Introduction to machine learning model-based classification

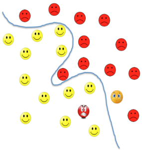
Naive Bayes.
Predictive Discriminant Analysis

Mathilde Mougeot

ENSIIE

2019

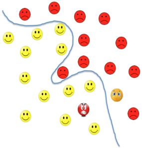
Machine Learning For classification



Naive Bayes Classifier

Y : target variable ; X : Explanatory multivariate variable.
Bayes formula.

Naive Bayes Classifier



Naive Bayes Classifier

Y : target variable ; X : Explanatory multivariate variable.

Bayes formula.

$$\begin{aligned}\mathbb{P}(Y = y/X = x) &= \frac{\mathbb{P}(X=x/Y=y)\mathbb{P}(Y=y)}{\mathbb{P}(X=x)} \\ &= \frac{\mathbb{P}(X=x/Y=y)\mathbb{P}(Y=y)}{\sum_{y'} \mathbb{P}(X=x/Y=y')\mathbb{P}(Y=y')}\end{aligned}$$

$\rightarrow$ The distribution of Y/X is completely defined by the knowledge of the distribution of Y and X/Y.

Bayes classifier.

- If we model $\mathbb{P}(X/Y)$ and $\mathbb{P}(Y)$, then we are able to estimate $\mathbb{P}(Y/X)$

Remark.

Different models on the distribution of X/Y lead to different classifiers.
Bayes classifier. Discriminant Analysis (LDA, QDA)....

Naive Bayes Classifier

The first model proposes a crude modeling for $\mathbb{P}(X/Y)$.

1. Assume that the features X^j are independent conditionally to Y :

$$\mathbb{P}(X = x/Y = y) = \prod \mathbb{P}(X^j = x^j/Y = y)$$

2. Model the univariate distribution X^j/Y. Ftor instance assume that

$$\mathbb{P}(X^j = x^j/Y = y) = \text{Normal}(\mu_{j,y}, \sigma^2_{j,y})$$

parameters $\mu_{j,y}$ and $\sigma^2_{j,y}$ are easily estimated by MLE.

$\rightarrow$ Leads to a classifier which is very easy to compute

$\rightarrow$ Requires only the computation of some averages... (MLE)

Maximum A Posteriori

What decision for a new observation : x_{new} ?

Bayes formula.

$$\mathbb{P}(Y = y/X = x_{new}) \quad = \frac{\mathbb{P}(X{=}x_{new}/Y{=}y)\ \mathbb{P}(Y{=}y)}{\mathbb{P}(X{=}x_{new})}$$

Make the decision

Use of the discriminant function to classify observation x_{new}.

1. Compute for the different modalities of y :

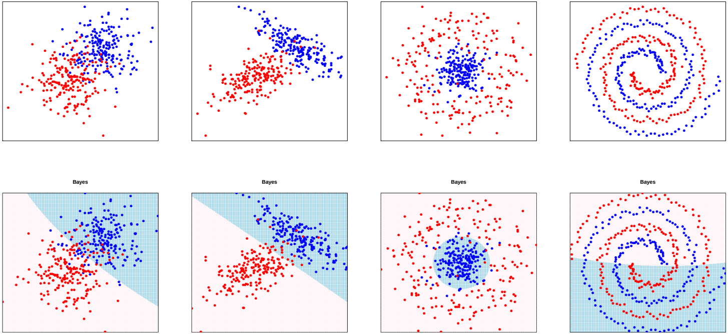
$$\delta_y(x_{new}) = \log \mathbb{P}(X = x_{new}/Y = y) + \log \mathbb{P}(Y = y)$$

2. choose the Maximum A Posteriori value for your decision.

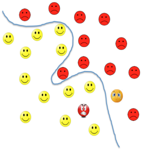
Remark. The term $\mathbb{P}(X = x)$ is similar for all comparisons.

Bayes Classifier

Illustration for several set of simulated data :



Discriminant Analysis



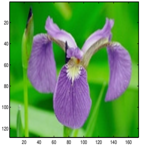
Linear Discriminant Analysis (LDA)
Quadratic Discriminant Anaysis (QDA)

Discriminant Analysis

- Very used classification method
- based on the work on Sir R. Fisher, biologist and Statistician in 1936
- also used in Credit Scoring
    - For example in Banque de France

Discriminant Analysis in 1935

- Historical Data "Fisher iris" :



- Data collection
  - The irises of the Gaspe Peninsula.
  - Different Species of Iris : Setosa, Versicolor et Virginica (Y)
  - Variables : length et width of sepals et petals ($p = 4$ variables)
  - Sample of size : $n = 150$.

  SCientific paper : Anderson E. Bulletin of the American Iris Society, 59 :2-5, 1935.

Statistics in 1935

- "iris" data set :

n	SepalLength	SepalWidth	PetalLength	PetalWidth	Species
1	5.1	3.5	1.4	0.2	setosa
...	...	...	...	...	...
50	5.3	3.7	1.5	0.2	setosa
51	7	3.2	4.7	1.4	versicolor
...	...	...	...	...	...
100	5.7	2.8	4.1	1.3	versicolor
101	6.3	3.3	6	2.5	virginica
...	...	...	...	...	
150	5.9	3	5.1	1.8	virginica

- **clean and reliable** data with **moderate size**.
- Goal :
  1. Being able to automatically classify the observation based on the values of the variables (PW,PL,SW,SL)
  2. Being able to understand the important variables
- Statistical analysis in 1935 : Linear Discriminant Analysis
  - The use of multiple measurements in taxonomic problems.

    R.A. Fisher. Annals of Eugenics, 7(2) :179-188, 1936.

Discriminant Analysis

- Two Approches **Factorial** (Geometry)
  - Descriptive (Factorial Analysis)
  - Predictive (ADL)

Discriminant Analysis

- Two Approches **Factorial** (Geometry)
  - Descriptive (Factorial Analysis)
  - Predictive (ADL)
- Approche **Predictive**
  - Linear DA
  - Quadratic DA

Discriminant Analysis. Introduction

Notations :
Definitions :

- *n* **observations**, Statistical units. $X_1, X_2, \ldots, X_n$ defined by
- *p* **Quantitative variables** $X^1, X^2, \ldots, X^p$ and with a qualitative target
- **Qualitative** target variable with *K* **modalities**, $G_1, G_2, \ldots, G_K$

$\rightarrow$ The qualitative target leads to define bf K sub- populations

Probabilistic Discriminant Analysis

This approach is also named "Bayesian approach" for the use of the Bayes theorem. We consider :

- x a multivariate observation, $x \in \mathbb{R}^p$
- K groups $G_1, \ldots, G_K$

Probabilistic Discriminant Analysis

This approach is also named "Bayesian approach" for the use of the Bayes theorem. We consider :

- x a multivariate observation, $x \in \mathbb{R}^p$
- K groups $G_1, \ldots, G_K$

$$P(G_j/x)$$

- $P(G_j/x)$: posterior probability to belong to group G_j knowing x (characteristics of x)

Probabilistic Discriminant Analysis

This approach is also named "Bayesian approach" for the use of the Bayes theorem. We consider :

- x a multivariate observation, $x \in \mathbb{R}^p$
- K groups $G_1, \ldots, G_K$

$$P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$$

- $P(G_j/x)$: posterior probability to belong to group G_j knowing x (characteristics of x)
- $P(G_j) = p_j$: prior probability for G_j (prop. of G_j in the population)

Probabilistic Discriminant Analysis

This approach is also named "Bayesian approach" for the use of the Bayes theorem. We consider :

- x a multivariate observation, $x \in \mathbb{R}^p$
- K groups $G_1, \ldots, G_K$

$$P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$$

- $P(G_j/x)$: posterior probability to belong to group G_j knowing x (characteristics of x)
- $P(G_j) = p_j$: prior probability for G_j (prop. of G_j in the population)
- $P(x/G_j) = f_j(x)$: conditional density of x given G_j group.

Probabilistic Discriminant Analysis

What is the group of a new observation :

Probabilistic Discriminant Analysis

What is the group of a new observation :

For all the groups G_j, $j = 1, \ldots, K$, one computes :

$$P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$$

The group of x is the group j_0 such that $j_0 = \arg\max_j P(G_j/x)$

For that, it is necessary to evaluate : $P(x/G_k) = f_k(x)$

Probabilistic Discriminant Analysis : conditionnal density

Assumption of Normality in the different groups, $\mathcal{N}(\mu_k, \Sigma_k)$:

Probabilistic Discriminant Analysis : conditionnal density

Assumption of Normality in the different groups, $\mathcal{N}(\mu_k, \Sigma_k)$:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

It is necessary to compute and compare :

Probabilistic Discriminant Analysis : conditionnal density

Assumption of Normality in the different groups, $\mathcal{N}(\mu_k, \Sigma_k)$:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

It is necessary to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$

Probabilistic Discriminant Analysis : conditionnal density

Assumption of Normality in the different groups, $\mathcal{N}(\mu_k, \Sigma_k)$:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

It is necessary to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$
- Or : $p_j * f_j(x)$

Probabilistic Discriminant Analysis : conditionnal density

Assumption of Normality in the different groups, $\mathcal{N}(\mu_k, \Sigma_k)$:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

It is necessary to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$
- Or : $p_j * f_j(x)$
- Or : $log(p_j) - \frac{p}{2} log(2\pi) - \frac{1}{2} log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)$

Probabilistic Discriminant Analysis : conditionnal density

Assumption of Normality in the different groups, $\mathcal{N}(\mu_k, \Sigma_k)$:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

It is necessary to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$
- Or : $p_j * f_j(x)$
- Or : $log(p_j) - \frac{p}{2}log(2\pi) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)$
- Or : $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)$

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.

All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.

All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

- $log(p_j) - \frac{1}{2} log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)$

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.

All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1}(x - \mu_j)$
- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}x\Sigma_j^{-1}x - \frac{1}{2}\mu_j\Sigma_j^{-1}\mu_j + x^T\Sigma_j^{-1}\mu_j$

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.

All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)$
- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}x\Sigma_j^{-1}x - \frac{1}{2}\mu_j\Sigma_j^{-1}\mu_j + x^T\Sigma_j^{-1}\mu_j$

In this particular case $\Sigma = \Sigma_j$, and one want to maximize

$$log(p_j) - \frac{1}{2}\mu_j\Sigma^{-1}\mu_j + x^T\Sigma^{-1}\mu_j$$

Some remarks :

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.
All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)$
- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}x\Sigma_j^{-1}x - \frac{1}{2}\mu_j\Sigma_j^{-1}\mu_j + x^T\Sigma_j^{-1}\mu_j$

In this particular case $\Sigma = \Sigma_j$, and one want to maximize

$$log(p_j) - \frac{1}{2}\mu_j\Sigma^{-1}\mu_j + x^T\Sigma^{-1}\mu_j$$

Some remarks :

- Normality and homoscedascity assumptions led to a Linear discrimination rules.

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.
All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

- $log(p_j) - \frac{1}{2} log(det(\Sigma_j)) - \frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1}(x - \mu_j)$
- $log(p_j) - \frac{1}{2} log(det(\Sigma_j)) - \frac{1}{2} x \Sigma_j^{-1} x - \frac{1}{2} \mu_j \Sigma_j^{-1} \mu_j + x^T \Sigma_j^{-1} \mu_j$

In this particular case $\Sigma = \Sigma_j$, and one want to maximize

$$log(p_j) - \frac{1}{2} \mu_j \Sigma^{-1} \mu_j + x^T \Sigma^{-1} \mu_j$$

Some remarks :

- Normality and homoscedascity assumptions led to a Linear discrimination rules.
- Except for the $log(p_j)$ term, it is the Discriminant Function proposed by Sir R. Fischer for the FDA (geometrical approach)

PDA : homoscedasticity

Homoscedascity assumptions : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K$.

All the groups are supposed to have the same Variance-covariance matrices. One compute and compare :

- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)$
- $log(p_j) - \frac{1}{2}log(det(\Sigma_j)) - \frac{1}{2}x\Sigma_j^{-1}x - \frac{1}{2}\mu_j\Sigma_j^{-1}\mu_j + x^T\Sigma_j^{-1}\mu_j$

In this particular case $\Sigma = \Sigma_j$, and one want to maximize

$$log(p_j) - \frac{1}{2}\mu_j\Sigma^{-1}\mu_j + x^T\Sigma^{-1}\mu_j$$

Some remarks :

- Normality and homoscedascity assumptions led to a Linear discrimination rules.
- Except for the $log(p_j)$ term, it is the Discriminant Function proposed by Sir R. Fischer for the FDA (geometrical approach)
- If equiprobability (for classes), this is exactly the Fischer Discriminant function (score with 2 groups).

PDA. Homoscedasticity with two groups

For the particularly case with 2 groups :

PDA. Homoscedasticity with two groups

For the particularly case with 2 groups :

- $p(G_1/x)^{-1} = 1 + (p_2/p_1)e^{-g(x)}$

PDA. Homoscedasticity with two groups

For the particularly case with 2 groups :

- $p(G_1/x)^{-1} = 1 + (p_2/p_1)e^{-g(x)}$
- $g(x) = (\mu_1 - \mu_2)^T \Sigma^{-1} x - 1/2(\mu_1^T \Sigma^{-1} \mu_1 - \mu_2^T \Sigma^{-1} \mu_2)$

PDA. Homoscedasticity with two groups

For the particularly case with 2 groups :

- $p(G_1/x)^{-1} = 1 + (p_2/p_1)e^{-g(x)}$
- $g(x) = (\mu_1 - \mu_2)^T \Sigma^{-1} x - 1/2(\mu_1^T \Sigma^{-1} \mu_1 - \mu_2^T \Sigma^{-1} \mu_2)$
- $g(x)$ is the Fischer Scoring fonction.

x **belongs to** G_1 if

- $P(G_1/x) > 0.5$, if
- $(p_2/p_1)e^{-g(x)} < 1$
- $g(x) > log(p_2/p_1)$

For 2 groups, the bayesian rule is similar to the score function $g(x) > log(p_2/p_1)$.
Il is a generalization of $g(x) > 0$ when the prior probabilities p_1 et p_2 are different.

PDA. Homoscedasticity with two groups

Particular case : 2 groups

PDA. Homoscedasticity with two groups

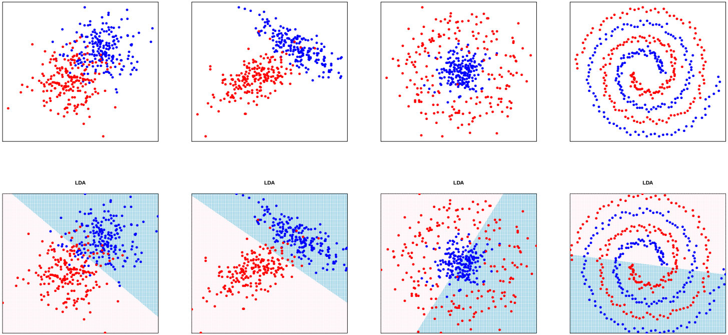
Particular case : 2 groups

- $p(G_1/x)^{-1} = 1 + (p_2/p_1)e^{-g(x)}$

The posterior probability $P(G_1/x)$ may be written as :

$P(G_1/x) = \frac{1}{1+\frac{p_2}{p_1}e^{-g(x)}} = \frac{e^{g(x)}}{\frac{p_2}{p_1}+e^{g(x)}}$ generalization of the logistic regression.

Linear Discriminant Analysis



Probabilistic Discriminant Analysis : heteroscedasticity

K groupes $G_1, \ldots, G_K$
Heteroscedasticity assumption : $\Sigma_1 \neq \Sigma_2 \neq \ldots \neq \Sigma_K$.

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

One want to compute and compare :

Probabilistic Discriminant Analysis : heteroscedasticity

K groupes $G_1, \ldots, G_K$
Heteroscedasticity assumption : $\Sigma_1 \neq \Sigma_2 \neq \ldots \neq \Sigma_K$.

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

One want to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$

Probabilistic Discriminant Analysis : heteroscedasticity

K groupes $G_1, \dots, G_K$
Heteroscedasticity assumption : $\Sigma_1 \neq \Sigma_2 \neq \dots \neq \Sigma_K$.

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

One want to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$

Probabilistic Discriminant Analysis : heteroscedasticity

K groupes $G_1, \ldots, G_K$
Heteroscedasticity assumption : $\Sigma_1 \neq \Sigma_2 \neq \ldots \neq \Sigma_K$.

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

One want to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$
- $p_j * f_j(x)$

Probabilistic Discriminant Analysis : heteroscedasticity

K groupes $G_1, \ldots, G_K$
Heteroscedasticity assumption : $\Sigma_1 \neq \Sigma_2 \neq \ldots \neq \Sigma_K$.

$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

One want to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$
- $p_j * f_j(x)$

Probabilistic Discriminant Analysis : heteroscedasticity

K groupes $G_1, \ldots, G_K$
Heteroscedasticity assumption : $\Sigma_1 \neq \Sigma_2 \neq \ldots \neq \Sigma_K$.

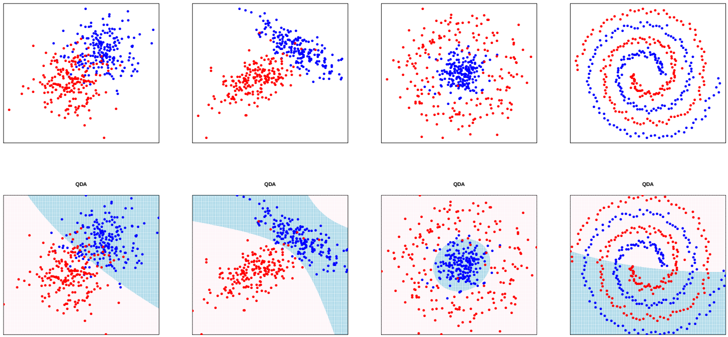
$$f_k(x) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$$

One want to compute and compare :

- $P(G_j/x) = \frac{P(G_j)P(x/G_j)}{\sum_k P(G_k)P(x/G_k)}$
- $p_j * f_j(x)$
- $log(p_j) - \frac{p}{2} log(2\pi) - \frac{1}{2} log(det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)$

The classification function is quadratic in x

Quadratic Discriminant Analysis



Probabilistic Discriminant Analysis

Sometimes, it is necessary to introduce a Cost matrix depending of the different misclassification cases :

Probabilistic Discriminant Analysis

Sometimes, it is necessary to introduce a Cost matrix depending of the different misclassification cases :

- C_{ij} is the miss classification cost for classify G_i instead of G_j.

Probabilistic Discriminant Analysis

Sometimes, it is necessary to introduce a Cost matrix depending of the different misclassification cases :

- C_{ij} is the miss classification cost for classify G_i instead of G_j.
- $C_{ii} = 0$

Probabilistic Discriminant Analysis

Sometimes, it is necessary to introduce a Cost matrix depending of the different misclassification cases :

- C_{ij} is the miss classification cost for classify G_i instead of G_j.
- $C_{ii} = 0$
- One defines an average cost for G_i with the sum $\sum_j C_{ij} P(G_j/x)$

Probabilistic Discriminant Analysis

Sometimes, it is necessary to introduce a Cost matrix depending of the different misclassification cases :

- C_{ij} is the miss classification cost for classify G_i instead of G_j.
- $C_{ii} = 0$
- One defines an average cost for G_i with the sum $\sum_j C_{ij} P(G_j/x)$
- One classify x in the group G_i for the minimal cost

Probabilistic Discriminant Analysis

Sometimes, it is necessary to introduce a Cost matrix depending of the different misclassification cases :

- C_{ij} is the miss classification cost for classify G_i instead of G_j.
- $C_{ii} = 0$
- One defines an average cost for G_i with the sum $\sum_j C_{ij} P(G_j/x)$
- One classify x in the group G_i for the minimal cost
- Example for classes.
  One classifies x in G_1 if $C_{12} P(G_2/x) < C_{21} P(G_1/x)$

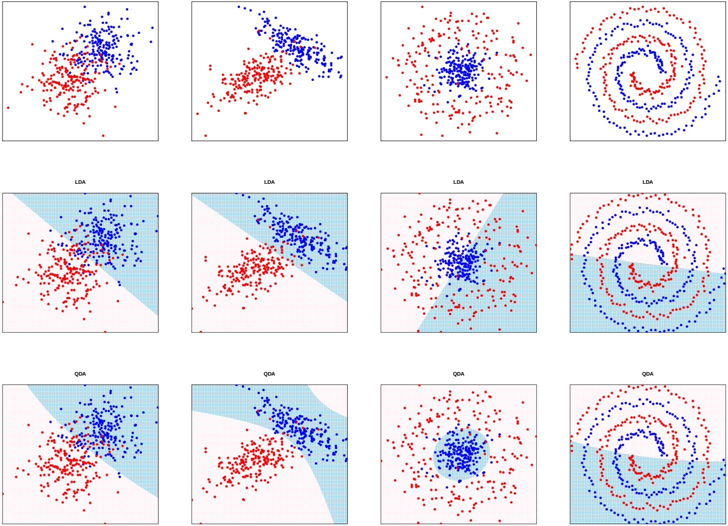
Probabilistic Discriminant Analysis

To resume :

Assumption	Bayesian rule
multi Normality	quadratic
multi Normality + homoscedasticity	linear 2 groupes $g(x) > log(p_2/p_1)$ where $g(x)$ is the fonction score of Fisher
multi Normality + homoscedasticity + equiprobability	linear et similar to the geometrical DA rule for 2 groups, we have $g(x) > 0$. the probability $P(G_1/x)$ may be written like logistic

Linear and Quadratic discriminant Analysis

From top to bottom : data set ; LDA ; QDA



Classifiers. From top to bottom : data set ; Naive Bayes classifier ; LDA ; QDA