Machine Learning

12. week

- Feature Extraction-Selection
- Dimension reduction
- PCA – LDA

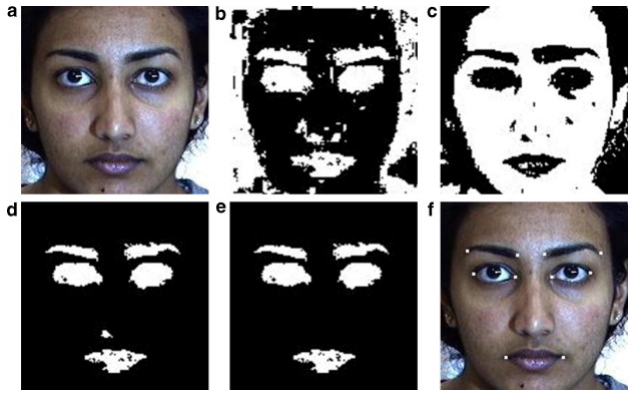
Feature Extraction

- Any problem can be solved by machine learning methods in case of that the system must be appropriately represented.
- The problem to be solved may not always have the possibility to be directly attributed to machine learning methods.
- In particular, it is necessary to extract appropriate features from data such as time series and images.

Feature Extraction

- The science of dealing with images and videos is called «image processing», and the science focused on the time series is named as «signal processing».
- The aim of studies in «pattern recognition» field is to extract features .

Feature Extraction



Feature Extraction

- The general approach of the doctors base on the following ECG (ElectroCardioGram) is done by calculating the heart rate and the duration of each special waves in the ECG signal.

5

Feature Selection

University of California Machine Learning Repository

Name	Data Types	Default Task	Attribute Types	Instances	Attributes	Year
URL Reputation	Multivariate, Time-Series	Classification	Integer, Real	2396130	**3231961**	2009
Gas sensor arrays in open sampling settings	Multivariate, Time-Series	Classification	Real	18000	**1950000**	2013
YouTube Multiview Video Games Dataset	Multivariate, Text	Classification, Clustering	Integer, Real	120000	**1000000**	2013
Twin gas sensor arrays	Multivariate, Time-Series, Domain-Theory	Classification, Regression	Real	640	**480000**	2016
Gas sensor array exposed to turbulent gas mixtures	Multivariate, Time-Series	Classification, Regression	Real	180	**150000**	2014
ElectricityLoadDiagrams	Time-Series	Regression, Clustering	Real	370	**140256**	2015
PEMS-SF	Multivariate, Time-Series	Classification	Real	440	**138672**	2011
Gas sensor array under flow modulation	Multivariate, Time-Series	Classification, Regression	Real	58	**120432**	2014
Bag of Words	Text	Clustering	Integer	8000000	**100000**	2008
Dorothea	Multivariate	Classification	Integer	1950	**100000**	2008
Farm Ads	Text	Classification		4143	**54877**	2011
Dexter	Multivariate	Classification	Integer	2600	**20000**	2008

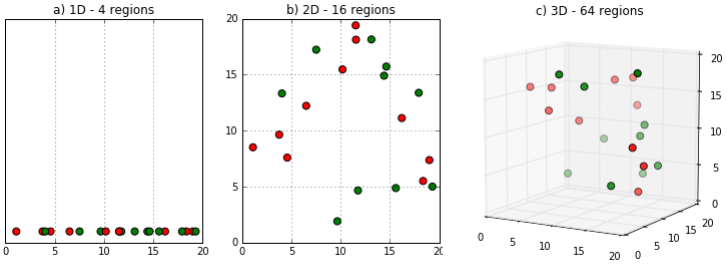
6

Feature Selection

- There are many feature extraction methods proposed in the literature. But the cost of machine learning methods increase with the amount of attributes exist in the dataset of the problem.
  - This is an undesired situation.
- Therefore, it is necessary to **use the minimum number of features** to represent the problem in the most appropriate way.

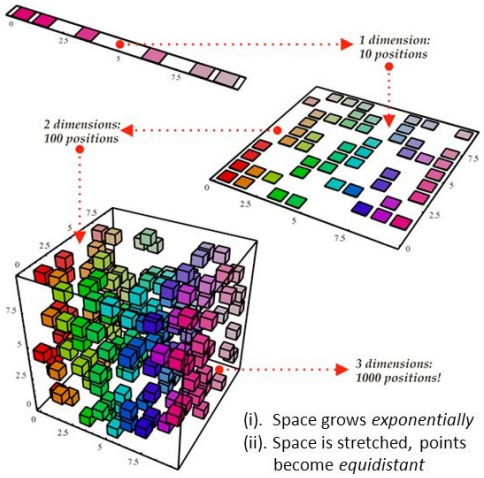
Feature Selection

- With feature selection machine learning algorithms will have shorter training times with reasonable resource needs.



Feature Selection

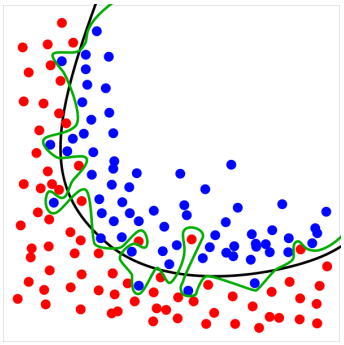
- With feature selection Curse of Dimensionality can be avoided.



(i). Space grows *exponentially*
(ii). Space is stretched, points become *equidistant*

Feature Selection

- With feature selection, we can reduce Overfitting



Feature Selection

- There are many studies on which of the known feature extracting methods should be applied for the problem.
- These studies aim to find out the most appropriate methods of feature extraction as soon as possible.
- For example, if we have a problem of classification with 5 extracted features and we want to find the most useful attribute, it will be necessary to make a classification over 31 sub-sets by trial and error and this work will take too long.

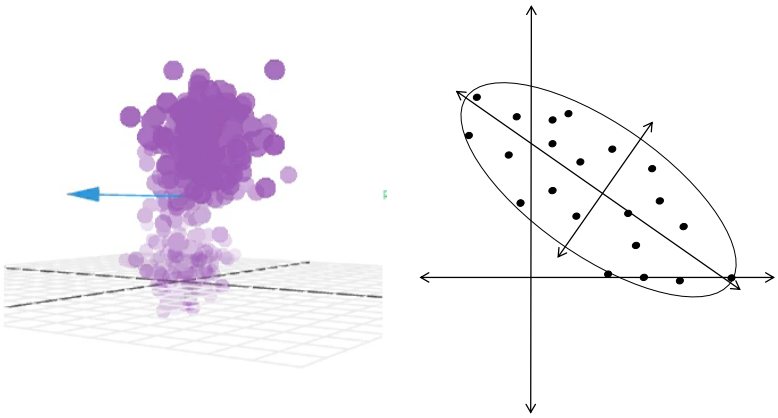
Dimension reduction

- Dimension reduction methods can also be used for feature selection.
- The main purpose of dimension reduction is to determine and remove the unnecessary attributes of the given data.
- The most known and used methods are PCA (Principle Components Analysis) and LDA (Linear Discriminant Analysis) methods.

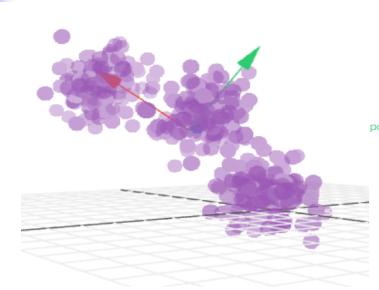
Principale Component Analysis - PCA

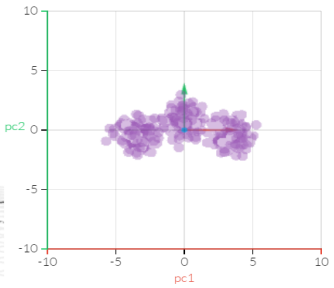
- In the PCA method, it is assumed that the covariance matrix representing the relation between the feature vectors of the data is equal to the eigen-vector and the eigen-value multiplication.
- The eigen-vectors found here are considered as new basic components and the new components of the data are calculated.
- If PCA is used for size reduction, the dimension with the smallest deviation is deleted and if necessary the data is recycled to its size.

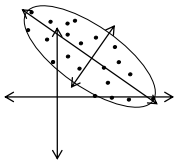
Principale Component Analysis - PCA



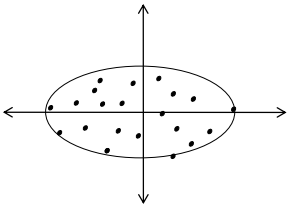
Principale Component Analysis - PCA







PCA 1st stage

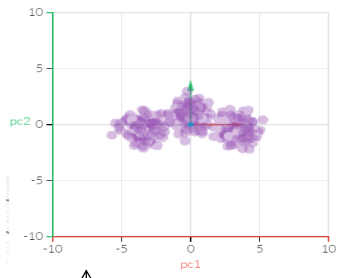


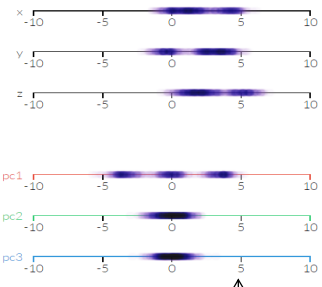
Umut ORHAN, PhD.

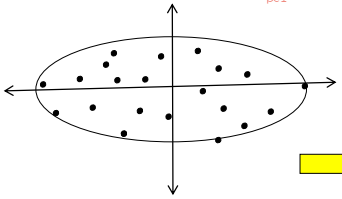
15

15

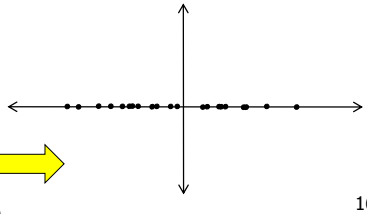
Principale Component Analysis - PCA







PCA 2nd stage

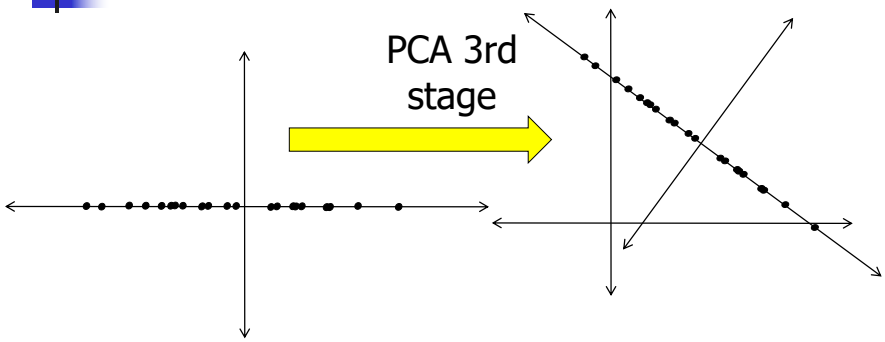


Umut ORHAN, PhD.

16

16

Principale Component Analysis - PCA

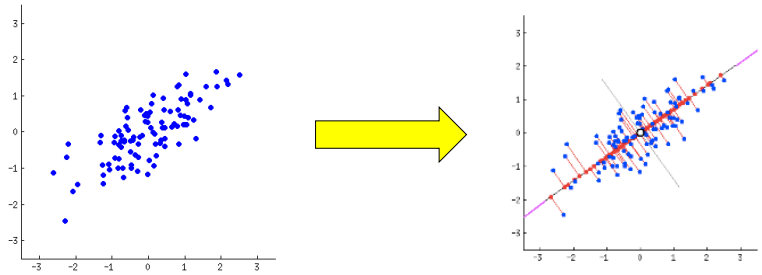


If you reduce the number of features in the target, which is targeted by size reduction, it is not recycled at this stage. It is enough to delete unnecessary features in basic components.

17

Principale Component Analysis - PCA

- Each straight line represents a "principal component," or a relationship between an independent and dependent variable. While there are as many principal components as there are dimensions in the data, PCA's role is to prioritize them.



18

Principale Component Analysis - PCA

- Variance is simply standard deviation squared, and is often expressed as *s^2*.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$$

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{(n-1)}$$

Principale Component Analysis - PCA

- Also, in the equation below, you'll notice that there is only a small difference between covariance and variance.

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

vs.

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{(n-1)}$$

Principale Component Analysis - PCA

- Imagine that we compose a square matrix of numbers that describe the variance of the data, and the covariance among variables. This is the *covariance matrix*. It is an empirical description of data we observe.
- Finding the eigenvectors and eigenvalues of the covariance matrix is the equivalent of fitting those straight, principal-component lines to the variance of the data.

MATLAB Application

```
>edit PCA_ornek.m
```

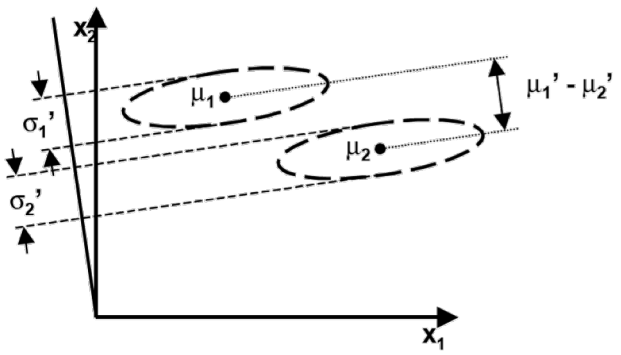
- The basic components analysis experiment is done by loading different prepared datasets. Different examples should be made on Matlab using these codes.

Linear Discriminant Analysis - LDA

- LDA (Linear Discriminant Analysis) is a method that reduces the size by maximizing the linear separability of groups belonging to different classes in the data.
- The variance within each group is kept at a minimum and the averages of the groups are maximized from each other.

Linear Discriminant Analysis - LDA

$$\max \frac{\mu_1' - \mu_2'}{\sigma_1'^2 - \sigma_2'^2}$$

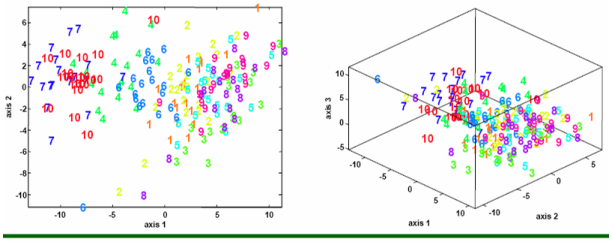


Linear Discriminant Analysis - LDA

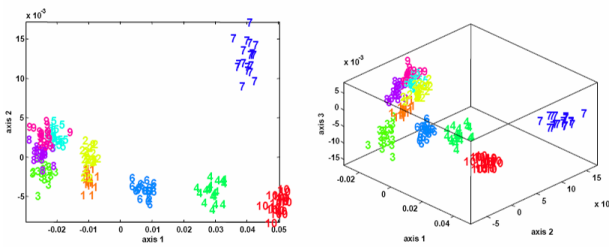
- Compute the d-dimensional mean vectors for the different classes from the dataset.
- Compute the scatter matrices (in-between-class and within-class scatter matrix).
- Compute the eigenvectors and corresponding eigenvalues for the scatter matrices.
- Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k dimensional matrix W.
- Use this d×k eigenvector matrix to transform the samples onto the new subspace.

PCA ve LDA

PCA



LDA



Presentation Task

Search an ICA (Independent Component Analysis) application used for size reduction method.