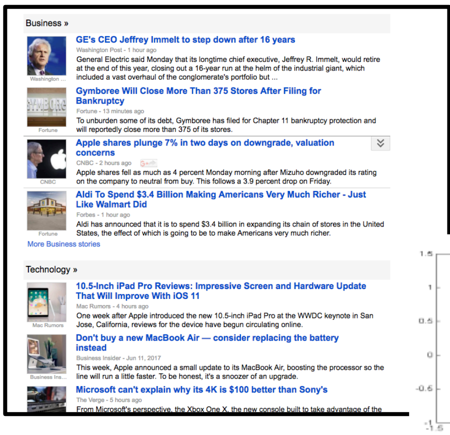
Unsupervised Learning

Stephan Robert

1

Unsupervised Learning

- Learning without the presence of an expert (e.g k-means, clustering approaches ,…)


Business »

GE's CEO Jeffrey Immelt to step down after 16 years
Washington Post - 1 hour ago
General Electric said Monday that its longtime chief executive, Jeffrey R. Immelt, would retire at the end of this year, closing out a 16-year run at the helm of the industrial giant, which included a vast overhaul of the conglomerate's portfolio but ...

Gymboree Will Close More Than 375 Stores After Filing for Bankruptcy
Fortune - 13 minutes ago
To unburden some of its debt, Gymboree has filed for Chapter 11 bankruptcy protection and will reportedly close more than 375 of its stores.

Apple shares plunge 7% in two days on downgrade, valuation concerns
CNBC - 2 hours ago
Apple shares fell as much as 4 percent Monday morning after Mizuho downgraded its rating on the company to neutral from buy. This follows a 3.9 percent drop on Friday.

Aldi To Spend $3.4 Billion Making Americans Very Much Richer - Just Like Walmart Did
Forbes - 1 hour ago
Aldi has announced that it is to spend $3.4 billion in expanding its chain of stores in the United States, the effect of which is going to be to make Americans very much richer.

More Business stories

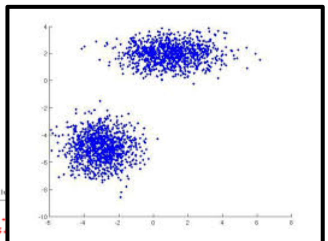
Technology »

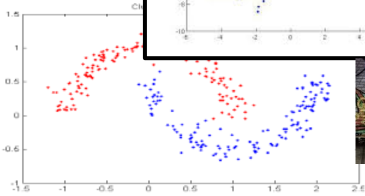
10.5-Inch iPad Pro Reviews: Impressive Screen and Hardware Update That Will Improve With iOS 11
Mac Rumors - 4 hours ago
One week after Apple introduced the new 10.5-inch iPad Pro at the WWDC keynote in San Jose, California, reviews for the device have begun circulating online.

Don't buy a new MacBook Air — consider replacing the battery instead
Business Insider - Jun 11, 2017
This week, Apple announced a small update to its MacBook Air, boosting the processor so the line will run a little faster. To be honest, it's a snoozer of an upgrade.

Microsoft can't explain why its 4K is $100 better than Sony's
The Verge - 5 hours ago
From Microsoft's perspective, the Xbox One X, the new console built to take advantage of the

news.google.com

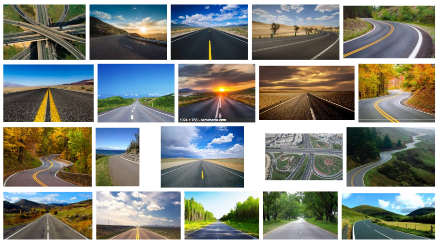


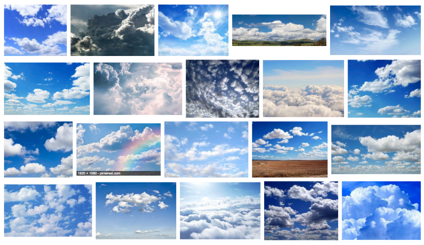


2

Unsupervised Learning

- Clustering images, group as:
  - Clouds
  - Mountains
  - Roads





- Users on Web sites (groups, activities)
- Words (First names, last names, location words, ...)

3

Unsupervised Learning

- Supervised Learning. Training set:

$$\{((x)_1,(y)_1),((x)_2,(y)_2),((x)_3,(y)_3),\ldots,((x)_m,(y)_m)\}$$

- Unsupervised Learning. Training set:

$$\{(x)_1,(x)_2,(x)_3,\ldots,(x)_m\}$$

4

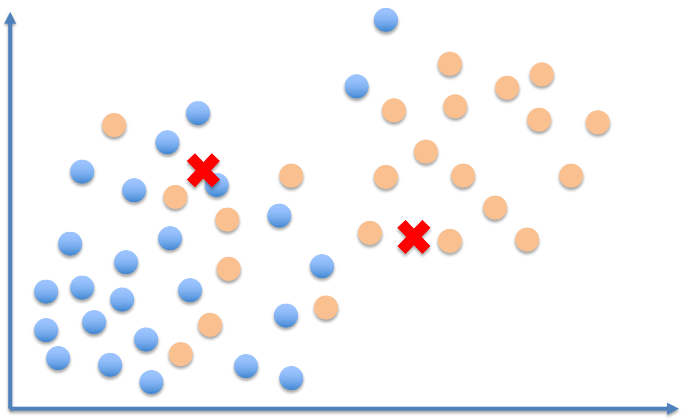
Algorithms

- K-Means
- Principal Component Analysis (PCA)

K-Means Algorithm

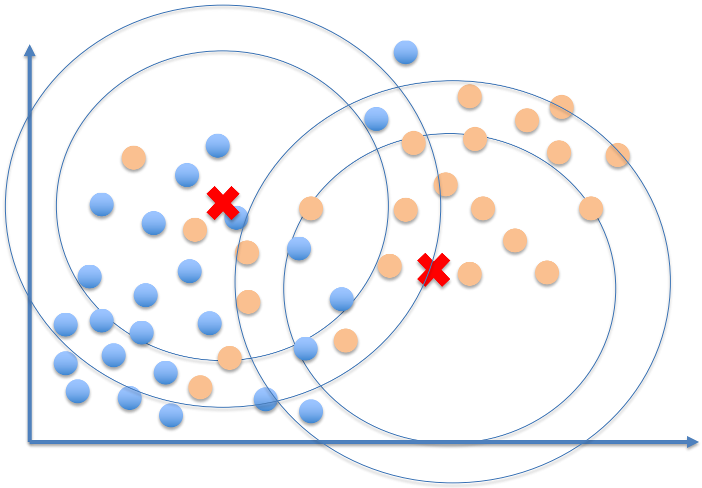
- Randomly chose K examples as initial centroids
- While true:
  - create k clusters by assigning each example to closest centroid
  - compute k new centroids by averaging examples in each cluster
  - if centroids don't change:
- Break

K-Means Algorithm
Random placement of the centroids



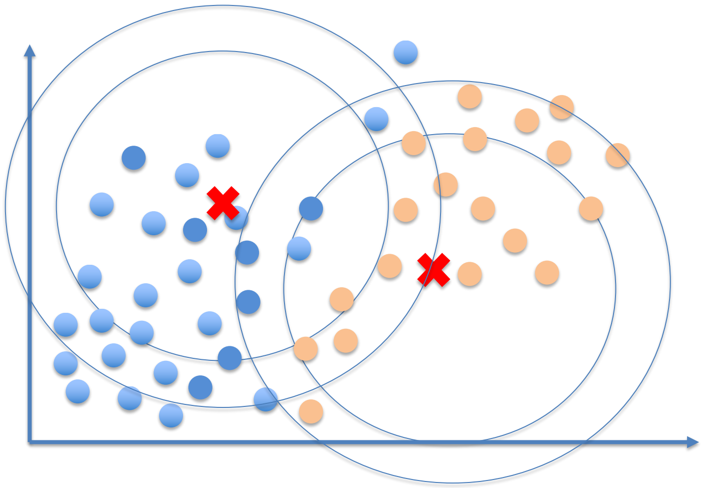
7

K-Means Algorithm: iteration 1



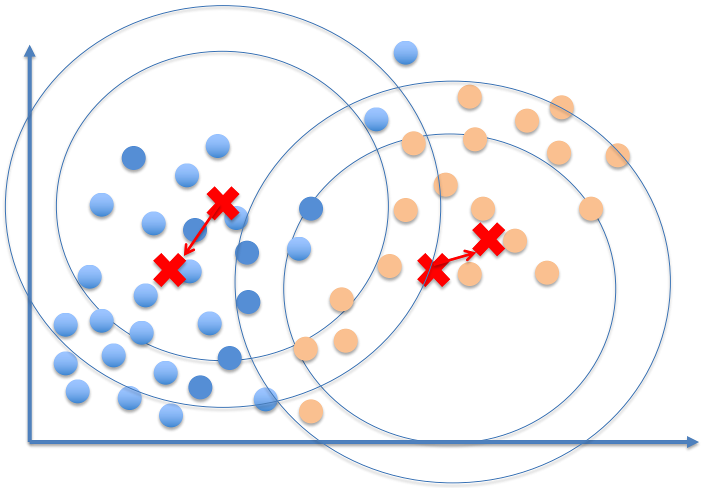
8

K-Means Algorithm: iteration 1



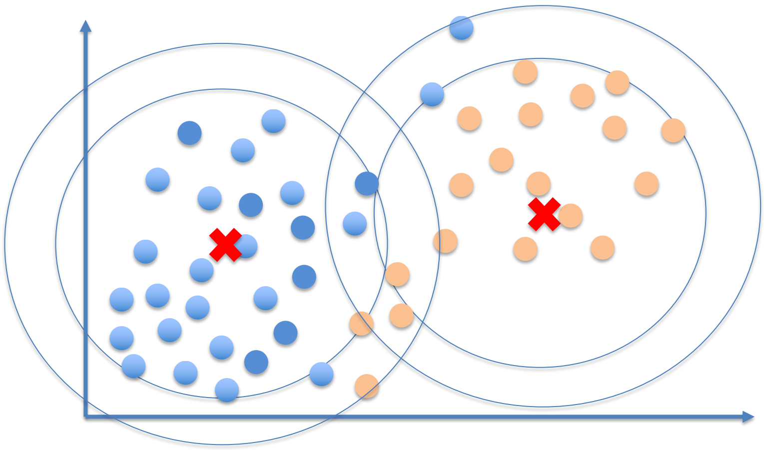
9

K-Means Algorithm: iteration 2

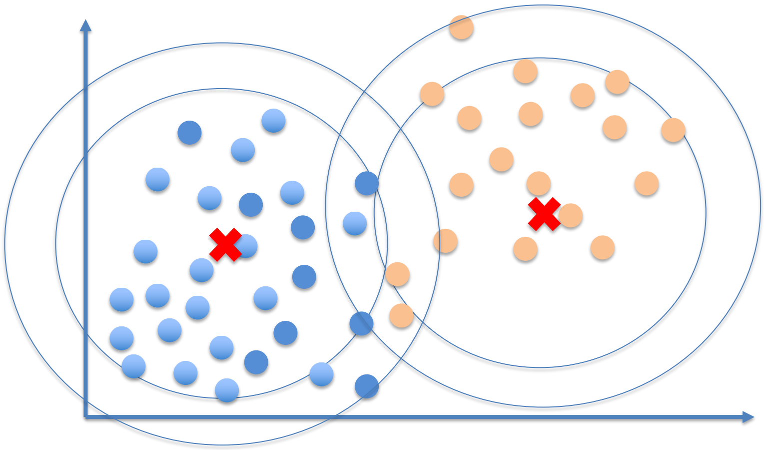


10

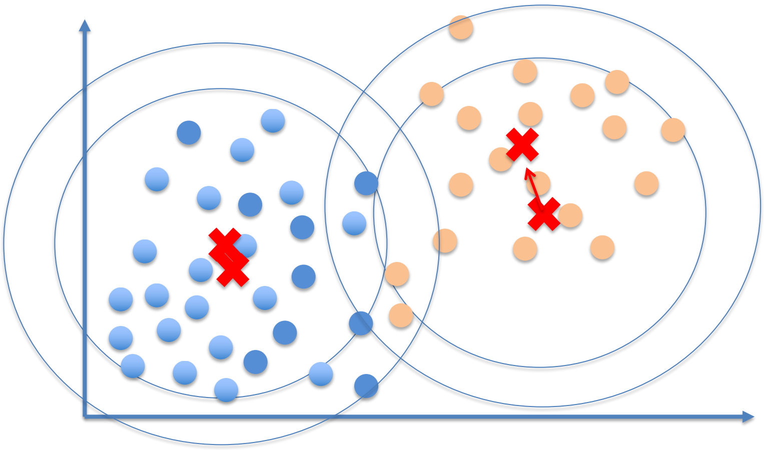
K-Means Algorithm: iteration 3



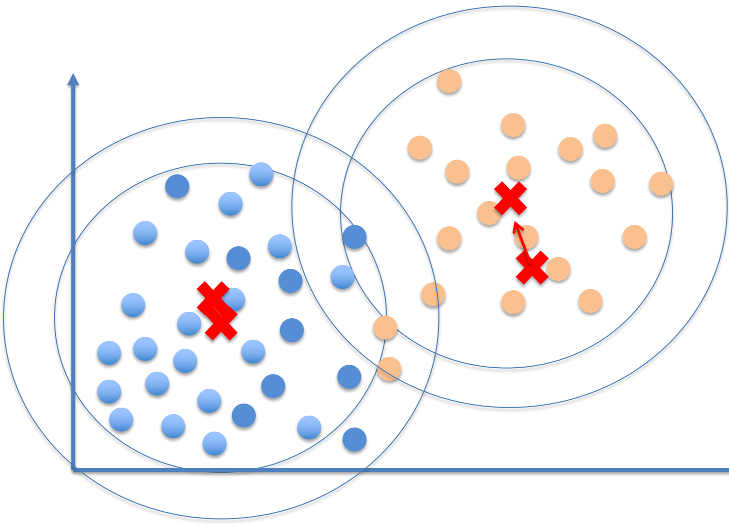
K-Means Algorithm: iteration 3



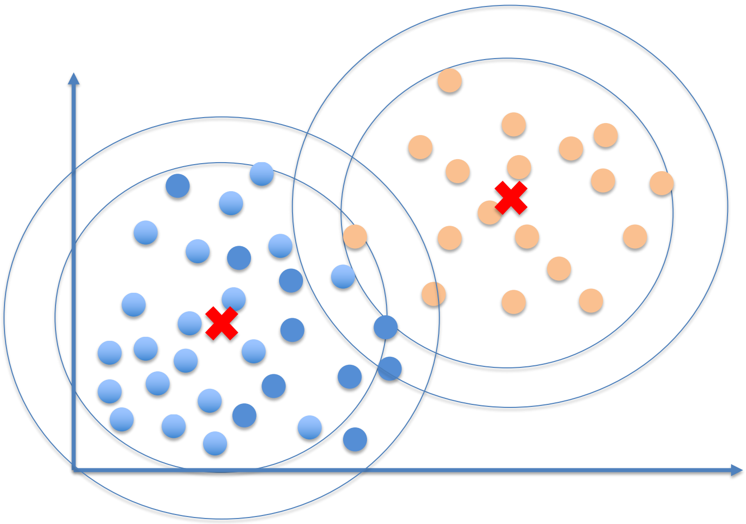
K-Means Algorithm: iteration 4



K-Means Algorithm: iteration 4

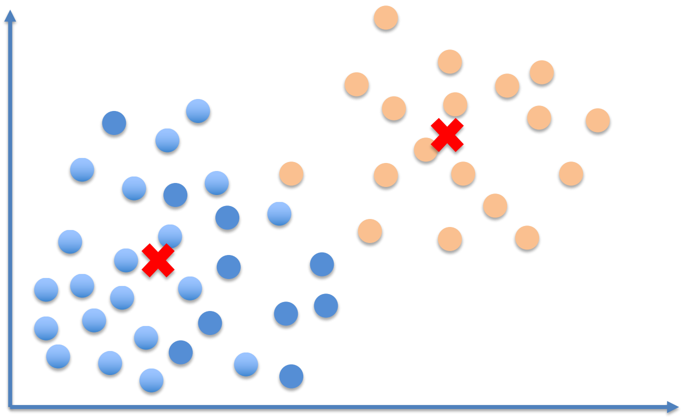


K-Means Algorithm: iteration 5



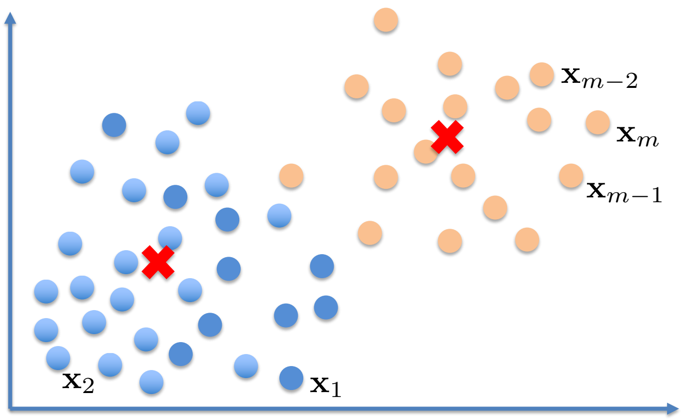
15

K-Means Algorithm: iteration 5



16

K-Means Algorithm: stop



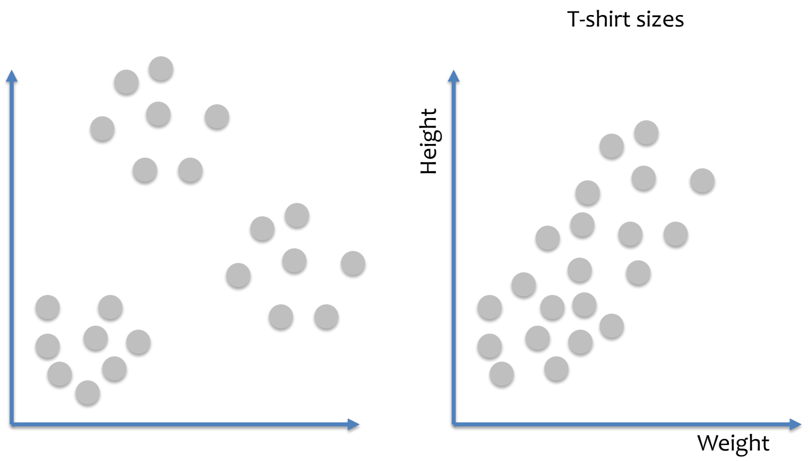
17

K-Means Algorithm

- randomly chose K examples as initial centroids
- for i=1 to m (number of vectors **x**):
  - Create K clusters by assigning each example to closest centroid. Label each **x** with $(c)_i \in \mathbb{N}^K$
- For k=1 to K
  - compute k new centroids by averaging examples in each cluster (l is the number of x associated to the centroid) $$\boldsymbol{\mu}_k = \frac{1}{l}\sum_{i=1}^{l} \mathbf{x}_i \in \mathbb{R}^n$$
- Repeat if centroids do change, else finished

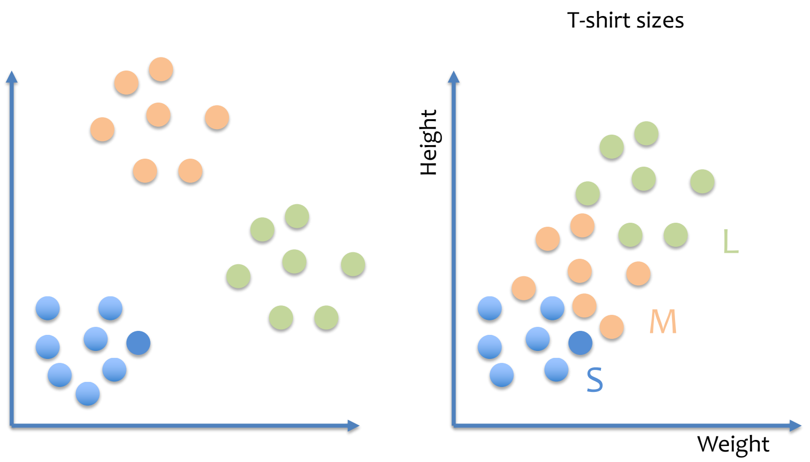
18

Clusters for non-separted clusters



19

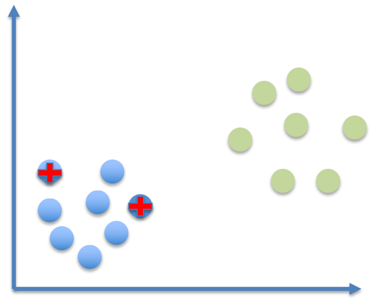
Clusters for non-separted clusters



20

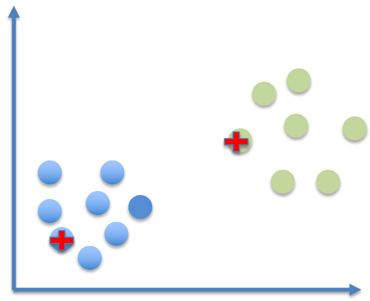
Random initialisation

- Should have K < m!
- Randomly pick K training examples
- Set $\mu_1, \mu_2, \ldots, \mu_k$ equal to these K examples



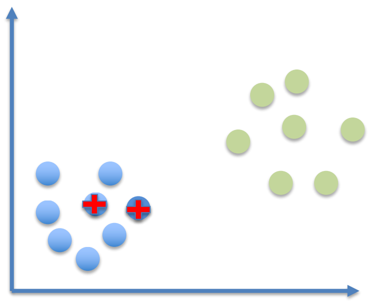
Random initialisation

- Should have K < m!
- Randomly pick K training examples
- Set $\mu_1, \mu_2, \ldots, \mu_k$ equal to these K examples



Random initialisation

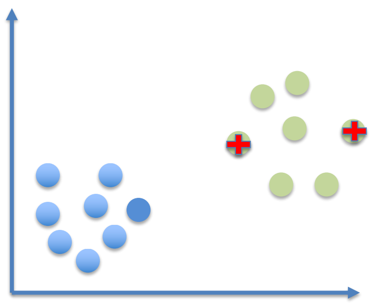
- Should have K < m!
- Randomly pick K training examples
- Set $\mu_1, \mu_2, \ldots, \mu_k$ equal to these K examples



23

Random initialisation

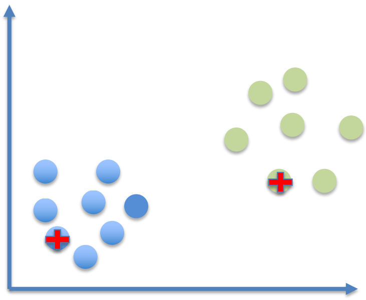
- Should have K < m!
- Randomly pick K training examples
- Set $\mu_1, \mu_2, \ldots, \mu_k$ equal to these K examples



24

Random initialisation

- Should have K < m!
- Randomly pick K training examples
- Set $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k$ equal to these K examples



25

Choosing the number of clusters

- Cost function:

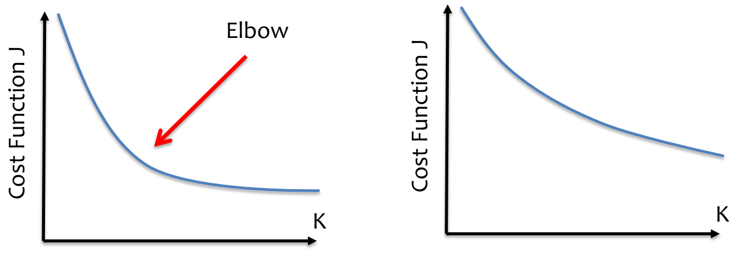
$$J = \frac{1}{m} \sum_{i=1}^{m} ||\mathbf{x}_i - \boldsymbol{\mu}_{(c)_i}||$$

$(c)_i$ is the index of the cluster to which $\mathbf{x}_i$ is assigned, $= 1, 2, \ldots, K$

$$\min_{\substack{(c)_1,(c)_2,\ldots,(c)_m \\ \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_K}} J$$

26

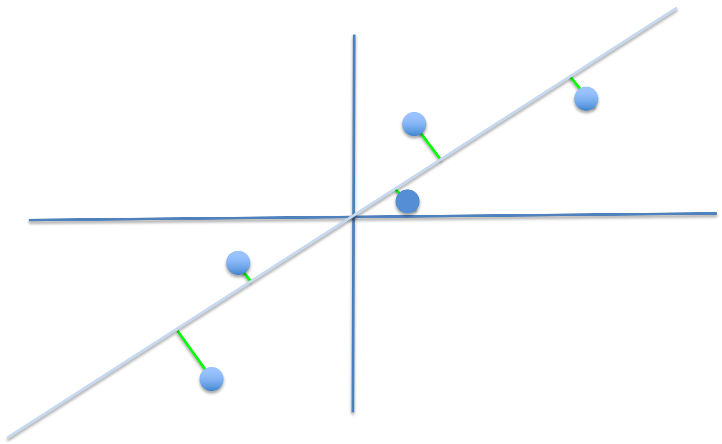
Choosing the number of clusters



27

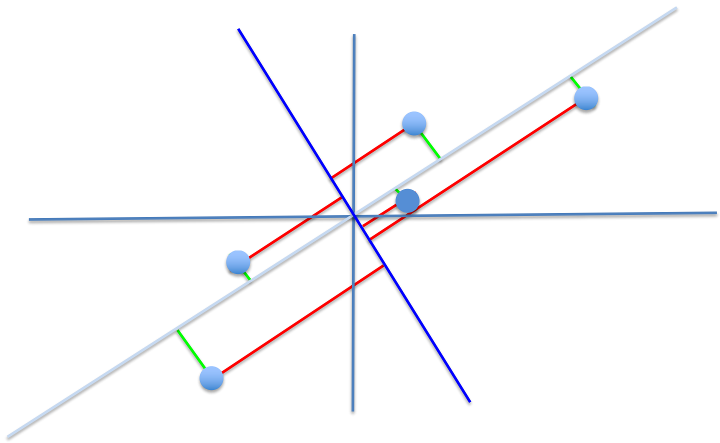
Principal Component Analysis (PCA)

Problem Formulation



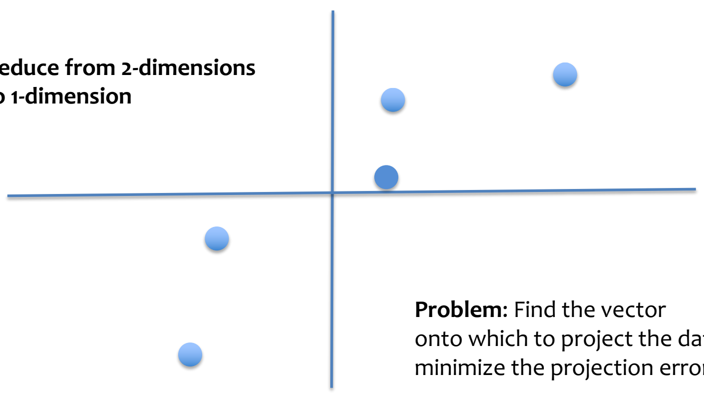
28

Principal Component Analysis (PCA)
Problem Formulation



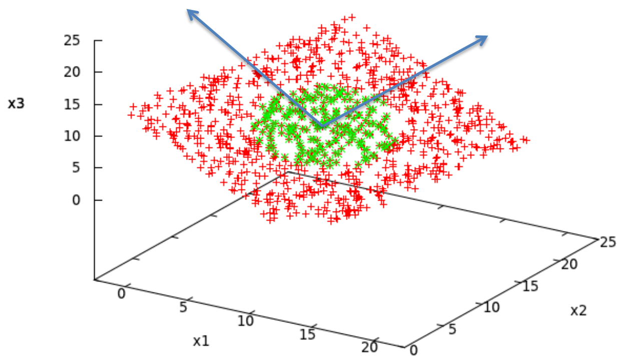
Principal Component Analysis (PCA)
Problem Formulation

Reduce from 2-dimensions to 1-dimension



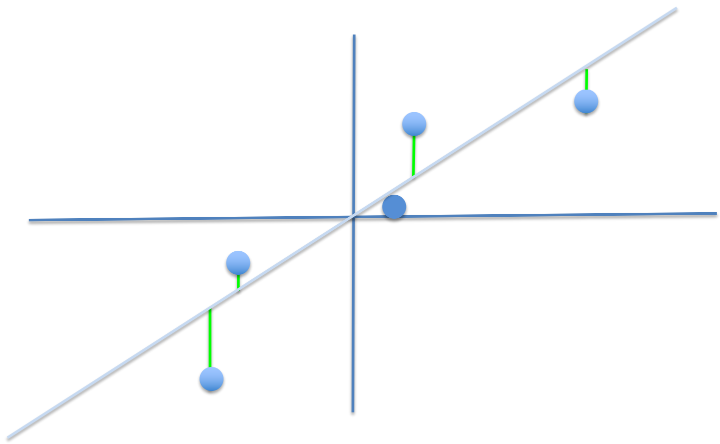
Problem: Find the vector onto which to project the data to minimize the projection error

PCA: 3D -> 2D



31

PCA: **Not** a linear regression



32

PCA: data preprocessing

- Training set: $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$
- Preprocessing:

$$\mu_0 = \frac{1}{m}\sum_{i=1}^{m}(x_0)_i$$
$$\mu_1 = \frac{1}{m}\sum_{i=1}^{m}(x_1)_i$$
$$\vdots$$

- Replace each $(x_j)_i$ with $(x_j)_i - \mu_j$
- Eventually scaling: replace each $(x_j)_i$ with

$$\frac{(x_j)_i - \mu_j}{\sigma_j}$$

33

PCA Algorithm

- Reduce data from *n*-dimension to *k*-dimension
- Covariance matrix:

$$(\mathbf{x})_i = \mathbf{x}_i = \begin{pmatrix} (x_0)_i \\ (x_1)_i \\ \vdots \\ (x_n)_i \end{pmatrix}$$

$$\mathbf{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i\mathbf{x}_i^t) =$$

$$\begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m}(x_0)_i(x_0)_i & \frac{1}{m}\sum_{i=1}^{m}(x_0)_i(x_1)_i & \ldots & \frac{1}{m}\sum_{i=1}^{m}(x_0)_i(x_n)_i \\ \frac{1}{m}\sum_{i=1}^{m}(x_1)_i(x_0)_i & \frac{1}{m}\sum_{i=1}^{m}(x_1)_i(x_1)_i & \ldots & \frac{1}{m}\sum_{i=1}^{m}(x_1)_i(x_n)_i \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{m}\sum_{i=1}^{m}(x_n)_i(x_0)_i & \frac{1}{m}\sum_{i=1}^{m}(x_n)_i(x_1)_i & \ldots & \frac{1}{m}\sum_{i=1}^{m}(x_n)_i(x_n)_i \end{pmatrix}$$

- Compute the eigenvectors of $\mathbf{\Sigma}$

$$\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$$

34

PCA Algorithm

- Eigenvectors: $\mathbf{u}_i, i = 1, \ldots, n$

$$\mathbf{U} = [\mathbf{u_1}\ \mathbf{u_2} \ldots\ \mathbf{u_n}]$$

- We construct a matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ with the eigenvectors and another one with the k first eigenvectors: $\mathbf{U}_{reduce} \in \mathbb{R}^{n \times k}$

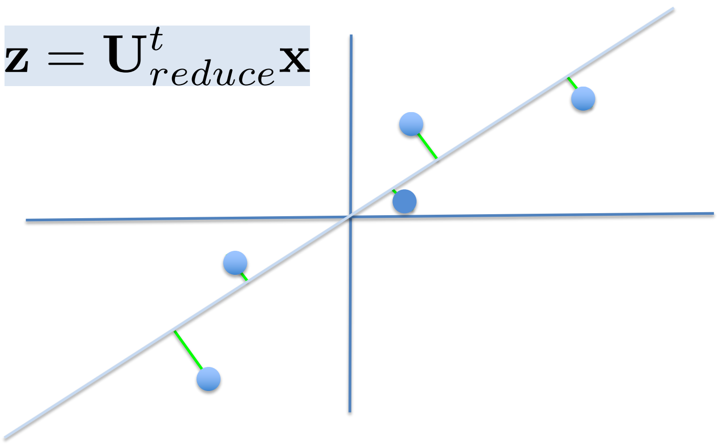
$$\mathbf{U}_{reduce} = [\mathbf{u_1}\ \mathbf{u_2} \ldots\ \mathbf{u_k}]$$

- Notice: Normalize and scale eventually! Every feature should be 0-mean.

35

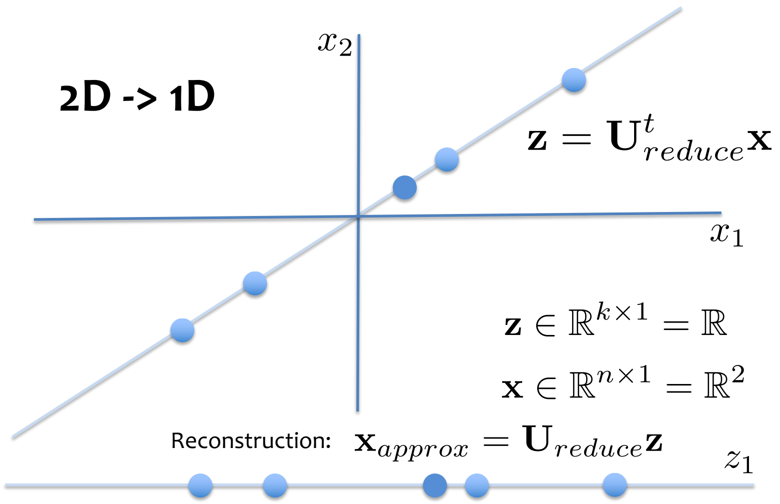
Principal Component Analysis (PCA)

$$\mathbf{z} = \mathbf{U}^t_{reduce}\mathbf{x}$$



36

Principal Component Analysis (PCA)



37

How many principal components?

- Average squared projection error:

$$\frac{1}{m}\sum_{i=1}^{m} ||(\mathbf{x})_i - (\mathbf{x}_{approx})_i||^2$$

- Total variation in the data:

$$\frac{1}{m}\sum_{i=1}^{m} ||(\mathbf{x})_i||^2$$

- Choose k to be the smallest value so that

$$\frac{\frac{1}{m}\sum_{i=1}^{m} ||(\mathbf{x})_i - (\mathbf{x}_{approx})_i||^2}{\frac{1}{m}\sum_{i=1}^{m} ||(\mathbf{x})_i||^2} \leq 0.01$$

38

Number of principal components

- Algorithm
  - Try PCA for $k = 1, 2, 3, 4, \ldots$
  - Compute

$$\mathbf{U}_{reduce}, (\mathbf{z})_1, (\mathbf{z})_2, \ldots, (\mathbf{z})_m,$$
$$(\mathbf{x}_{approx})_1, (\mathbf{x}_{approx})_2, \ldots, (\mathbf{x}_{approx})_m$$

  - Check if

$$\frac{\frac{1}{m}\sum_{i=1}^{m}||(\mathbf{x})_i - (\mathbf{x}_{approx})_i||^2}{\frac{1}{m}\sum_{i=1}^{m}||(\mathbf{x})_i||^2} \leq 0.01$$

39