INTRODUCTION TO MACHINE LEARNING

1 Warm-up: Bayes classifier for scalar Gaussian mixtures

Let $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ be independent variables in $\mathbb{R} \times \{0, 1\}$. Assume that $\mathbb{P}(Y_1 = 0) = 1/2$. Assume also that the distribution of X_1 given $\{Y_1 = 0\}$ (resp. $\{Y_1 = 1\}$) is Gaussian with mean μ_0 (resp. μ_1) and variance 1. The probability density function of X_1 is written g. Write

$$g_0 : x \mapsto (2\pi)^{-1/2} \exp(-(x - \mu_0)^2/2) \quad \text{and} \quad g_1 : x \mapsto (2\pi)^{-1/2} \exp(-(x - \mu_1)^2/2)\,.$$

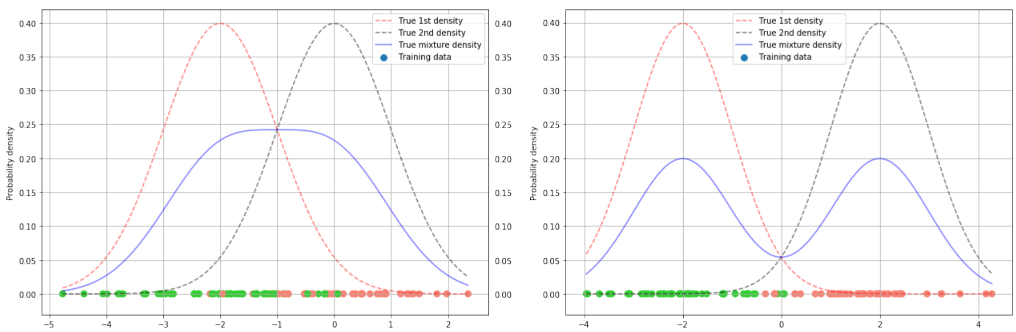

Figure 1: Samples and density when $\mu_0 = -2$ et $\mu_1 = 0$ (left) and $\mu_0 = -2$ and $\mu_1 = 2$ (right).

1. Recall the definition of the Bayes classifier.
2. Using Bayes rule, show that h_* depends only on g_1/g_0.
3. Show that the Bayes classifier uses the mean between μ_0 and μ_1 to classify samples.

2 Bayes classifier

2.1 Uniform distributions

Assume that $(X, Y) \in \mathbb{R} \times \{0, 1\}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}(Y = 1) = \pi \in (0, 1)$. Assume that conditionally on $\{Y = 0\}$ (resp. $\{Y = 1\}$) X has a uniform distribution on $[0, \theta]$ with $\theta \in (0, 1)$ (resp. on $[0, 1]$). Compute $\eta(X) = \mathbb{P}(Y = 1|X)$.

2.2 Weighted risk

Assume that $(X, Y) \in \mathbb{R} \times \{0, 1\}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Using $\omega_0, \omega_1 > 0$, with $\omega_0 + \omega_1 = 1$, we consider the weighted risk:

$$\mathsf{R}(h) = \mathbb{E}[2\omega_Y \mathbb{1}_{Y \neq h(X)}]\,.$$

Compute the Bayes classifier and its associated risk.

3 Additional exercises

3.1 Bayes classifier: excess risk

Let $(X, Y) \in \mathbb{R}^d \times \{0; 1\}$ be random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any classifier $h : \mathcal{X} \to \{0, 1\}$, define its classification error by

$$\mathsf{R}(h) = \mathbb{P}(Y \neq h(X)) .$$

The best classifier in terms of the classification error R is the Bayes classifier

$$h_*(x) = \mathrm{sign}(\eta(x) - 1/2) ,$$

where

$$\eta(X) = \mathbb{P}(Y = 1|X) .$$

1. Prove that
$$\mathsf{R}(h_*) = \mathbb{E}\left[\eta(X) \wedge (1 - \eta(X))\right] \leqslant \frac{1}{2} .$$

2. Prove that for all classifiers h, the excess risk is given by
$$\mathsf{R}(h) - \mathsf{R}(h_*) = \mathbb{E}\left[|1 - 2\eta(X)| \, |h(X) - h_*(X)|\right] .$$

3.2 Plug-in classifier

Let $(X, Y) \in \mathbb{R}^d \times \{-1; 1\}$ be random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any classifier $h : \mathcal{X} \to \{-1, 1\}$, define its classification error by

$$\mathsf{R}(h) = \mathbb{P}(Y \neq h(X)) .$$

The best classifier in terms of the classification error R is the Bayes classifier

$$h_*(x) = \mathrm{sign}(\eta(x) - 1/2) ,$$

where

$$\eta(X) = \mathbb{P}(Y = 1|X) .$$

Given n independent couples $\{(X_i, Y_i)\}_{1 \leqslant i \leqslant n}$ with the same distribution as (X, Y), an empirical surrogate for h_* is obtained from a possibly nonparametric estimator $\widehat{\eta}_n$ of η:

$$\widehat{h}_n : x \mapsto \mathrm{sign}(\widehat{\eta}_n(x) - 1/2) .$$

1. Prove that for any classifier $h : \mathcal{X} \to \{-1, 1\}$,
$$\mathbb{P}(Y \neq h(X)|X) = (2\eta(X) - 1)\mathbb{1}_{h(X)=-1} + 1 - \eta(X)$$
   and
$$\mathsf{R}(h) - \mathsf{R}(h_*) = 2\mathbb{E}\left[\left|\eta(X) - \frac{1}{2}\right| \mathbb{1}_{h(X) \neq h_*(X)}\right] .$$

2. Prove that
$$|\eta(x) - 1/2|\mathbb{1}_{\widehat{h}_n(x) \neq h_*(x)} \leqslant |\eta(x) - \widehat{\eta}_n(x)|\mathbb{1}_{\widehat{h}_n(x) \neq h_*(x)} ,$$
   where
$$\widehat{h}_n : x \mapsto \mathrm{sign}(\widehat{\eta}_n(x) - 1/2) .$$
   Deduce that
$$\mathsf{R}(\widehat{h}_n) - \mathsf{R}(h_*) \leqslant 2\mathbb{E}[|\eta(X) - \widehat{\eta}_n(X)|^2]^{1/2} .$$