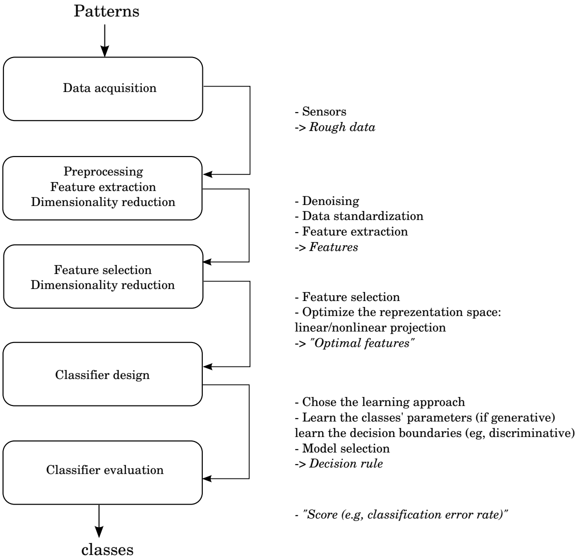
M2 Statistics & Data Science

Advanced Statistics & Machine Learning

Faicel Chamroukhi
Professeur
https://chamroukhi.com/

Overview

Pattern recognition system



Data analysis process

- **Data acquisition :** the sensors part in which the rough data (e.g, speech signals, images ...) are acquired (e.g., measured) through dedicated sensors.

Data analysis process

- **Data acquisition :** the sensors part in which the rough data (e.g, speech signals, images ...) are acquired (e.g., measured) through dedicated sensors.
- **Data Preprocessing :** *denoising*, *standardization*,... **feature extraction** for data *representation*.

Data analysis process

- **Data acquisition :** the sensors part in which the rough data (e.g, speech signals, images ...) are acquired (e.g., measured) through dedicated sensors.
- **Data Preprocessing :** *denoising*, *standardization*,... **feature extraction** for data *representation*.
- **Feature selection :** optimization of the representation space by applying for example, *linear* or *nonlinear dimensionality reduction* techniques, in a *supervised* or *unsupervised* context.

Data analysis process

- **Data acquisition :** the sensors part in which the rough data (e.g, speech signals, images ...) are acquired (e.g., measured) through dedicated sensors.
- **Data Preprocessing :** *denoising*, *standardization*,... **feature extraction** for data *representation*.
- **Feature selection :** optimization of the representation space by applying for example, *linear* or *nonlinear dimensionality reduction* techniques, in a *supervised* or *unsupervised* context.
- **Classifier design :** Given a training set of (selected) features (observations) $\rightarrow$ design of a *decision rule* with respect to a chosen *optimality criterion*

Data analysis process

- **Classifier evaluation :** Once the classifier is designed, its performance has to be assesses for example by computing the *classification error rate* on new data examples, in order to evaluate its generalization capabilities.

Data analysis process

- **Classifier evaluation :** Once the classifier is designed, its performance has to be assesses for example by computing the *classification error rate* on new data examples, in order to evaluate its generalization capabilities.

- these stages are not independent. They are highly interrelated and, depending on the results, one may go back to redesign earlier stages in order to improve the overall performance.

- There are some methods that combine stages, for example, a statistical learning at two stages (learning for feature extraction and learning for classification).

- Some stages can also be removed, for example when the classifier is directly built without feature extraction nor feature selection.

Data representation

Speech signal representation :

- Linear Predictive Coding (LPC)
- Mel Frequency Cepstrum Coding (MFCC)
- ...

Image representation :

- Histogram
- Local Binary Patterns (LBP)
- ...

Data Classification

- We talk about ''direct" classification, e.g., K-NN

Data Classification

- We talk about ”direct" classification, e.g., K-NN
- Very often, the classification task involves a learning problem (e.g., Neural Networks, Support Vector Machines, Gaussian Discriminant Analysis, Gaussian Mixtures, Hidden Markov Models,...)

  $\Rightarrow$ we learn a decision rule, or a functional mapping between possibly labeled features and the considered classes.

Data Classification

- We talk about "direct" classification, e.g., K-NN
- Very often, the classification task involves a learning problem (e.g., Neural Networks, Support Vector Machines, Gaussian Discriminant Analysis, Gaussian Mixtures, Hidden Markov Models,...)

  $\Rightarrow$ we learn a decision rule, or a functional mapping between possibly labeled features and the considered classes.
- The main questions concerning the classifier design : the type of the approach (generative, discriminative,..), linear/non-linear separation, and the type of the criterion.

Data Classification

- We talk about "direct" classification, e.g., K-NN
- Very often, the classification task involves a learning problem (e.g., Neural Networks, Support Vector Machines, Gaussian Discriminant Analysis, Gaussian Mixtures, Hidden Markov Models,...)

  $\Rightarrow$ we learn a decision rule, or a functional mapping between possibly labeled features and the considered classes.
- The main questions concerning the classifier design : the type of the approach (generative, discriminative,..), linear/non-linear separation, and the type of the criterion.

  $\Rightarrow$ In this course, we focus on probabilistic classifiers in a *maximum likelihood estimation (MLE)* framework.
- Probabilistic approaches can easily address problems related to missing information and allows for the integration of prior knowledge (Bayesian approaches), such as experts information.

Machine learning context

- The paradigm for automatically (without human intervention) learning from raw data is known as *machine learning*
- acquisition of knowledge from rough data for analysis, interpretation, prediction.
- *automatically* extracting useful information, possibly unknown, from rough data :
  - features
  - simplified models
  - classes,...

Machine learning context

- Statistical learning = machine learning + Statistics
- distinguished by the fact that the data are assumed to be realizations of random variables $\Rightarrow$ define probability densities over the data $\Rightarrow$ statistical (probabilistic) models.
- take benefit from the asymptotic properties of the estimators, e.g., consistency (e.g., Maximum likelihood)
- To make accurate decisions and predictions for future data, there is an important need to understand the *processes generating the data*.

Machine learning context

- Statistical learning = machine learning + Statistics

- distinguished by the fact that the data are assumed to be realizations of random variables $\Rightarrow$ define probability densities over the data $\Rightarrow$ statistical (probabilistic) models.

- take benefit from the asymptotic properties of the estimators, e.g., consistency (e.g., Maximum likelihood)
- To make accurate decisions and predictions for future data, there is an important need to understand the *processes generating the data*.
- $\Rightarrow$ This therefore leads us to *generative* learning

Machine learning context

Supervised/Unsupervised :

- *Supervised learning* : both the input (observation) and the output (target : class in classification) are available

- The objective is to predict the class of new data given predefined learned classes : *classification (discrimination)* problem

Machine learning context

Supervised/Unsupervised :

- *Supervised learning* : both the input (observation) and the output (target : class in classification) are available
- The objective is to predict the class of new data given predefined learned classes : *classification (discrimination)* problem
- In several application domains, we are confronted with the problem of missing information (class label missing, unknown, hidden).
- $\Rightarrow$ an *unsupervised* learning problem
- The objective is to discover possible classes (exploratory analysis)
- main models : latent data models : e.g., mixture models, HMMs
  $\Rightarrow$ a *clustering/segmentation* problem.

Machine learning context

Static/Dynamic :
Two contexts for the classification problem :

- Static context : The classification rules are taken from *static* modeling techniques because the data are assumed to be independent
- this hypothesis may be restrictive regarding some real phenomena
- *dynamical* framework : building decision rules from sequential data (or time series).

Objectives

- machine learning concepts for data analysis
- overview of some statistical learning approaches from the literature, with a particular focus on generative learning (see how they work).
- How can we define an accurate discrimination rule by considering both homogeneous and dispersed data ? (classification (discrimination))
- When expert information is missing, how can we automatically search for possible classes ? unsupervised learning for segmentation, clustering..
- How can we model the underlying (dynamical) behavior from sequential data ? (Sequential modeling)

Discriminative learning

- Two main approaches are generally used in the statistical learning literature : the *discriminative* approach and the *generative* approach
- Discriminative approaches (especially used in supervised learning (classification, regression)) learn a direct map from the inputs $\mathbf{x}$ to the output y, or they directly learn a model of the conditional distribution $p(y|\mathbf{x})$.
- From the conditional distribution $p(y|\mathbf{x})$, we can make predictions of y for any new value of $\mathbf{x}$ by using the Maximum A Posteriori (MAP) classification rule :

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} p(y|\mathbf{x}).$$

Generative learning

- Generative classifiers learn a model of the joint distribution $p(\mathbf{x}, y)$

  $\Rightarrow$ model the class conditional density $p(\mathbf{x}|y)$ together with the prior probability $p(y)$.
- The required posterior class probability is then computed using Bayes' theorem

$$p(y|\mathbf{x}) = \frac{p(y)p(\mathbf{x}|y)}{\sum_{y'} p(y')p(\mathbf{x}|y')}.$$

- the outputs y are not always available (i.e., they may be missing or hidden)

  $\Rightarrow$ generative approaches are more suitable for unsupervised learning.

$\Rightarrow$ In this course we focus on probabilistic models for data modeling, discrimination and clustering.