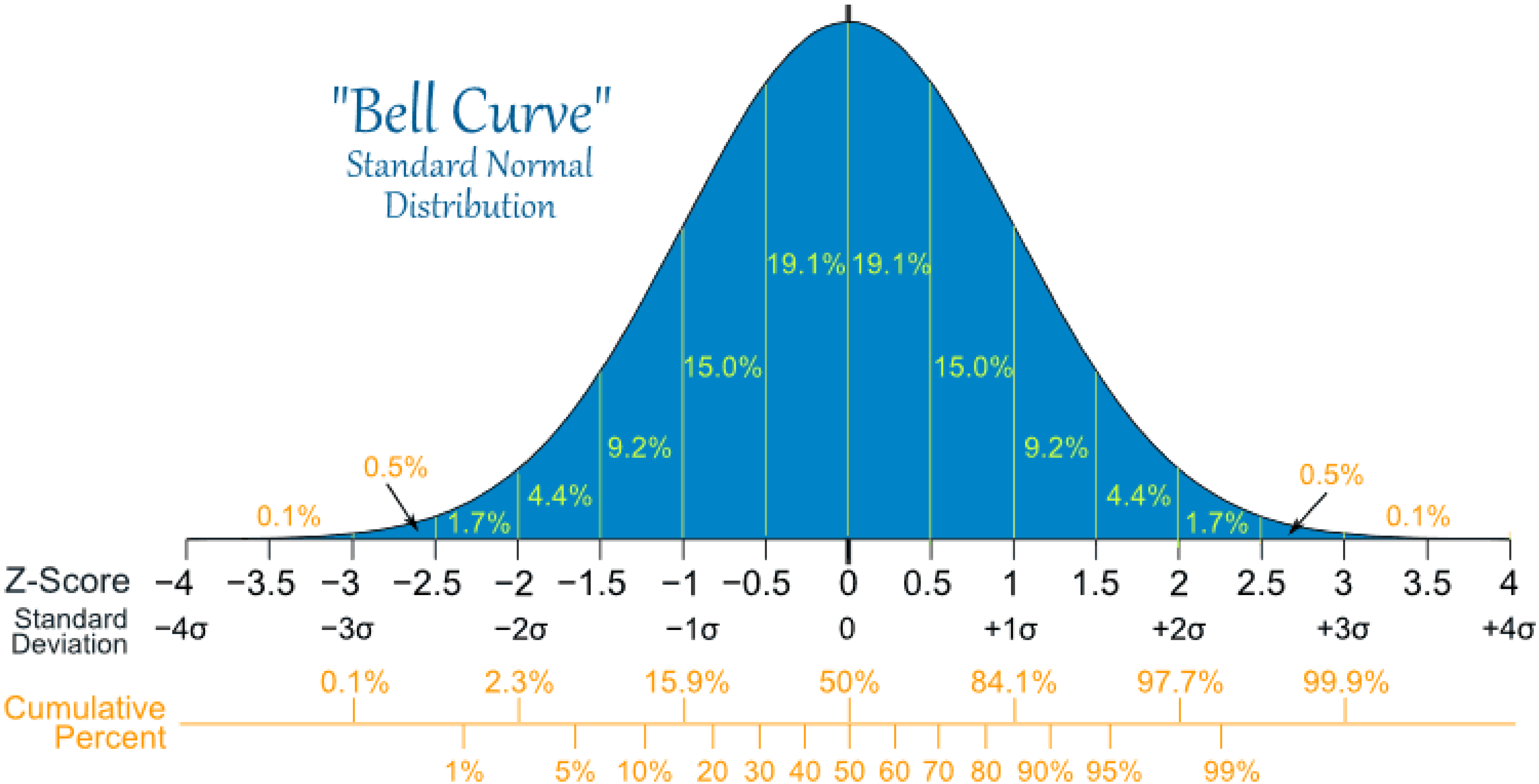


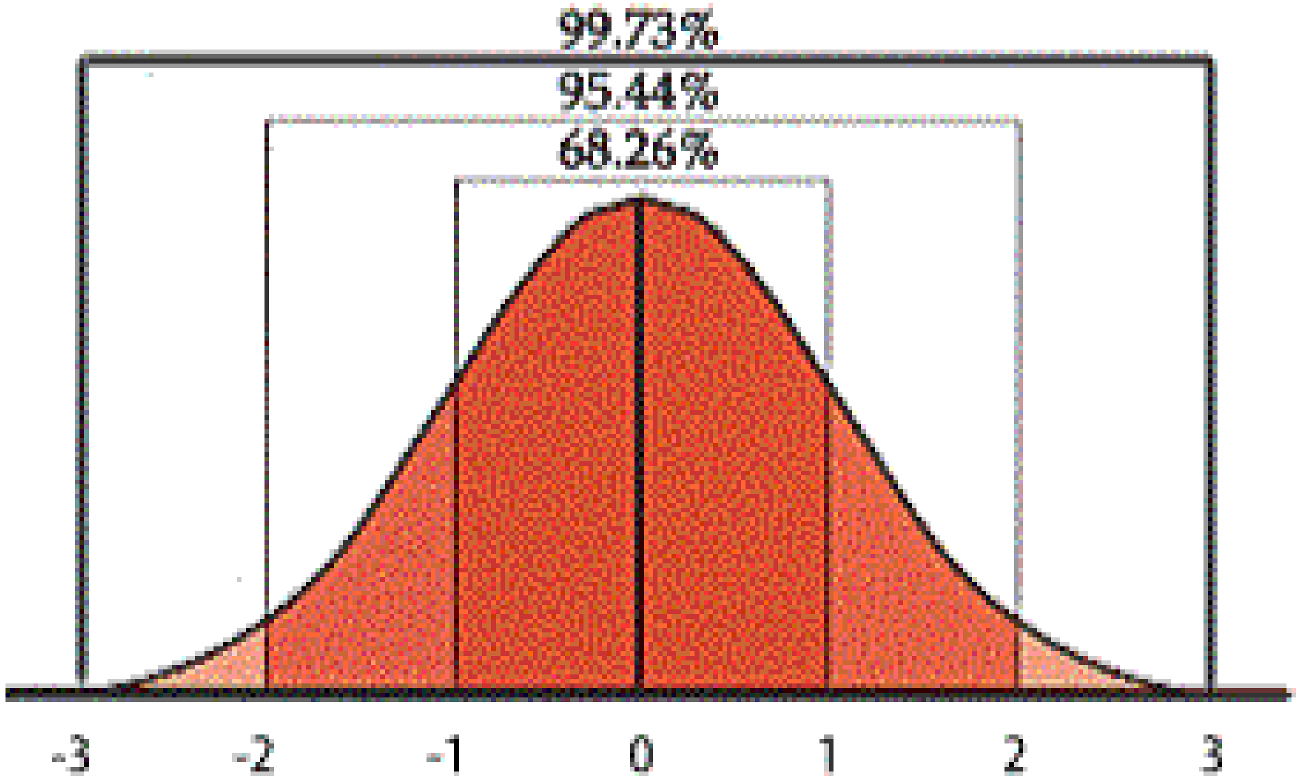




What does random mean in high-dimensions?

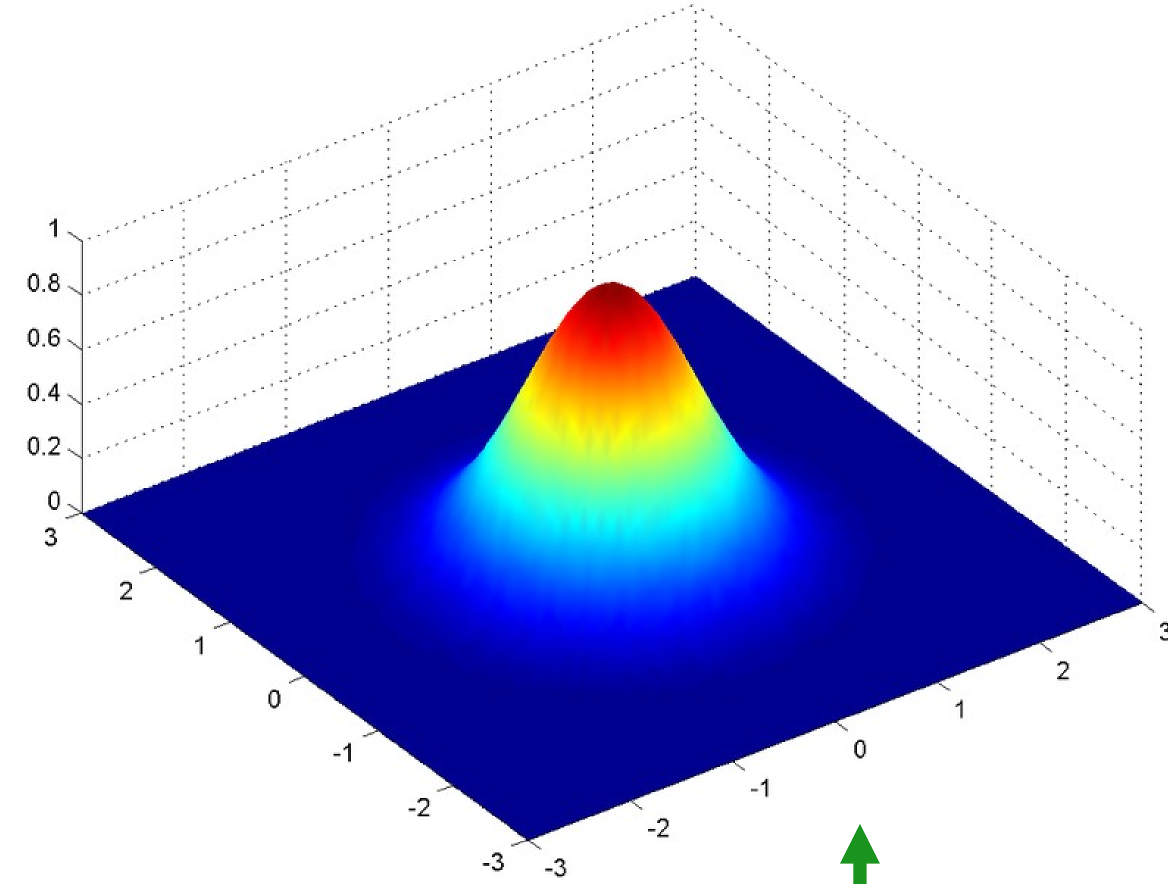
Question #3:
Gaussian Distribution (Center v.s. Surface)





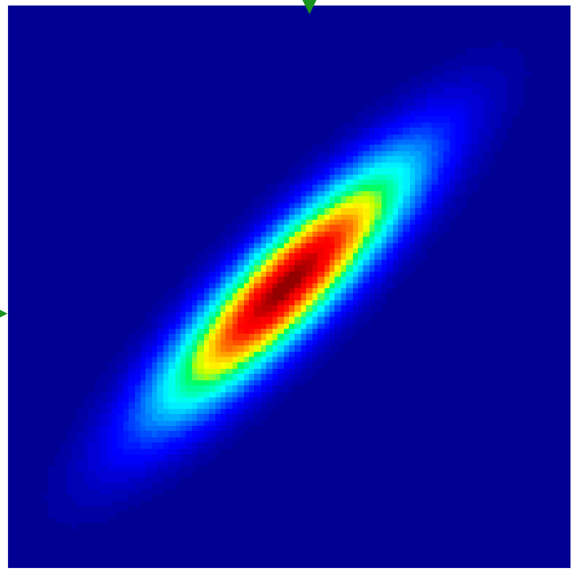
99.73%
95.44%
68.26%
-3 -2 -1 0 1 2 3

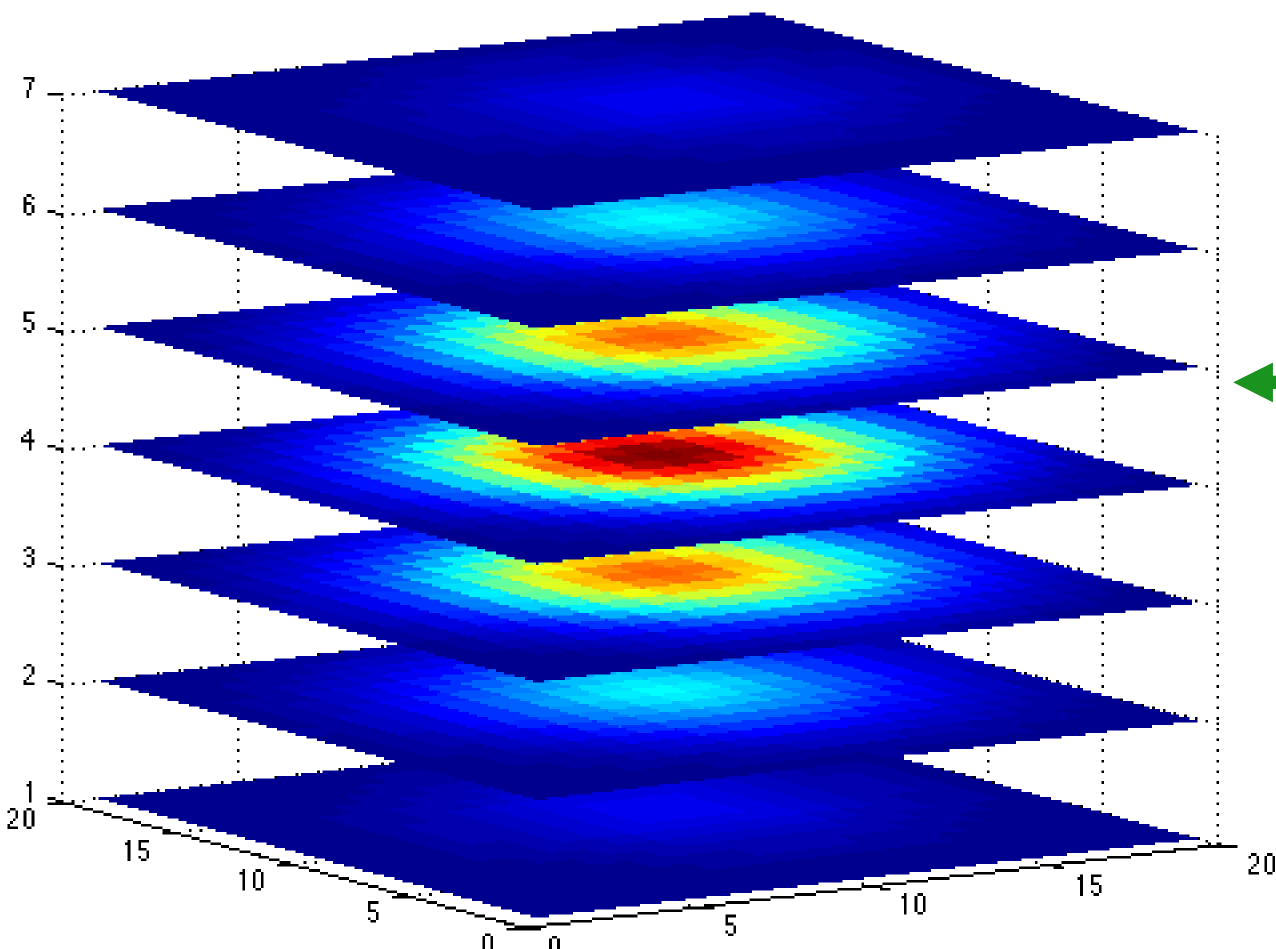
1 D



0 0.2 0.4 0.6 0.8 1
3 2 1 0 -1 -2 -3
-3 -2 -1 0 1 2

2 D

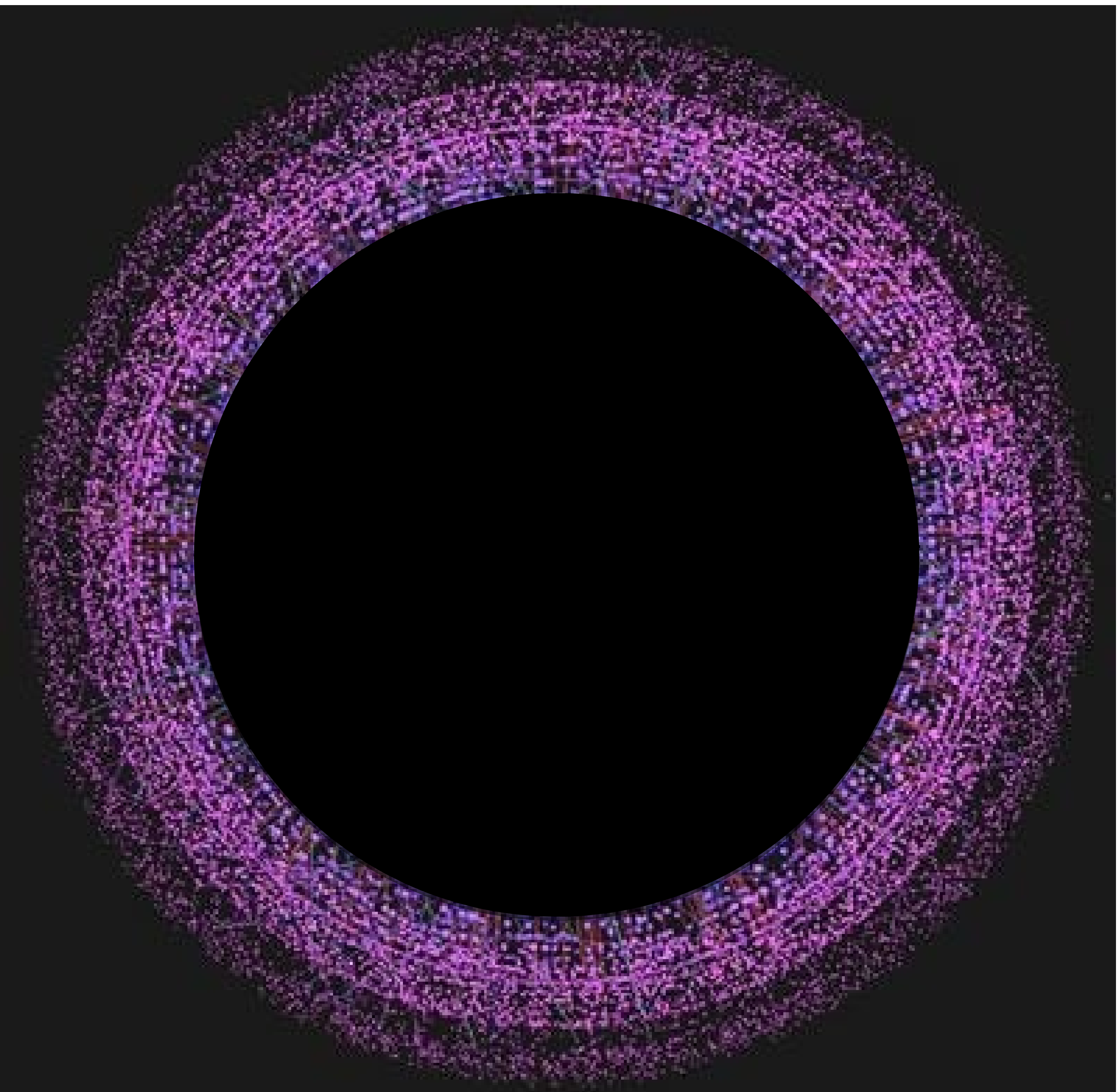


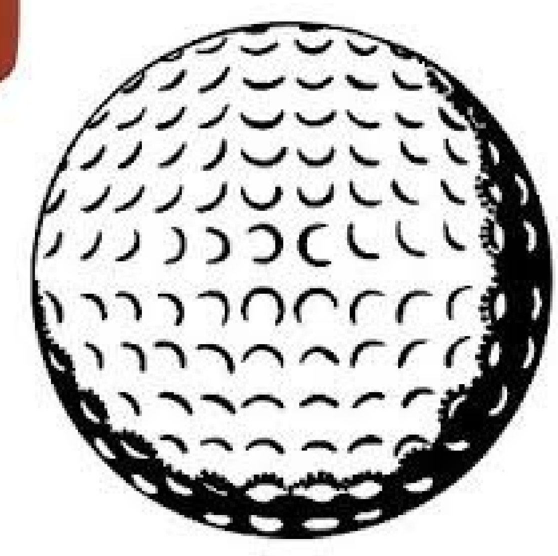


1 2 3 4 5 6 7
20 15 10 5 0
0 5 10 15 20

3 D

High Dimensional Gaussian Distribution





Question #3:
Gaussian Distribution (Center v.s. Surface)

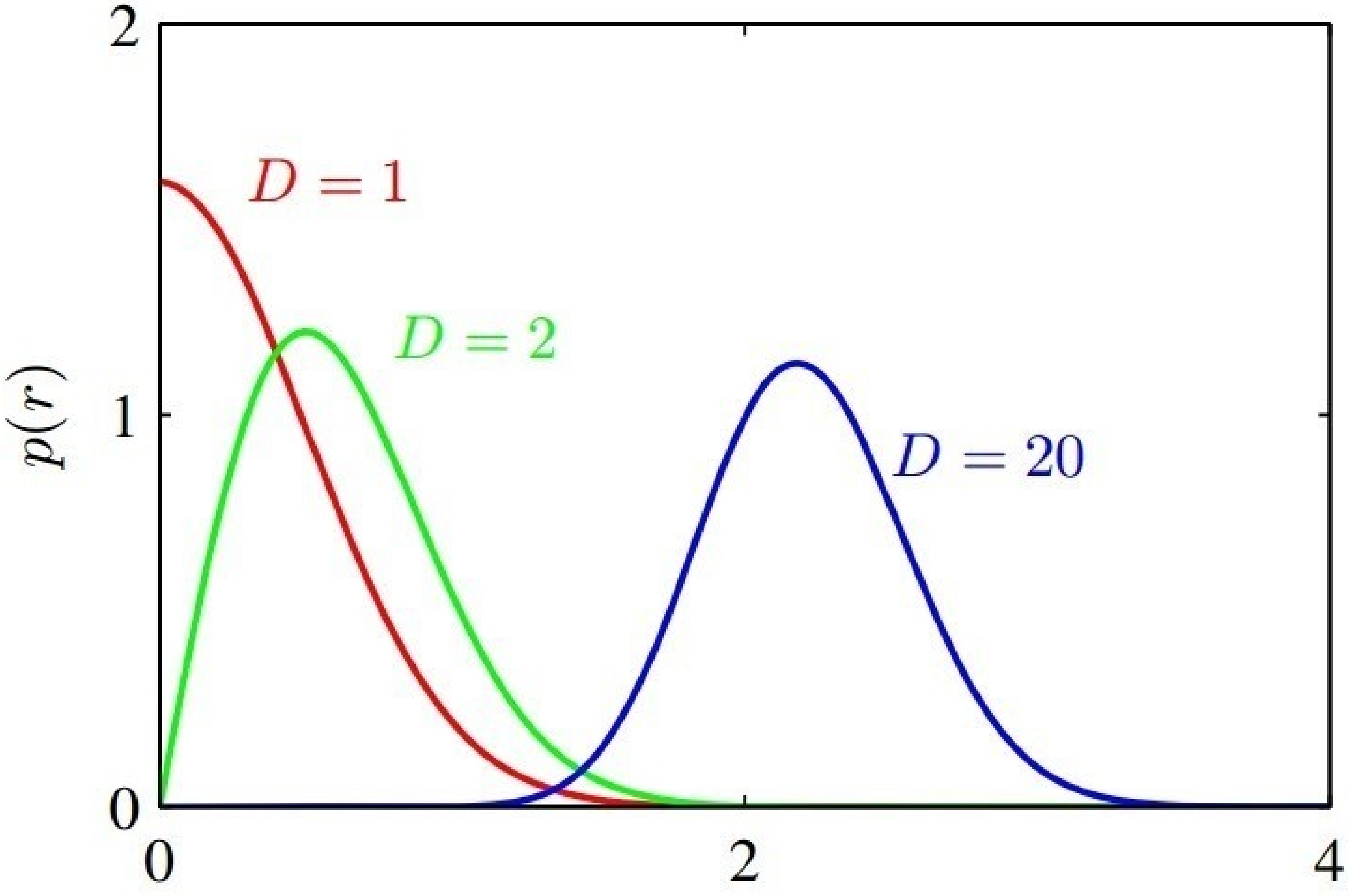
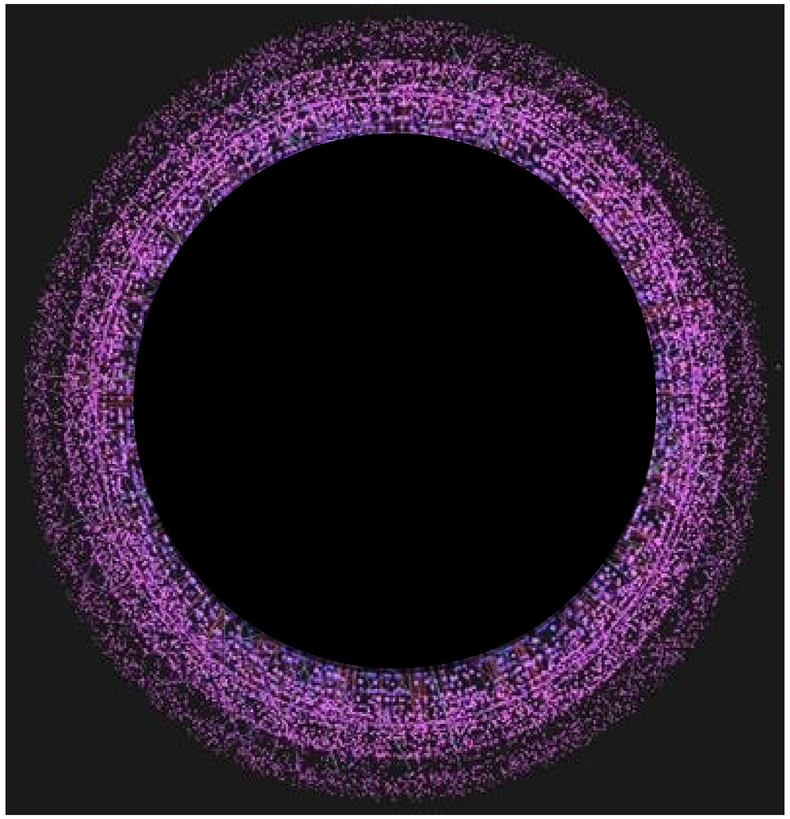


Figure 1.23 Plot of the probability density with respect to radius r of a Gaussian distribution for various values of the dimensionality D. In a high-dimensional space, most of the probability mass of a Gaussian is located within a thin shell at a specific radius.

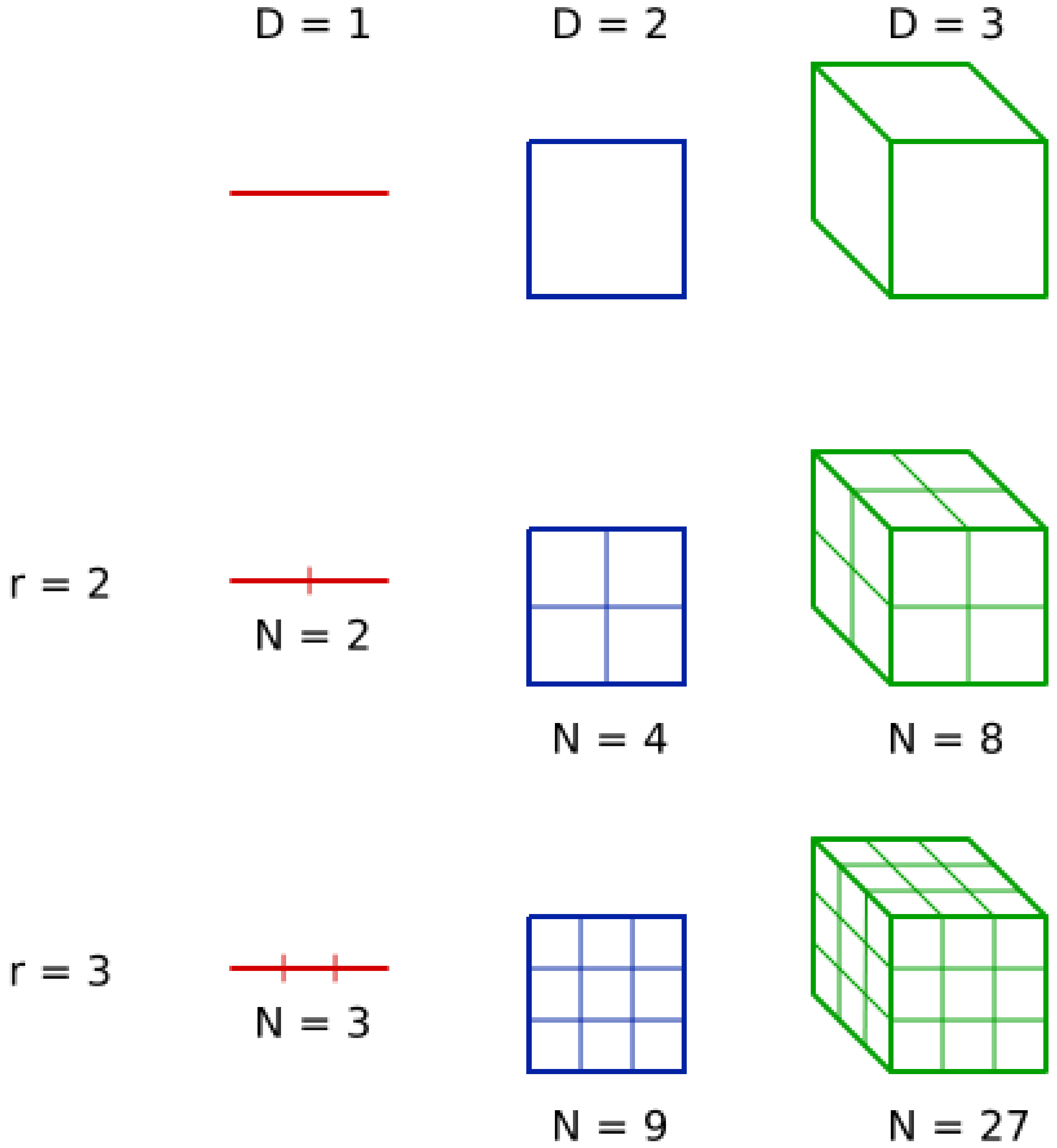
In High Dimensional Space: **Rule # 3**

- **All data points are on the surface**
  - By chance, data will never reach the inner space of the ball.



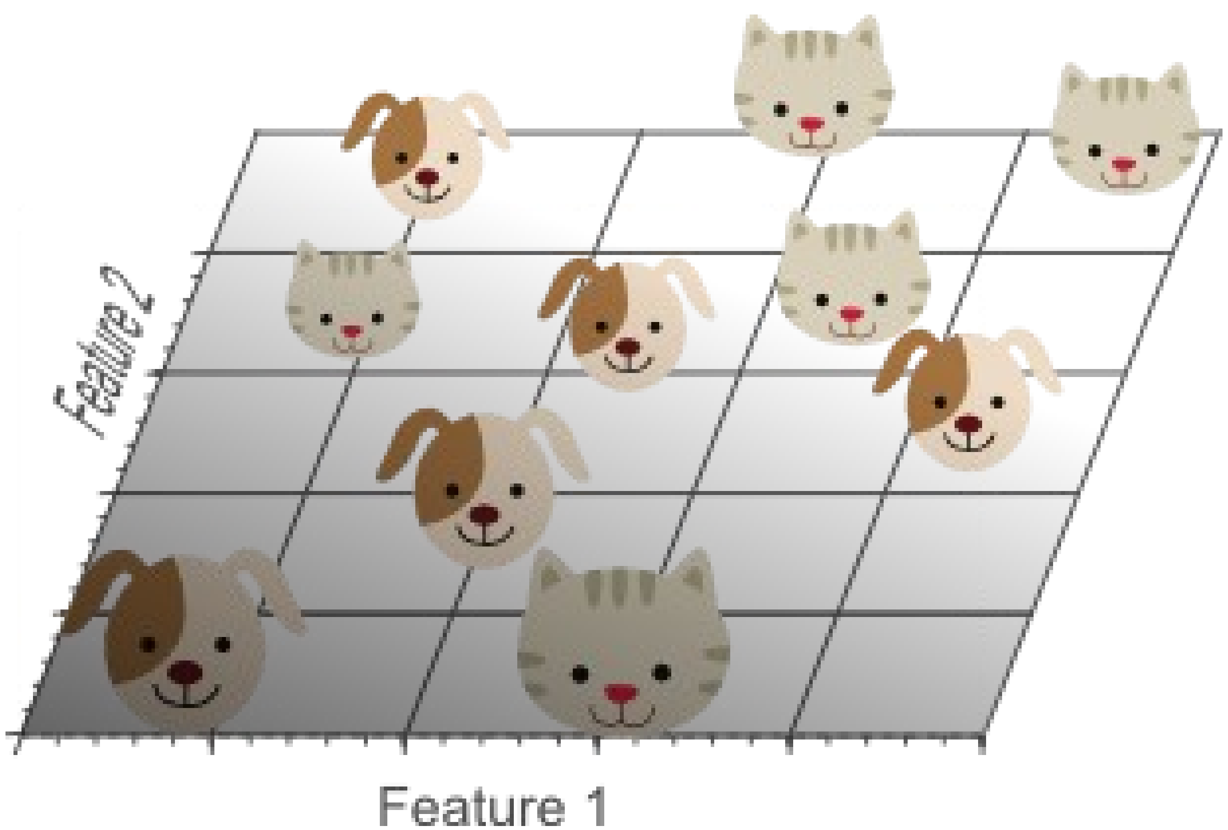
What does data mean in high-dimensions?

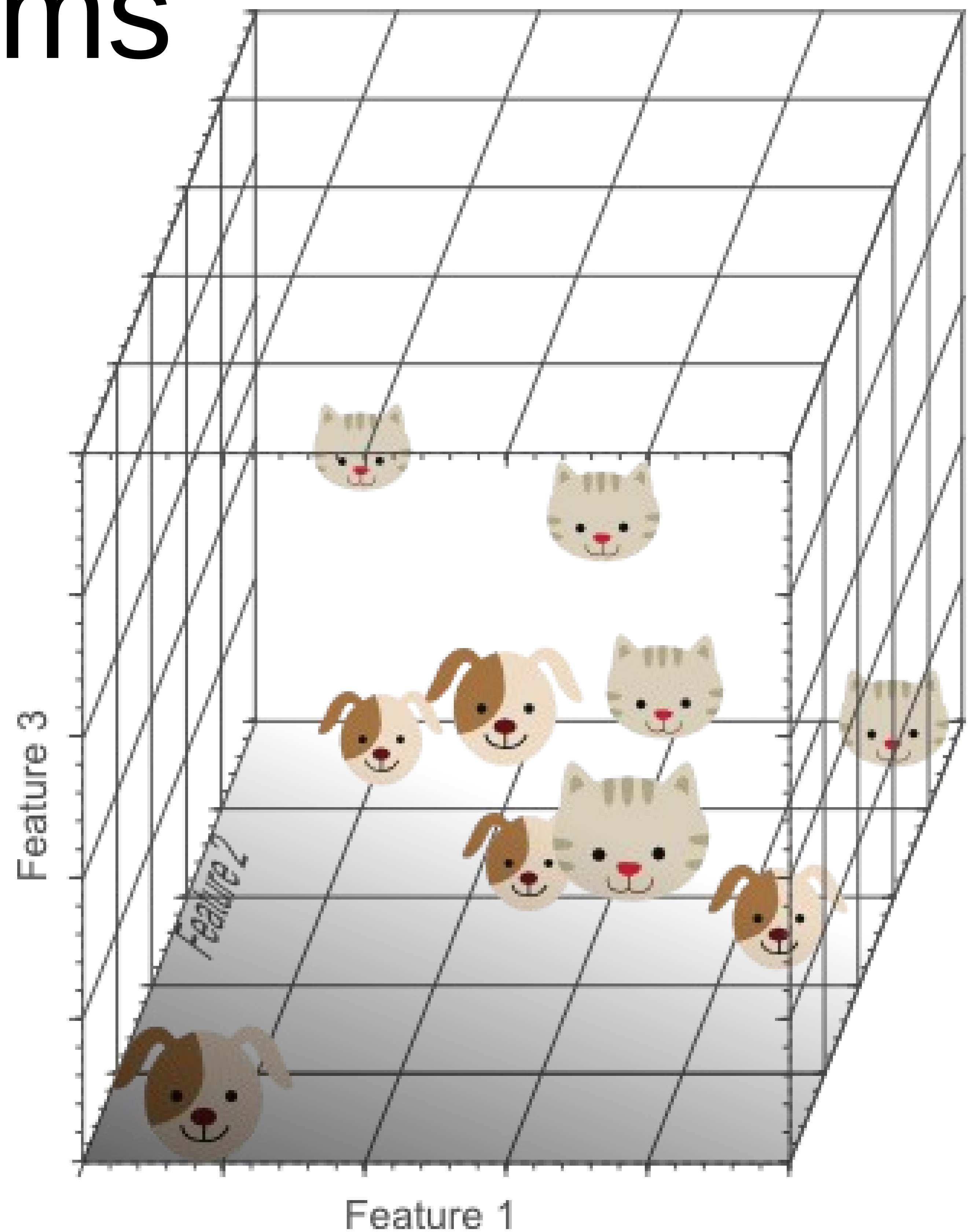
Question # 4: do we have enough data



Number of "Rooms"

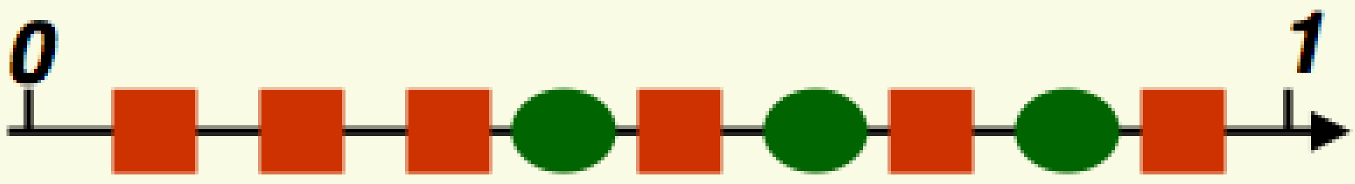
You need more data to fill the rooms





Data become sparser

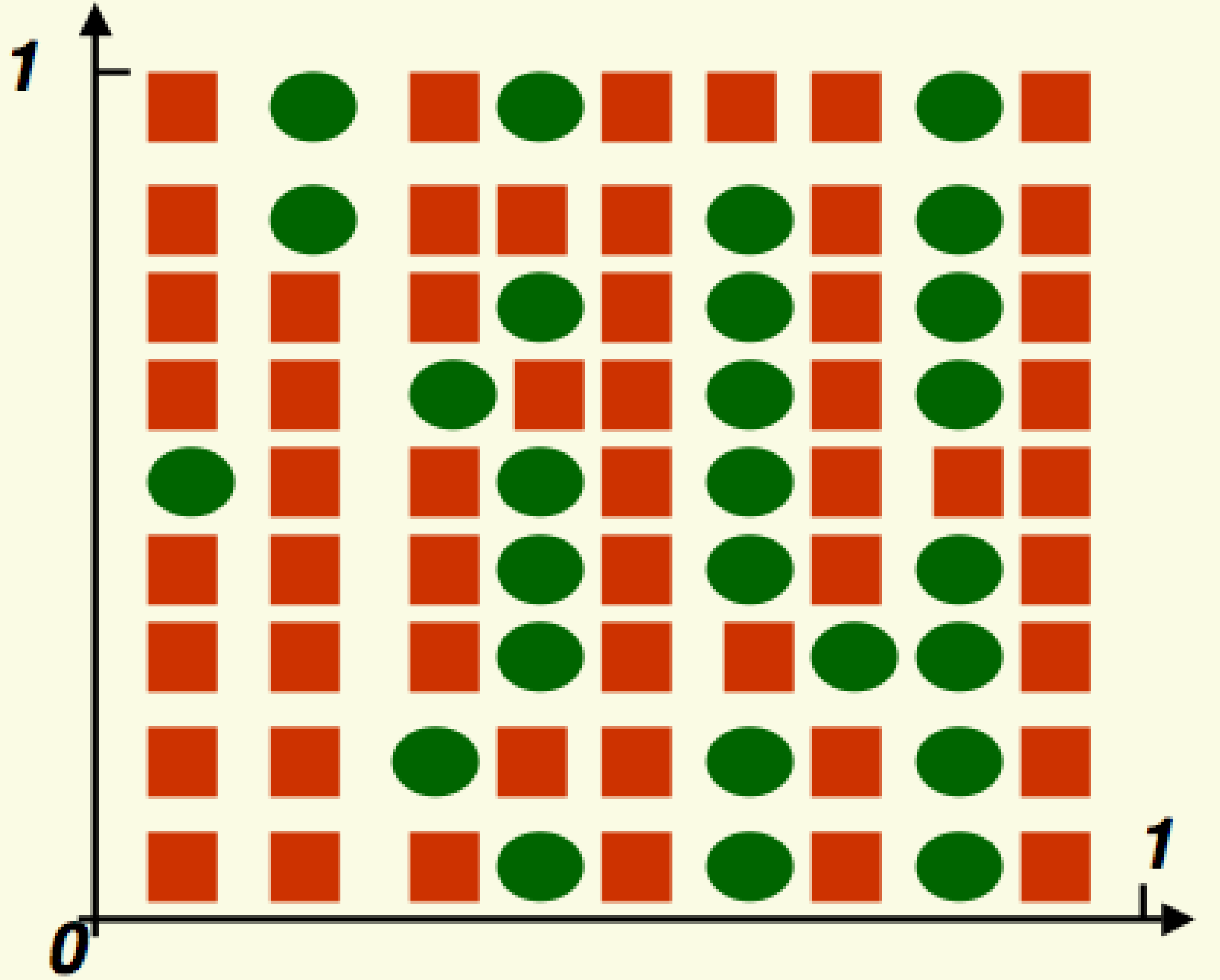
- Suppose we want to use the nearest neighbor approach with ***k*** = 1 (***1NN***)
- Suppose we start with only one feature



- This feature is not discriminative, i.e. it does not separate the classes well
- We decide to use 2 features. For the 1NN method to work well, need a lot of samples, i.e. samples have to be dense
- To maintain the same density as in 1D (9 samples per unit length), how many samples do we need?

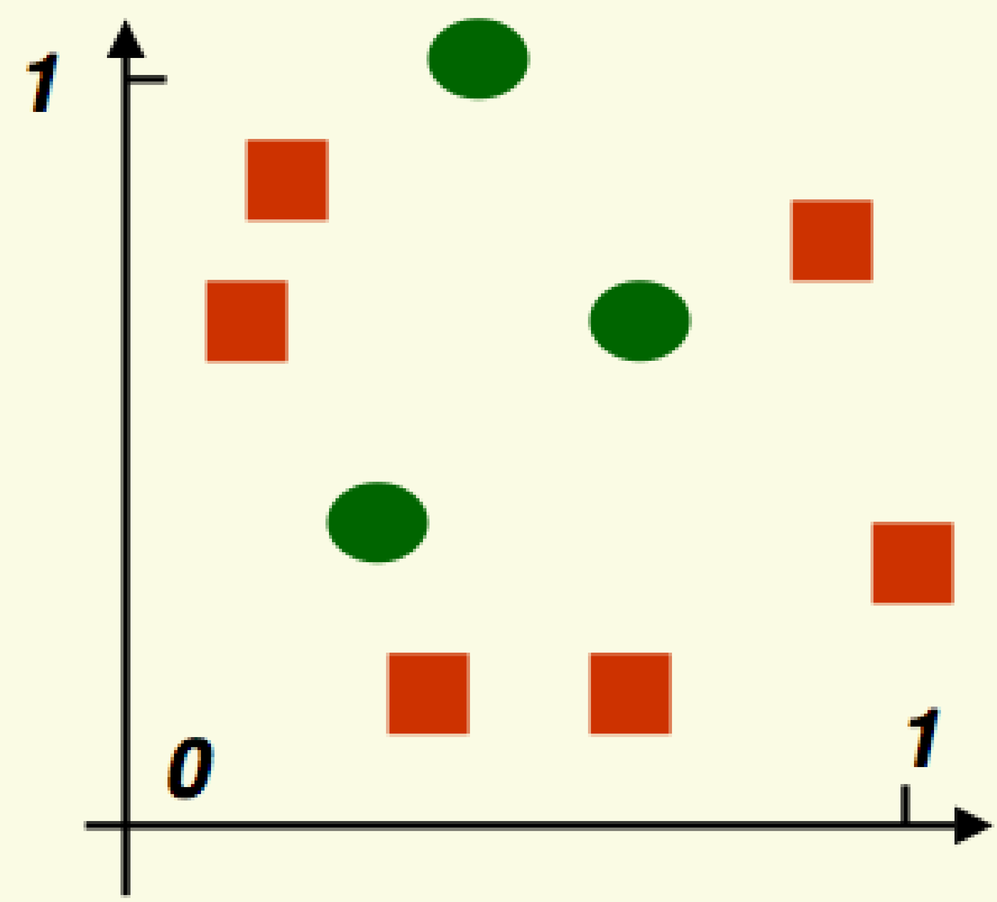
Data become Sparser: Number of Samples

- We need 9^2 samples to maintain the same density as in *1D*



Curse of Dimensionality: Number of Samples

- Of course, when we go from 1 feature to 2, no one gives us more samples, we still have 9



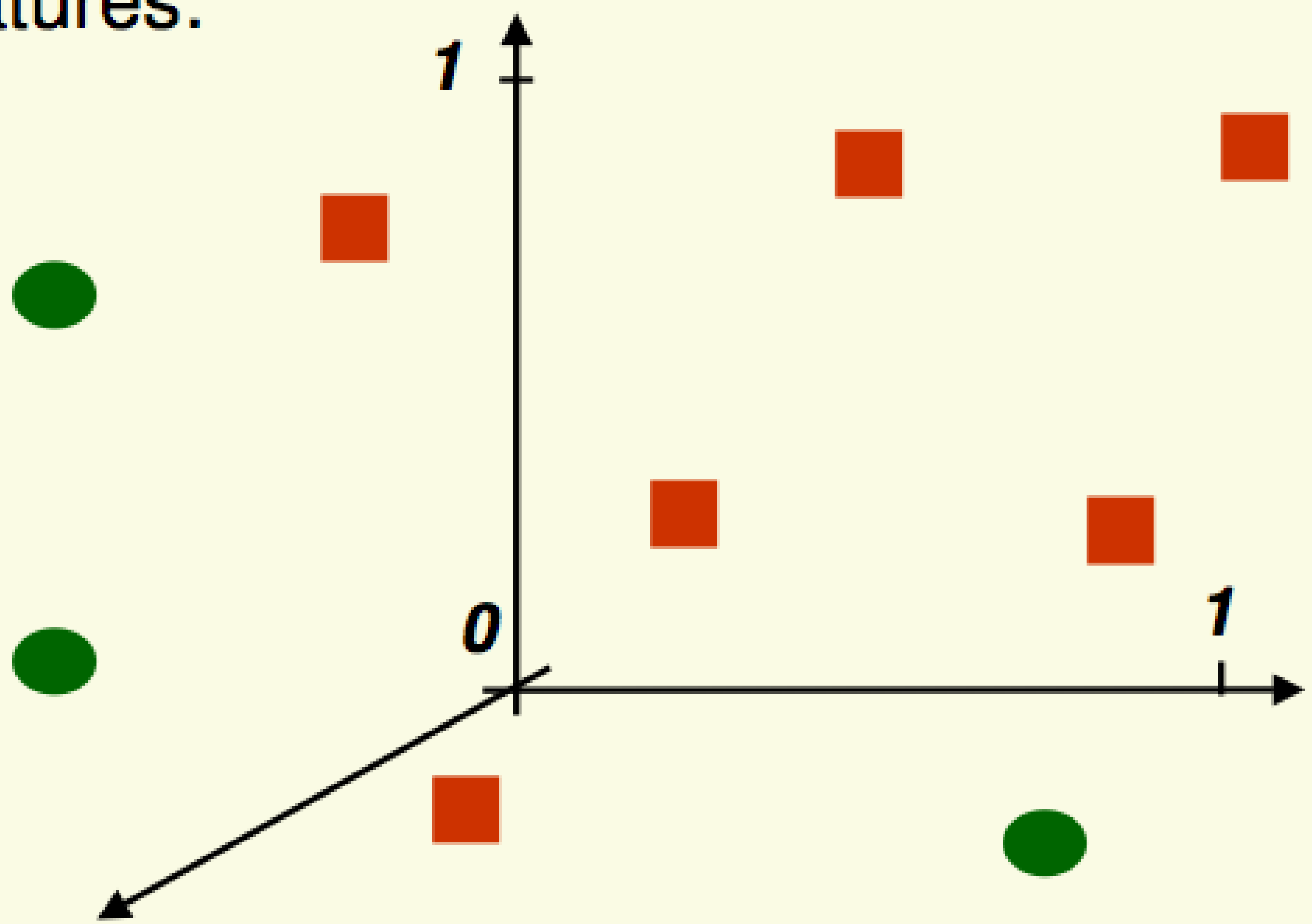
K^{d} ← High Dim

low Dim
Data
needed

- This is way too sparse for ***1NN*** to work well

Curse of Dimensionality: Number of Samples

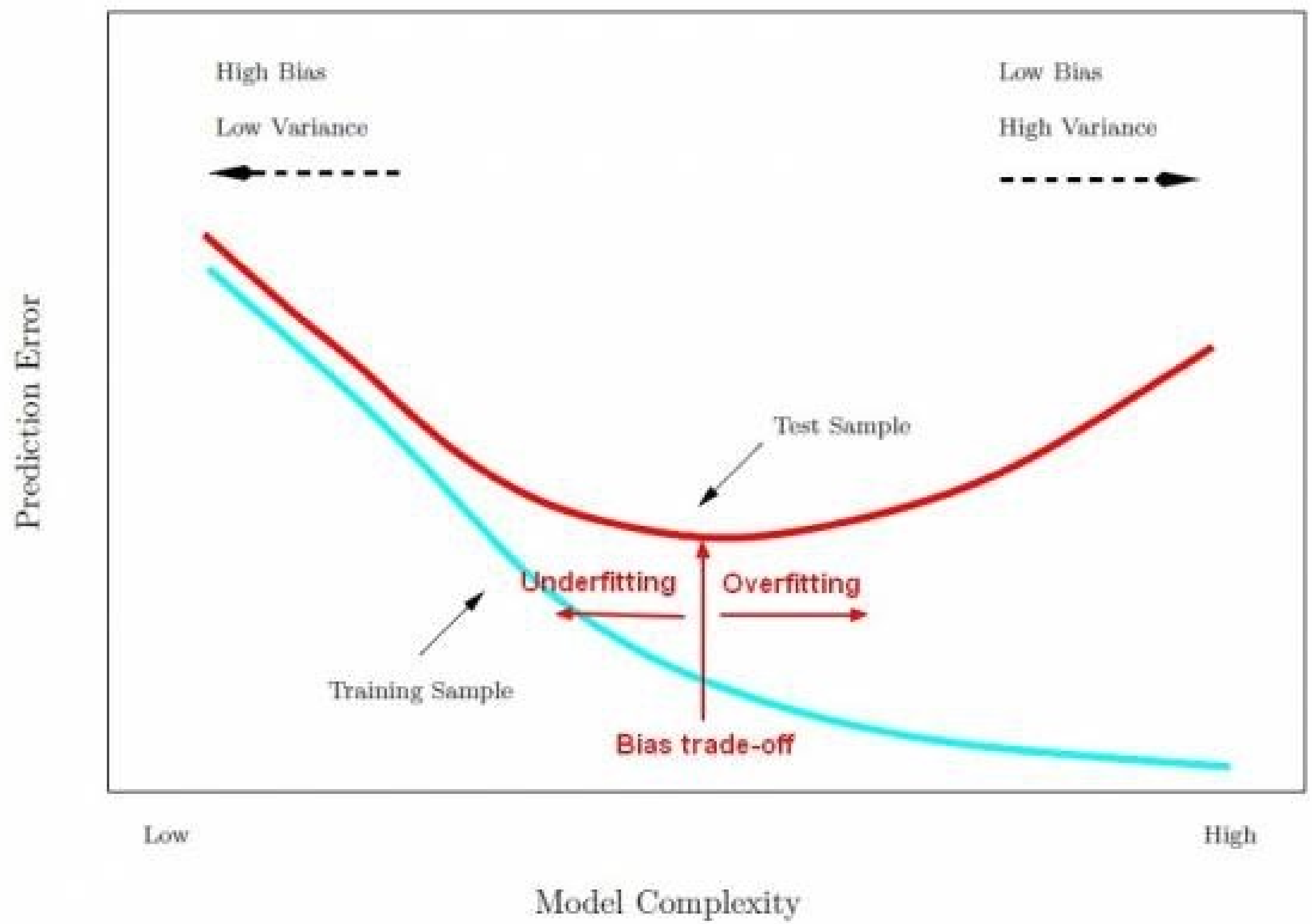
- Things go from bad to worse if we decide to use 3 features:



- If *9* was dense enough in 1D, in 3D we need $9^3=729$ samples!

Rule #4:
need **MORE** data to train your model

- models tend to **overfit** with a “small” amount of data

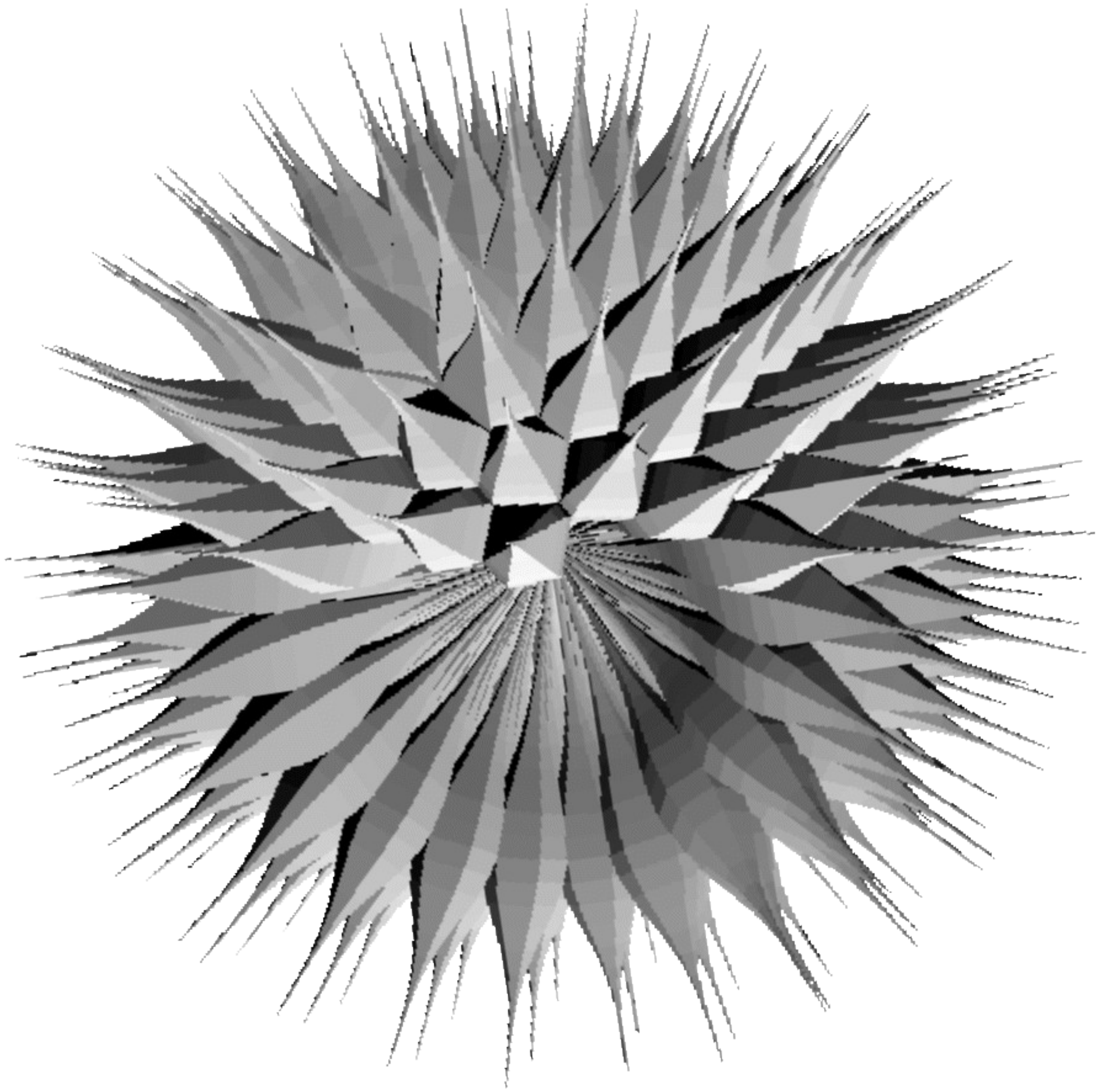


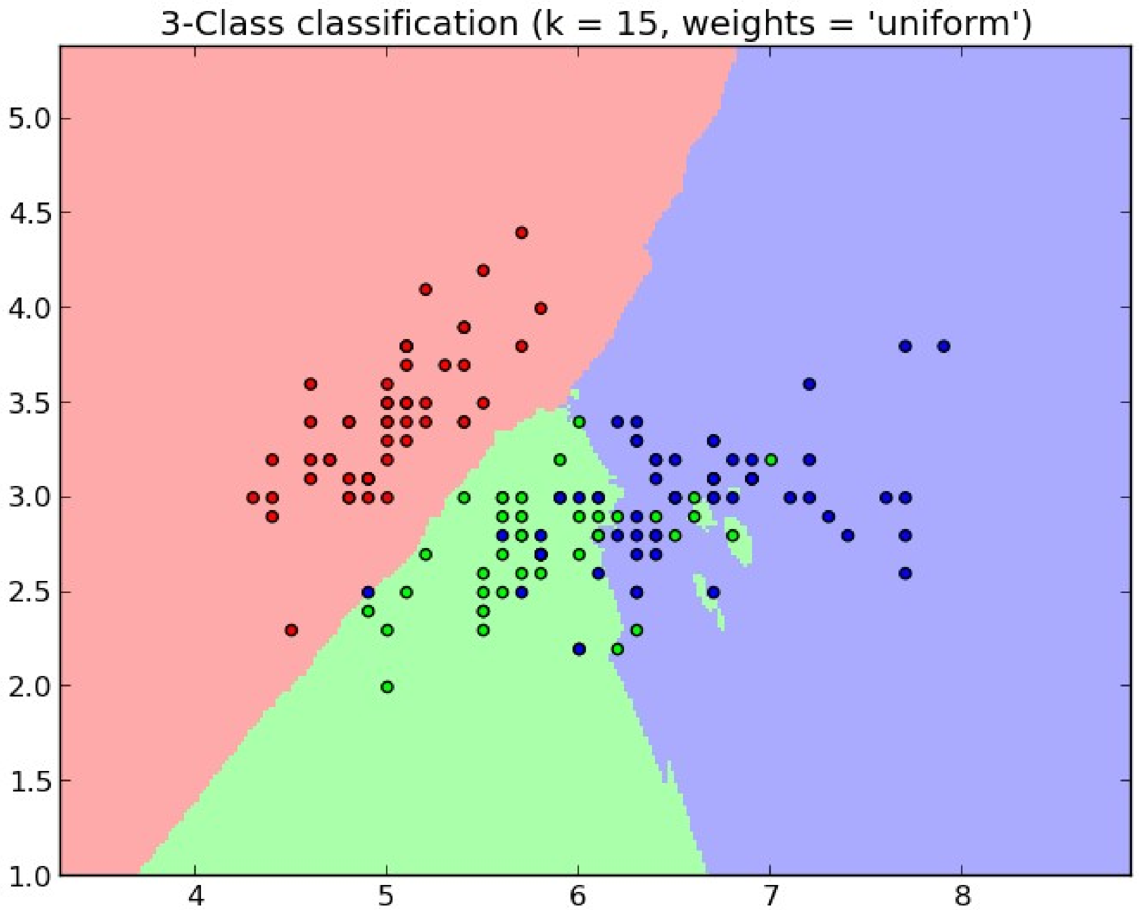
Curse of Dimensionality: Number of Samples

- In general, if ***n*** samples is dense enough in ***1D***
- Then in ***d*** dimensions we need n^d samples!
- And n^d grows really really fast as a function of ***d***
- Common pitfall:
  - If we can't solve a problem with a few features, adding more features seems like a good idea
  - However the number of samples usually stays the same
  - The method with more features is likely to perform worse instead of expected better

Now you know what is really happening

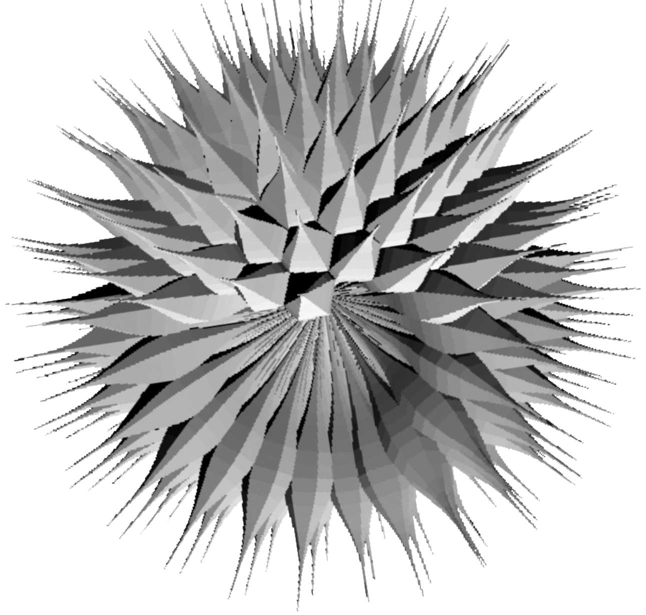
in High Dimensional Space

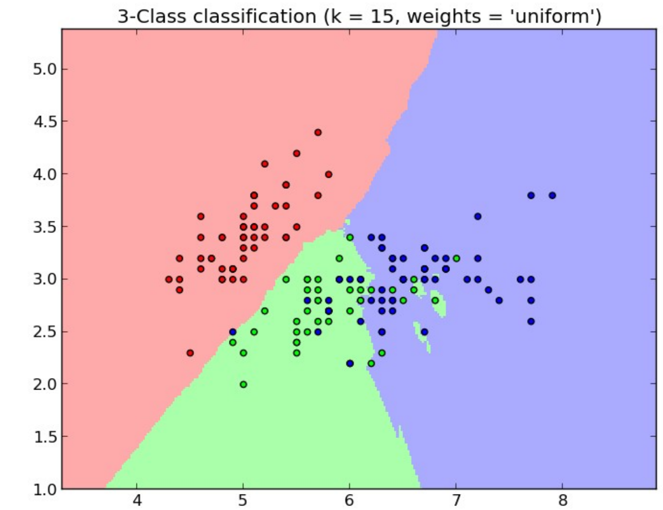




Real world

Perfect world

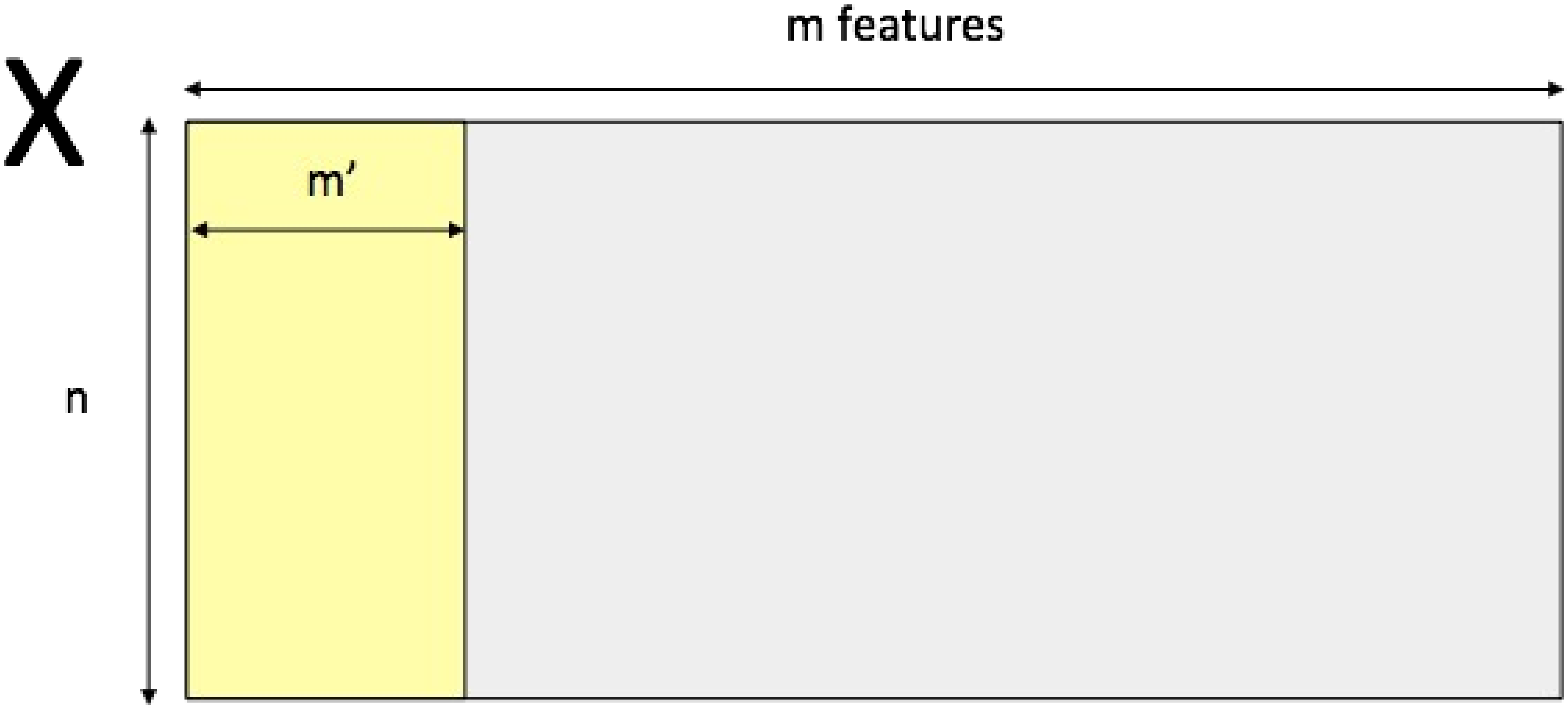




- Feature Selection
  - Filtering approach
  - Wrapper approach
  - Embedded methods

- Dimensionality Reduction
  - Principal Components Analysis (PCA)
  - Nonlinear PCA (Kernel PCA, CatPCA)
  - Multi-Dimensional Scaling (MDS)
  - Homogeneity Analysis

Feature Selection



Feature Selection

- Filtering approach:

  ranks features or feature subsets independently of the predictor (classifier).

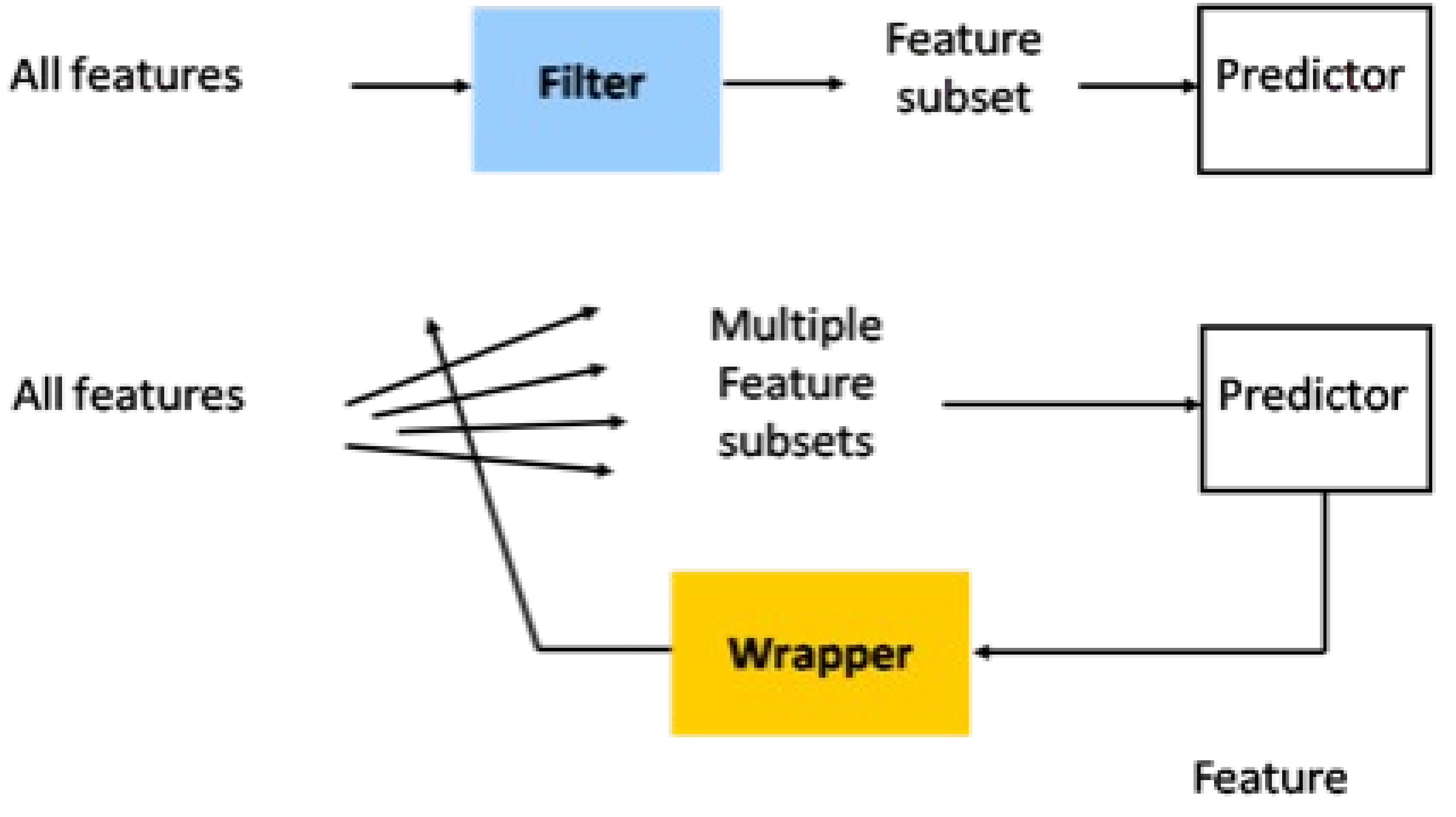
  - ...using univariate methods: consider one variable at a time
  - ...using multivariate methods: consider more than one variables at a time

- Wrapper approach:

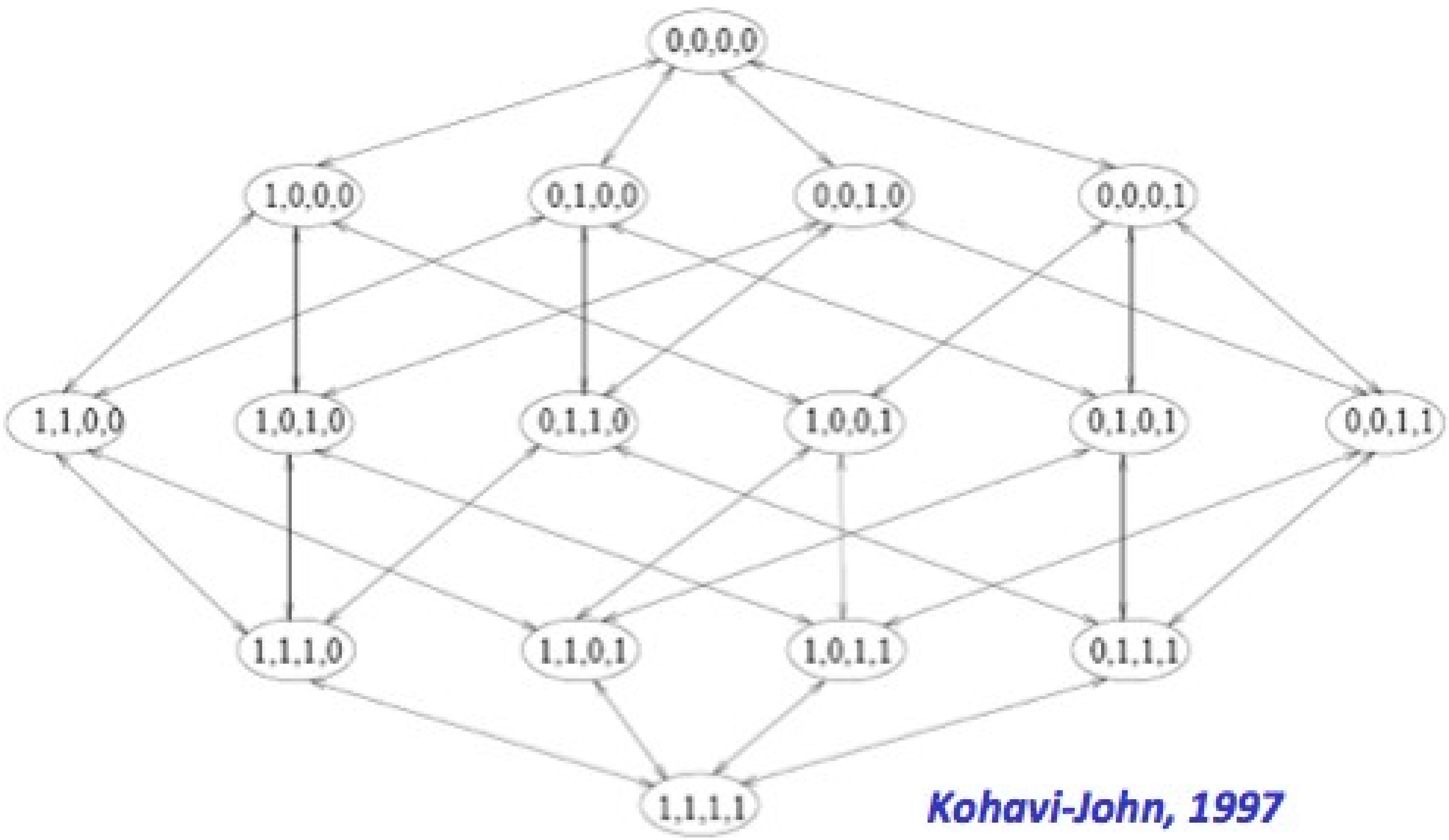
  uses a classifier to assess (many) features or feature subsets.

Feature selection: Filter vs. Wrapper methods

- **Main goal**: rank subsets of useful features



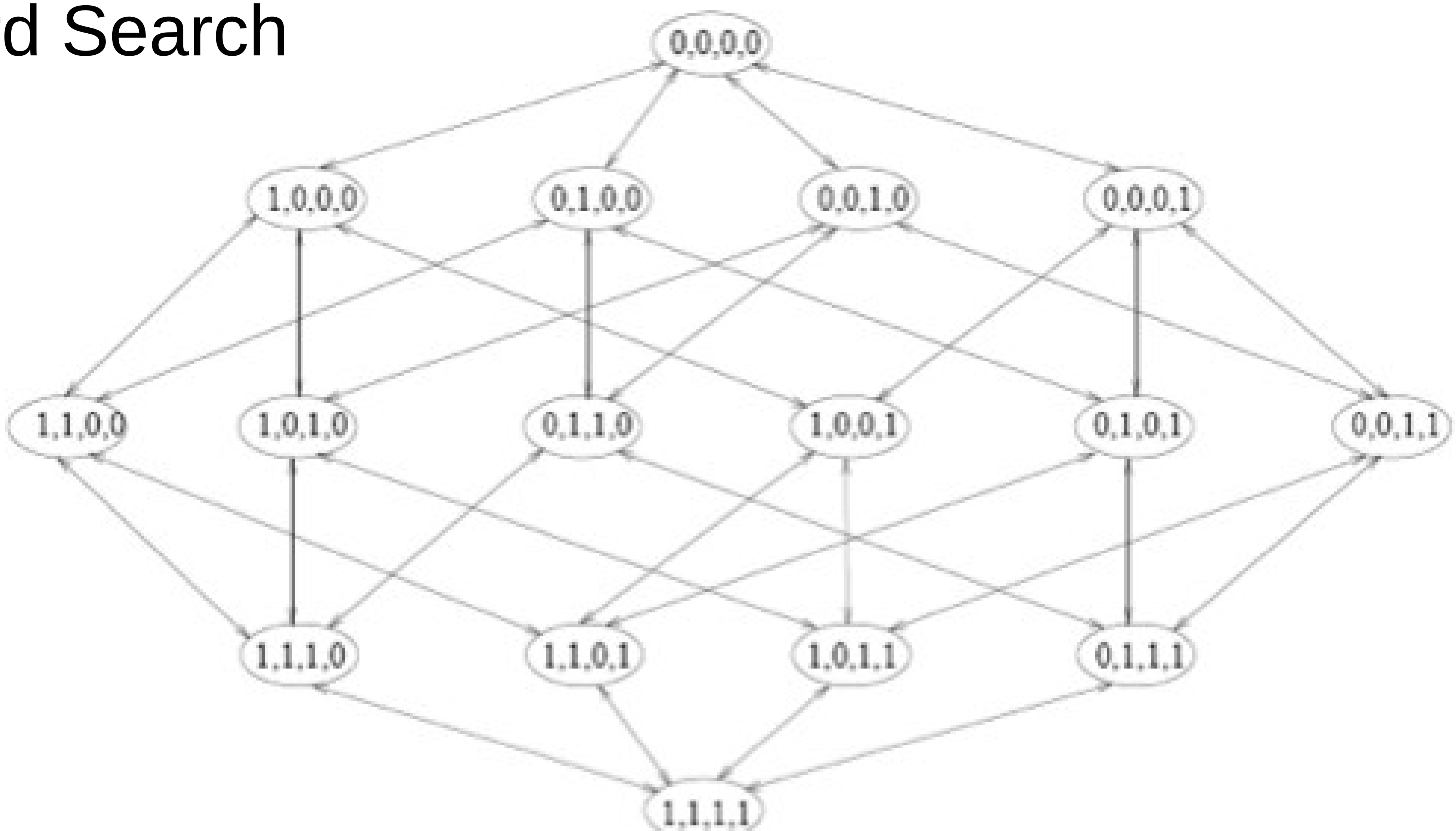
Feature Selection



N features, 2^N possible feature subsets!

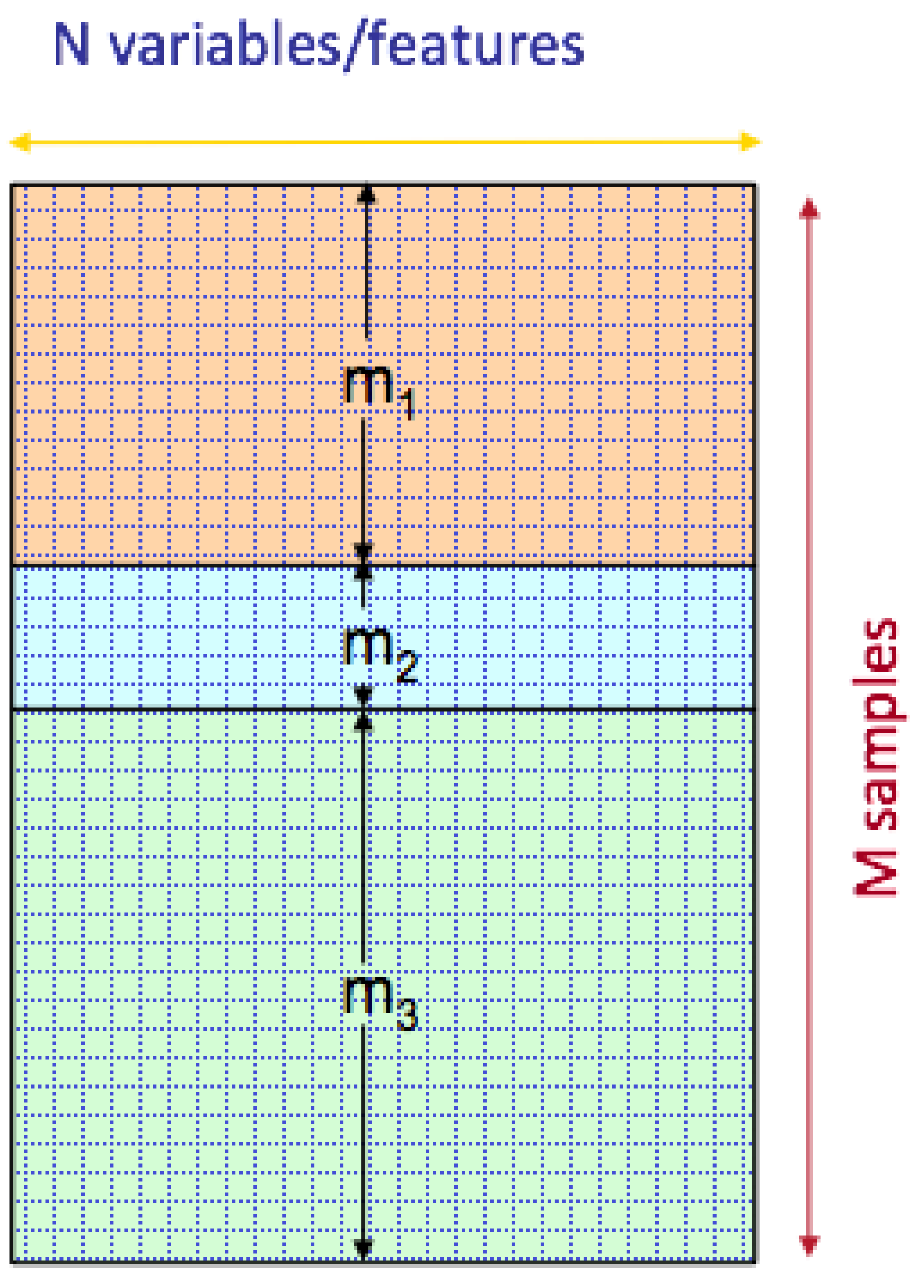
Search Strategies

Forward Search



Backward Search

Wrapper methods



Split data into 3 sets:

training, validation, and test set.

1) For each feature subset, train predictor on training data.

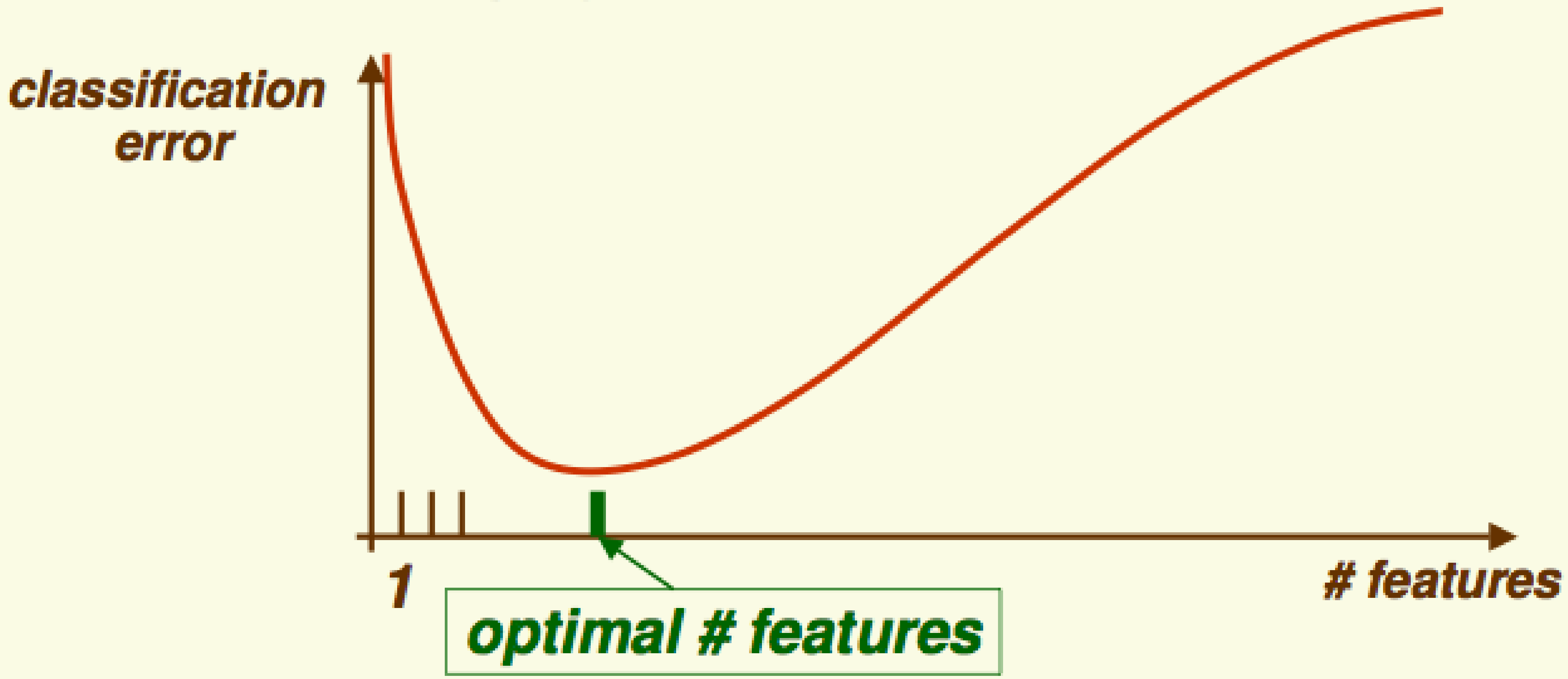
2) Select the feature subset, which performs best on validation data.
   - Repeat and average if you want to reduce variance (cross-validation).

3) Test on test data.

Danger of over-fitting with intensive search!

Over-fitting: Number of Samples

- For a fixed number of samples, as we add features, the graph of classification error:



- Thus for each fixed sample size ***n***, there is the optimal number of features to use

Dimensionality Reduction

- High dimensionality is challenging and redundant
- It is natural to try to reduce dimensionality
- Reduce dimensionality by feature combination: combine old features ***x*** to create new features ***y***

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \rightarrow f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}\right) = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} = \boldsymbol{y} \quad \text{with } k < d$$

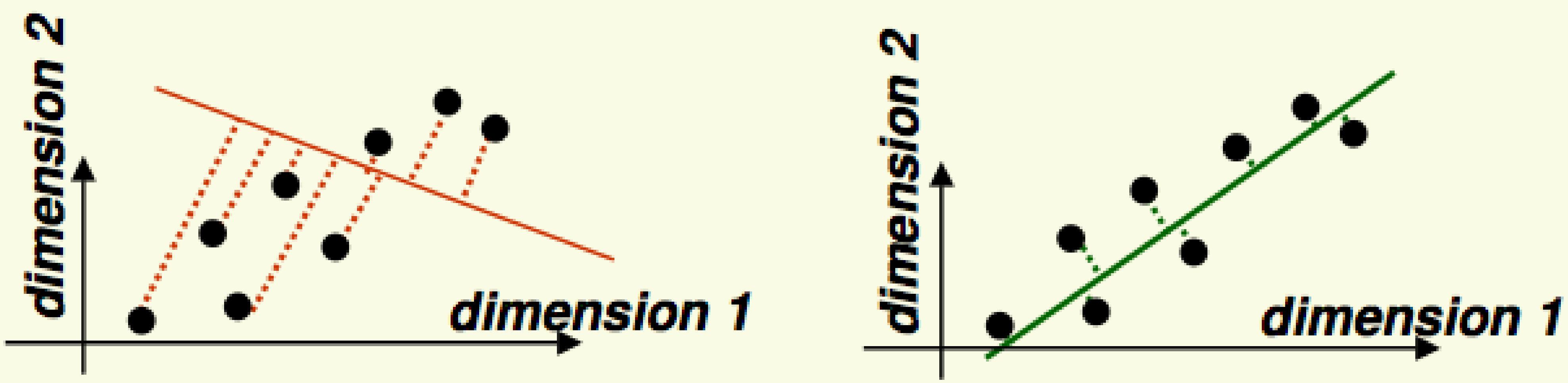
- For example,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 + x_2 \\ x_3 + x_4 \end{bmatrix} = \boldsymbol{y}$$

- Ideally, the new vector ***y*** should retain from ***x*** all information important for classification

PCA:Principle Component Analysis

- Main idea: seek most accurate data representation in a lower dimensional space
- Example in 2-D
  - Project data to 1-D subspace (a line) which minimize the projection error

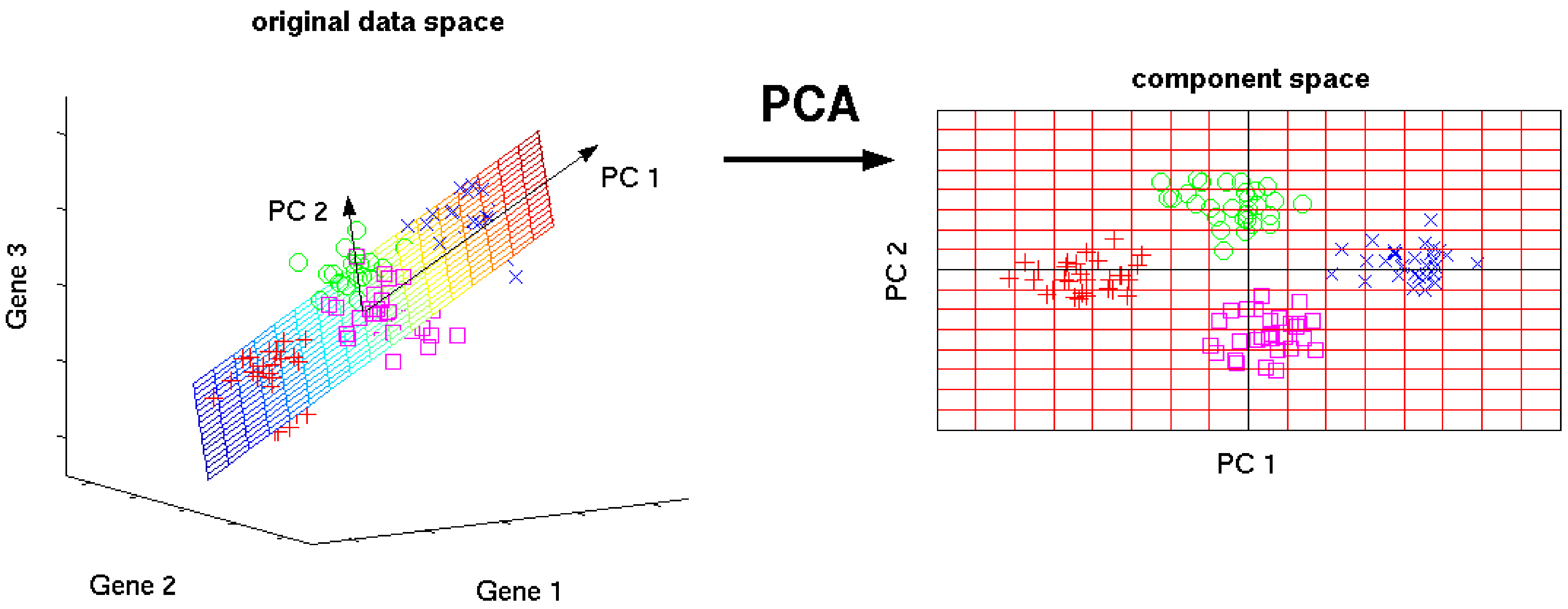


large projection errors, bad line to project to

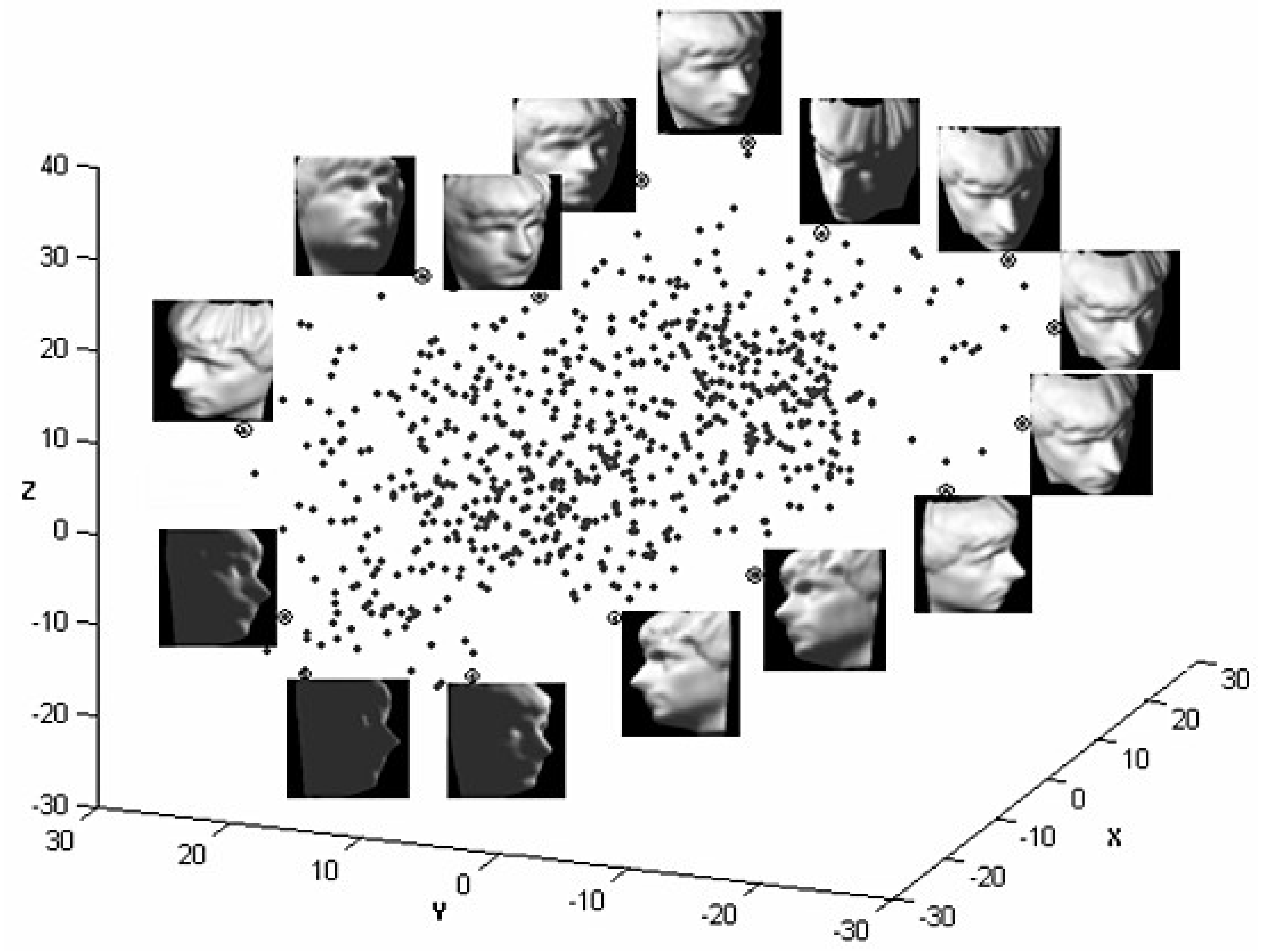
small projection errors, good line to project to

- Notice that the the good line to use for projection lies in the direction of largest variance

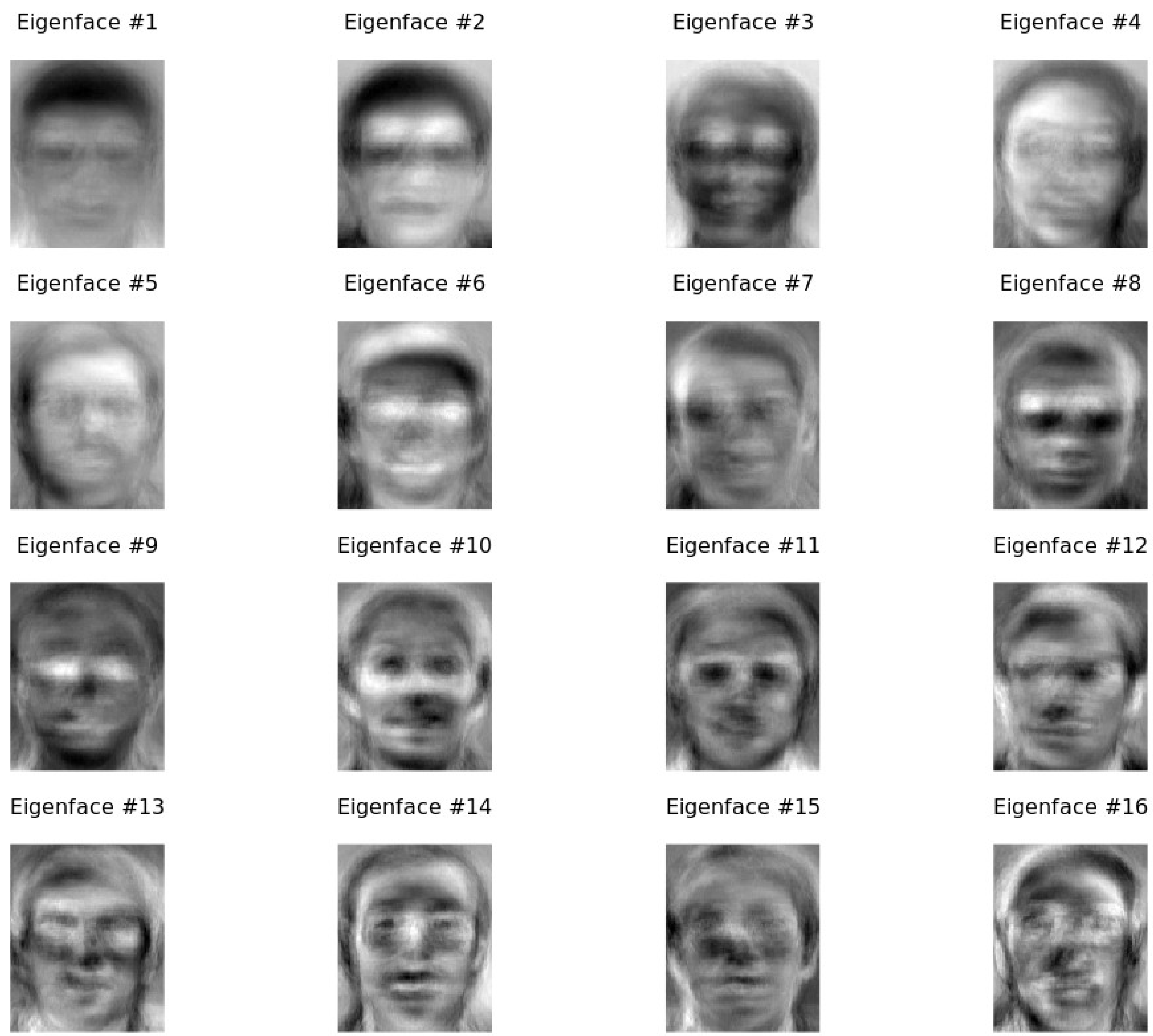
PCA in 3D


original data space
PC 1
PC 2
Gene 3
Gene 2
Gene 1
PCA
component space
PC 2
PC 1

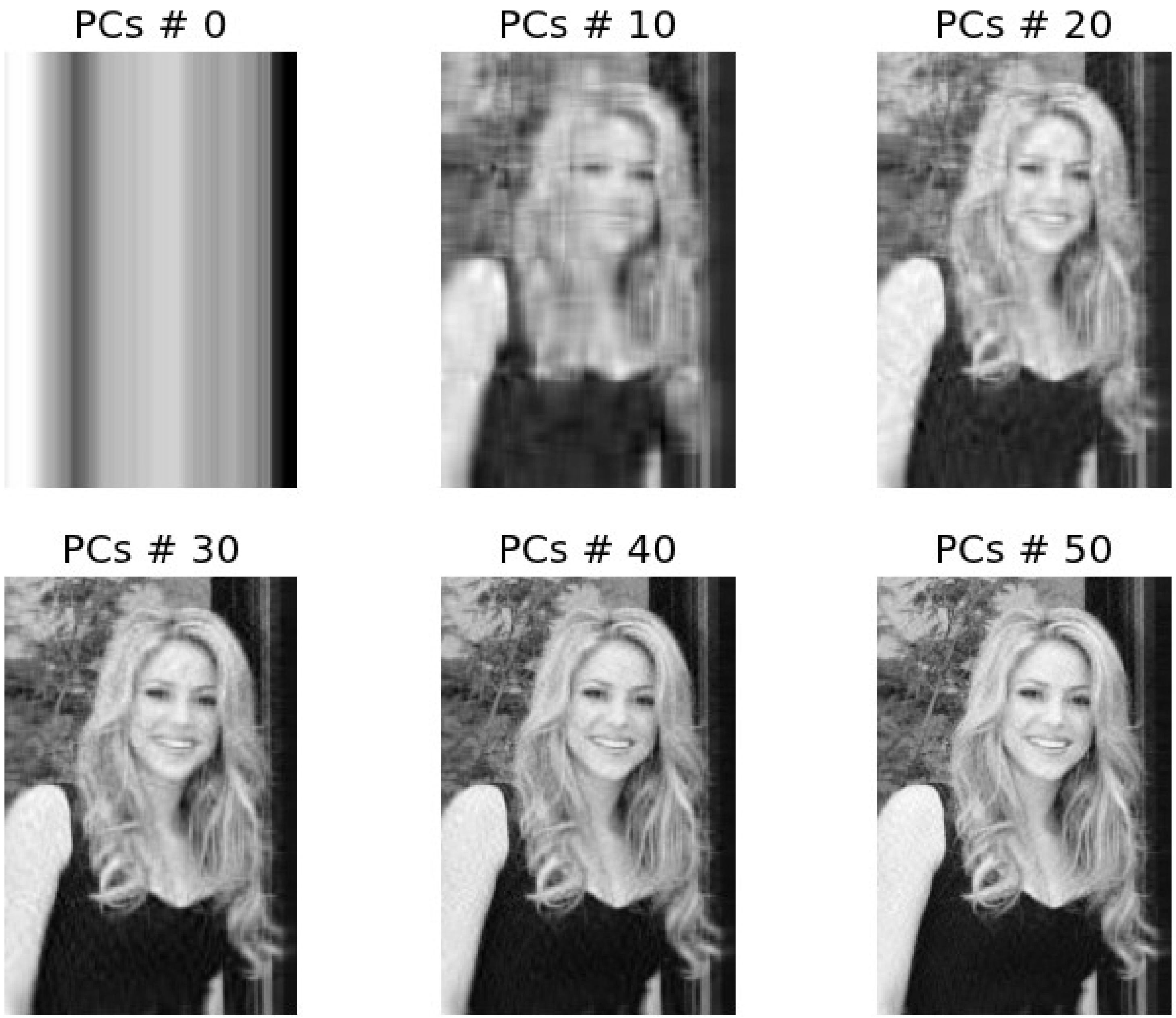
Neat application: Eigenfaces!



http://doi.ieeecomputersociety.org/cms/Computer.org/dl/trans/tp/2008/05/figures/ttp200805079619.gif



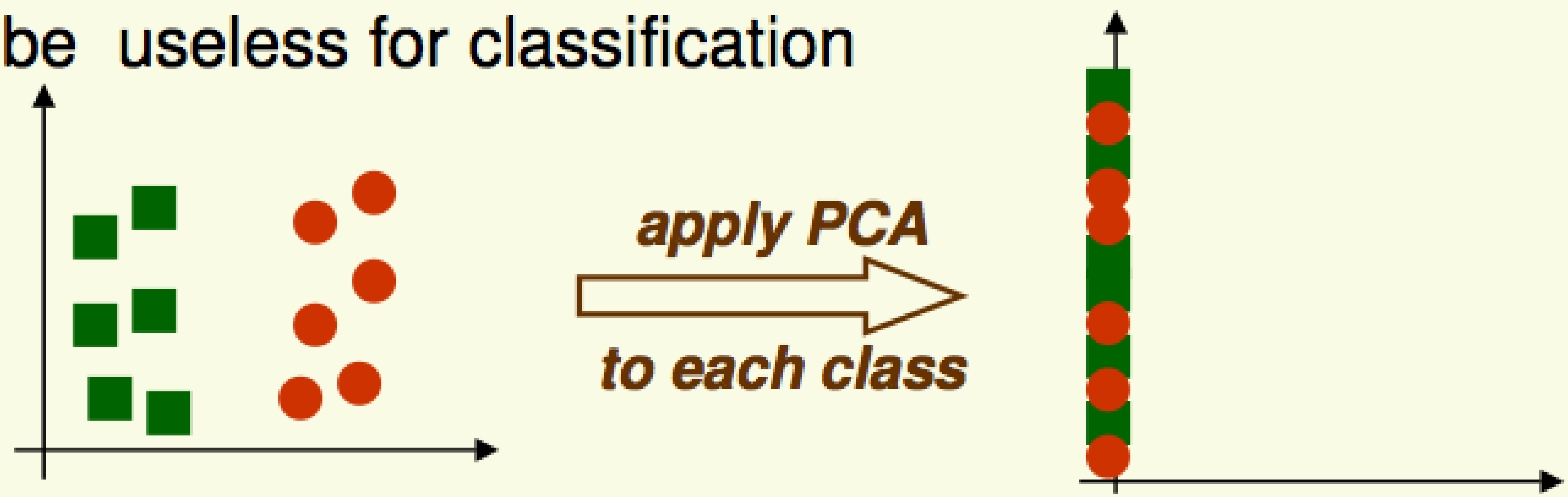
http://bytefish.de/static/images/blog/eigenfaces/subplot_eigenfaces.png



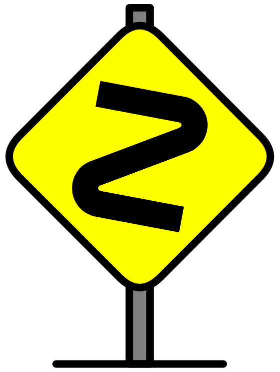
http://3.bp.blogspot.com/-eTWWNr-CyeI/Ti1l2h9ldnI/AAAAAAAAAE0/fUF2hl-hTMQ/s1600/pca_shakira.png

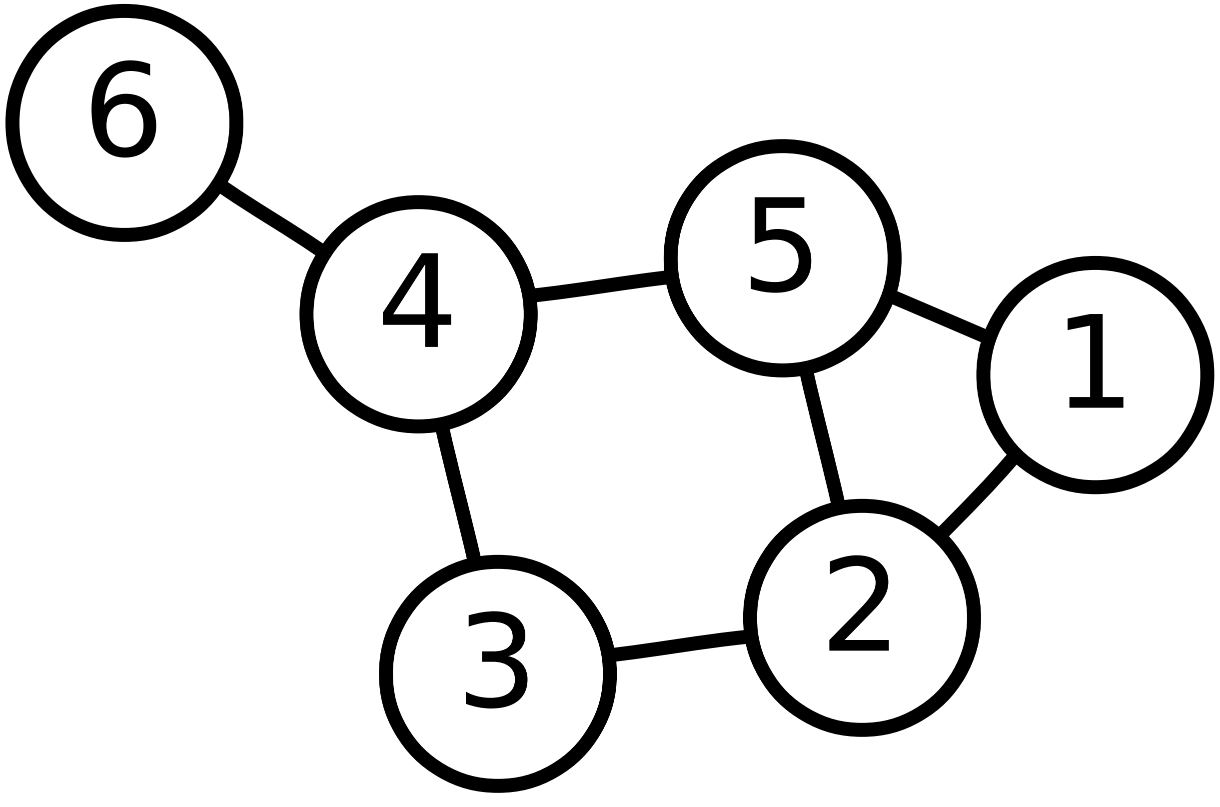
Drawbacks of PCA

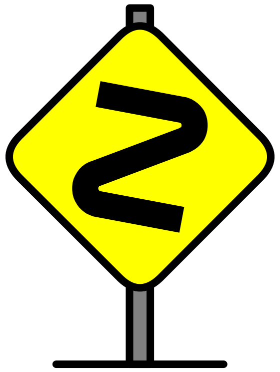
- PCA was designed for accurate *data representation*,' not for data classification
  - Preserves as much variance in data as possible
  - If directions of maximum variance is important for classification, will work
- However the directions of maximum variance may be useless for classification



Distance preserving: Johnson-Lindenstrauss







Do not worry about your difficulties in Mathematics. I can assure you mine are still greater.

- Albert Einstein

Read more at http://www.brainyquote.com/quotes/quotes/a/alberteins125370.html#65g20AeMDxm86kRt.99

Theorem: Johnson-Lindenstrauss

Given $0 < \epsilon < 1$, a set X of m points in $\mathbb{R}^N$, and a number $n > \frac{8\log(m)}{\epsilon^2}$, there is a linear map $f : \mathbb{R}^N \to \mathbb{R}^n$ such that

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2$$

for all $u, v \in X$

"Fear has big eyes" by Robbie Grubbs from Houston - What????. Licensed under Creative Commons Attribution-Share Alike 2.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Fear_has_big_eyes.jpg#mediaviewer/File:Fear_has_big_eyes.jpg

WPI