

## 1 Warm-up: Bayes classifier for scalar Gaussian mixtures

Let  $(X_i, Y_i)_{1 \leq i \leq n}$  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  $\mu_0$  (resp.  $\mu_1$ ) and variance 1. The probability density function of  $X_1$  is written  $g$ . Write

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

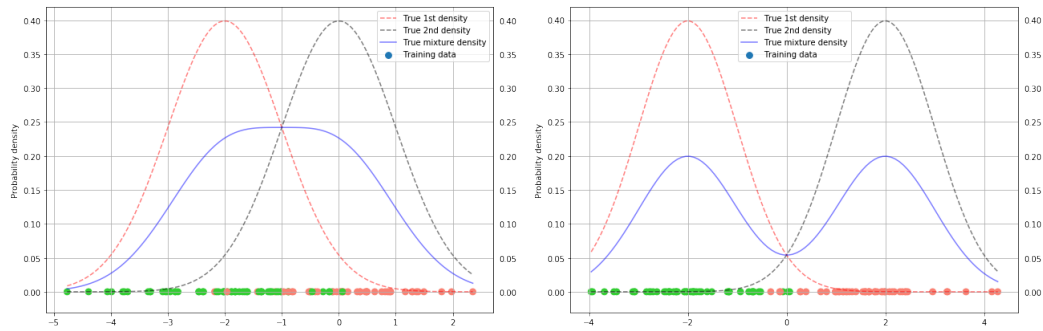


Figure 1: Samples and density when  $\mu_0 = -2$  et  $\mu_1 = 0$  (left) and  $\mu_0 = -2$  and  $\mu_1 = 2$  (right).

1. Provide an expression of a classifier  $h_*$  minimizing  $h \mapsto \mathbb{P}(h(X) \neq Y)$ .
2. Using Bayes rule, show that  $h_*$  depends only on  $g_1/g_0$ .
3. Show that the Bayes classifier uses the mean between  $\mu_0$  and  $\mu_1$  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:

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

Compute a classifier  $h_*$  minimizing  $h \mapsto R(h)$  and  $R(h_*)$ .

### 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  $h : \mathcal{X} \rightarrow \{0, 1\}$ , define its classification error by

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

The classifier  $h_*$  defined by:

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

where

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

minimizes  $h \mapsto R(h)$ .

1. Prove that

$$R(h_*) = \mathbb{E}[\eta(X) \wedge (1 - \eta(X))] \leq \frac{1}{2}.$$

2. Prove that for all classifiers  $h$ , the excess risk is given by

$$R(h) - R(h_*) = \mathbb{E}[|1 - 2\eta(X)| |h(X) - h_*(X)|].$$

#### 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  $h : \mathcal{X} \rightarrow \{-1, 1\}$ , define its classification error by

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

The classifier  $h_*$  defined by:

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

where

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

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

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

1. Prove that for any classifier  $h : \mathcal{X} \rightarrow \{-1, 1\}$ ,

$$\mathbb{P}(Y \neq h(X)|X) = (2\eta(X) - 1)\mathbb{1}_{h(X)=-1} + 1 - \eta(X)$$

and

$$R(h) - R(h_*) = 2\mathbb{E}\left[\left|\eta(X) - \frac{1}{2}\right| \mathbb{1}_{h(X) \neq h_*(X)}\right].$$

2. Prove that

$$|\eta(x) - 1/2| \mathbb{1}_{\hat{h}_n(x) \neq h_*(x)} \leq |\eta(x) - \hat{\eta}_n(x)| \mathbb{1}_{\hat{h}_n(x) \neq h_*(x)},$$

where

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

Deduce that

$$R(\hat{h}_n) - R(h_*) \leq 2\mathbb{E}[|\eta(X) - \hat{\eta}_n(X)|^2]^{1/2}.$$