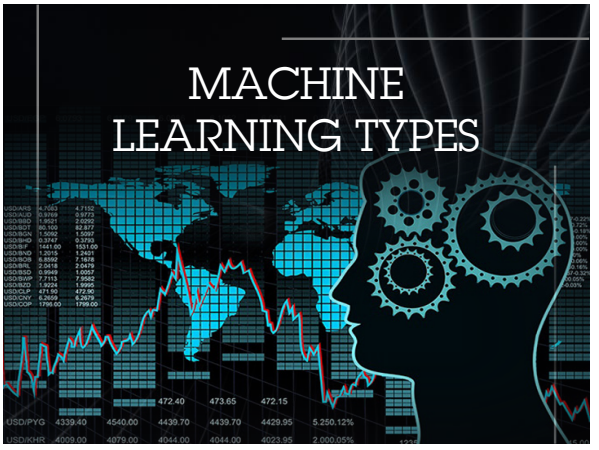

## Outline

- Machine Learning types
- Unsupervised learning:
  - Principal Components Analysis




## What is Machine Learning ?

- Machine learning is the integration of different techniques
- Its goal is to extract value out of data

## Machine Learning Classes

- Supervised Learning
- Unsupervised Learning
- Deep Learning

## Supervised Learning

Given $(x_i, y_i), i = 1,\ldots,n$, learn a function $f : X \rightarrow Y$

- Where X is the vector of features
- Y is the vector of labels
- Categorical Y: classification
- Continuous Y: regression

$$Y = f(X) + \varepsilon$$

## Supervised ML methods

- PARAMETRIC METHODS
  - Define the functional form of the function f(X) & fit the model

- NON PARAMETRIC METHODS
  - Find the best fitting function f(X)
- MODEL EVALUATION (MSE)

## Supervised ML and Input Data

- Quantitative data: Regression problems
- Qualitative (categorical) data: Classification problems
- But… not a crisp classification

## Unsupervised Learning

- Given only $(x_i)$, $i = 1, \ldots, n$, can we infer the underlying structure of X ?
- Used for exploratory data analysis

- Limits of UML
  - No single goal
  - Subjective
  - No way of cross-validating

## Review questions

- Which are the different categories of Machine Learning?
- What is the difference between Unsupervised and Supervised ML?

## Review questions

- Which sort of input data are generally needed by Regression problems?
- And Classification ones?

## Review questions

- Do you need Labels for Unsupervised ML ?
- What is the typical usage of Unsupervised ML ?

## UNSUPERVISED LEARNING

Principal Components Analysis

## PCA goal

- PCA derives a low-dimensional set of features from a large set of variables (p) given (n) observations

- The idea is that each of the n-observations lives in the p-dimensional space but not all of these dimensions are equally interesting

## PCA background

- It reduces the dimension of a n x p data matrix X by:
  - Identifying directions in feature space along which original data are highly variable

- It finds a low-dimensional representation of a data set that contains as much as possible of the variation

## First principal component

- First PC is the normalized linear combination of the features that has the largest variance

$$Z_1=\phi_{11} X_{1+}\phi_{21} X_{2+}+ \ldots+\phi_{p1} X_p$$

- Normalized because:

$$\sum_{j=1}^{p} \phi^2_{j1} = 1$$

- Each of the X variables have to be standardized

- First PC vector:

$$\phi_1 = (\phi_{11}, \phi_{21},\ldots,\phi_{p1})^T$$

$$\text{Max} \left[\frac{\sum_{i=1}^{n}(\phi_1^T \cdot X_i)^2}{n}\right] \text{ subject to } \sum_{j=1}^{p} \phi^2_{j1} = 1$$

$$Z_i = \phi_1^T \cdot X_i$$

$$\text{Max} \left[\frac{\sum_{i=1}^{n}(Z_i)^2}{n}\right] \text{ subject to } \sum_{j=1}^{p} \phi^2_{j1} = 1$$

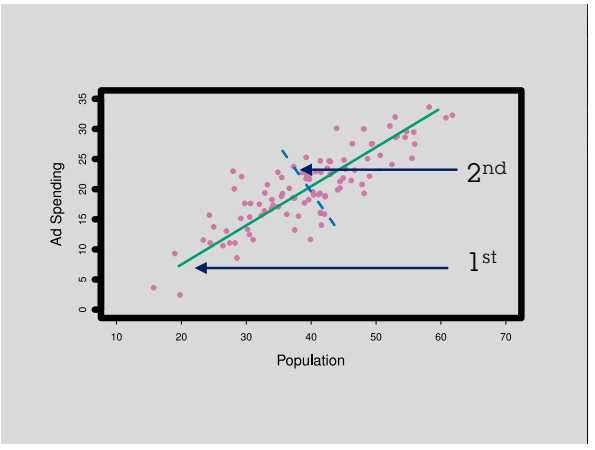
- We are maximizing the sample variance of the values of $Z_i$
- Solved via eigen decomposition

## $N^{th}$ principal component

- $2^{nd}$ PC is the linear combination of $X_i$ which has the maximal variance out of all linear combinations that are uncorrelated with $Z_1$

- $N^{th}$ PC has maximal variance out of all linear combinations that are uncorrelated with $Z_1 \dots Z_{n-1}$

- $1^{st}$ PC is the dimension along which the data varies the most
- It defines the line that is closest to all n observations




## PCA – Other interpretation

- $1^{st}$ PC is the line in p-dimensional space that is closest to the n observations (euclidean distance)

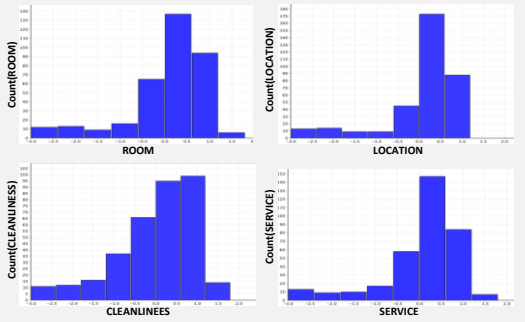
- The first 2 PC span the plane that is closest to the n observations
- The m first PC span the m-dim hyperplane that is closest to the n observations

## PCA – Example

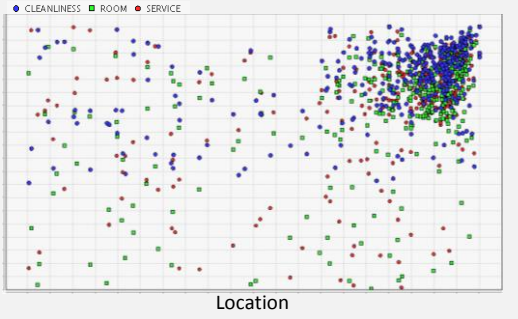
- Input data:
  - 360 hotel review scores in NYC, taken by scraping Tripadvisor data
  - 4 features considered:

CLEANLINESS, LOCATION, ROOM, SERVICE

Example Set: Normalized Statistics




Example set: Scatter

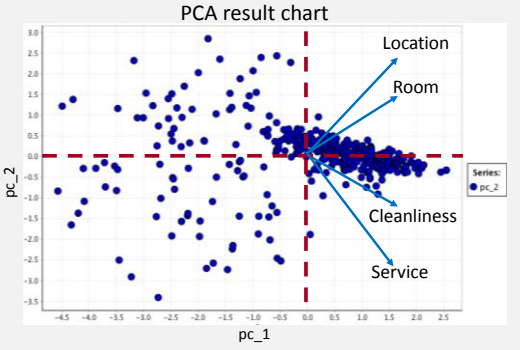



## PCA results

- Results will be 3 vectors with 360 scores each
- i.e. p=4, n=360, m=3
- Eigenvalues:

| | PC1 | PC2 | PC3 |
|---|---|---|---|
| CLEANLINESS | 0.501 | -0.336 | 0.746 |
| ROOM | 0.510 | 0.358 | -0.429 |
| SERVICE | 0.490 | -0.635 | -0.483 |
| LOCATION | 0.499 | 0.596 | 0.163 |

PCA result chart

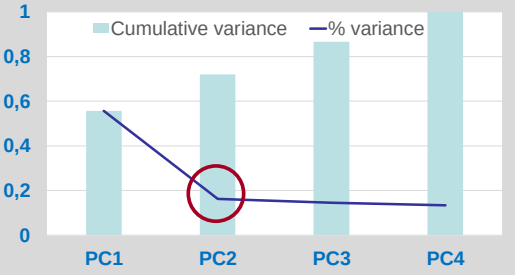



## Deciding how many PC

We decide how many PC based on the smallest number of PC required to explain a sizable amount of variation in the data

Use a "scree-plot", showing the proportion of variance explained by each PC

PCA result variance



## Review questions

- What is the purpose of PCA?
- Why is it useful for many ML problems?
- Why do you think it is necessary to standardize data before doing PCA?

## Review questions

- Provide a geometrical explanation of PCA
- Explain the "elbow" technique for choosing the number of PC to be used

## Review questions

- Using the NYC hotels example, explain why Cleanliness is not as important as Room and Location to Tripadvisor reviewers

## SUMMARY QUESTIONS

- Which are the different categories of Machine Learning?
- What is the difference between Unsupervised and Supervised ML?
- Which sort of input data are generally needed by Regression problems?

- And Classification ones?
- Do you need Labels for Unsupervised ML?
- What is the typical usage of Unsupervised ML?

- What is the purpose of PCA?
- Why is it useful for many ML problems?
- Why do you think it is necessary to standardize data before doing PCA?

- Provide a geometrical explanation of PCA
- Explain the "elbow" technique for choosing the number of PC to be used

- Using the NYC hotels example, explain why Cleanliness is not as important as Room and Location to Tripadvisor reviewers