Machine Learning

1 Introduction

1.1 Binary Classification problem

Find a binary classifier : $\begin{aligned} h : \mathbb{R}^n &\to \{-1, 1\} \\ x &\mapsto h(x) \end{aligned}$

So that $R_D(h) = \mathbb{P}_{(x,y)\sim D}[h(x) \neq y]$ is small.
With :

- X : space of input data (image, text, sound, etc.)
- Y : space of label (e.g. $Y = \{-1, 1\}$)
- D : joint probability distribution of $(x, y) \in X \times Y$

$\boxed{\to \textbf{ Objectif of ML : } \text{Risk Minimization for } h \in H}$

Définition - Test error of h

Let $h \in H$, the test error of h is defined as :

$$\begin{aligned} R_D(h) &= \mathbb{E}_{(x,y)\sim D}[\mathbb{1}_{h(x)\neq y}] \\ &= \int_{X\times Y} \mathbb{1}_{h(x)\neq y}\, D(dx, y) \end{aligned}$$

Remarque : Minimal $R_D(h)$: Bayes Classifier %D
$h_{\text{Bayes}} = \underset{y\in\{-1,1\}}{\operatorname{argmax}}\ \mathbb{P}_{(x,y)\sim D}[y|x]$

What is $P_{(x,y)\sim D}[y|x]$?

Example : Gaussians mixitures
Let $\mathbb{P}(y = -1) = \pi_0, \qquad \mathbb{P}(y = 1) = 1 - \pi_0$
$\mathbb{P}(x|y = -1) = \mathcal{N}(\mu_1, \Sigma_1), \qquad \mathbb{P}(x|y = 1) = \mathcal{N}(\mu_2, \Sigma_2)$

What are $\mathbb{P}(y = -1|x)$ and $\mathbb{P}(y = 1|x)$?

Using Bayes' theorem :
Texte manquant

Théorème

Let H be all measurable functions from X to $\{-1, 1\}$.
Then, $R_D(h) \geq R_D(h_{\text{Bayes}})$ for all $h \in H$.

- Assume $D(dx, y) = \mathbb{P}(x|y)dx \cdot \mathbb{P}(y) = \mathbb{P}(y|x)\mathbb{P}(x)dx$

$$\begin{aligned}
\text{Then :} \quad R_D(h) &= \mathbb{E}_{(x,y)\sim D}[\mathbb{1}_{h(x)\neq y}] \\
&= \sum_{y\in\{-1,1\}} \int_X \mathbb{1}_{h(x)\neq y} D(dx, y) \\
&= \sum_{y\in\{-1,1\}} \int_X \mathbb{1}_{h(x)\neq y} \mathbb{P}(y|x)\mathbb{P}(x)dx \\
&= \int_X \mathbb{P}(y=1|x)\mathbb{1}_{h(x)\neq 1}\mathbb{P}(x)dx + \int_X \mathbb{P}(y=-1|x)\mathbb{1}_{h(x)\neq -1}\mathbb{P}(x)dx
\end{aligned}$$

Why h_{Bayes} is optimal ?

Texte manquant

1.2 Linear Classification problem

In general, D is unknown and $\mathbb{P}(x|y)$ is hard to model, $\mathbb{P}(y)$ prior to choose.
Start from "simple" H : linear classifiers on $x \in \mathbb{R}^n$.

Définition - Linear classifier

A linear classifier is a function $h : \mathbb{R}^n \to \{-1, 1\}$ of the form :

$$\begin{aligned}
h(x) &= \text{sign}(\langle w, x\rangle + b) \\
&= \text{sign}(\sum_{i=1}^{n} w_i x_i + b)
\end{aligned}$$

with $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$.