Introduction to Statistical Learning

Exercise set #1

Exercise 1 - Consider the binary classification model where the random pair (X, Y) has distribution P over $\mathbb{R} \times \{0, 1\}$ and :

$$\begin{aligned} \mathcal{L}(X \mid Y = 0) &= \mathcal{U}([0, \theta]) \\ \mathcal{L}(X \mid Y = 1) &= \mathcal{U}([0, 1]) \\ p &= \mathbb{P}(Y = 1) \end{aligned}$$

with $p, \theta \in (0, 1)$ fixed. Compute the posterior probability $\eta(x) = \mathbb{P}(Y = 1 \mid X = x)$, for any $x \in \mathbb{R}$, as a function of p, θ. What if $\theta = 1/2$?

Exercise 2 - Consider the binary classification model where the random pair (X, Y) has distribution P over $\mathbb{R}_+ \times \{0, 1\}$ and :

- the marginal distribution of X over $\mathbb{R}_+$ is denoted P_X
- the conditional distribution of Y given $X = x$ is a Bernoulli distribution with parameter $\eta(x) = \dfrac{x}{x + \theta}$, for any $x \in \mathbb{R}_+$, and for fixed $\theta > 0$.

Find the Bayes classifier for this model (i.e. the minimizer of $L(g) = \mathbb{P}(Y \neq g(X))$ over all measurable classifiers $g : \mathbb{R}_+ \to \{0, 1\}$. Give the expression of the Bayes error $L^* = L(g^*)$ in the case where $P_X = \mathcal{U}([0, \alpha\theta])$ with $\alpha > 1$. What is the value of α that maximizes L^* ?

Exercise 3 - Let $X = (T, U, V)^T$ where T, U, V IID real-valued random variables with exponential distribution $\mathcal{E}(1)$. Define $Y = \mathbb{I}\{T + U + V < \theta\}$ with fixed $\theta > 0$.

1. Find the Bayes classifier $g^*(T, U)$ when V is not observed. Give the expression of the classification error of g^* (also called Bayes error). Compute it for $\theta = 9$.
2. Now assume that only T is observed, and address the same questions as above.
3. Propose a classifier for X when none of T, U, V are observed. What is its classification error ?

Exercise 4 - Find the expressions of f_+, f_- and η in the following probabilistic models :

- Discriminant Analysis : find η

$$f_+ = \mathcal{N}_d(\mu_+, \Sigma_+), f_- = \mathcal{N}_d(\mu_-, \Sigma_-)$$

- Logistic regression : find f_+, f_-

$$\log\left(\frac{\eta_\theta(x)}{1 - \eta_\theta(x)}\right) = h(x, \theta) \ , \quad \text{typically } h(x, \theta) = \theta^T x$$

Exercise 5 - Find the optimal elements in the following cases of error measures with binary classification data :

1. Asymmetric cost - set $\omega \in (0,1)$,

$$L_\omega(g) = 2\mathbb{E}\big((1-\omega)\mathbb{I}\{Y=+1\}\mathbb{I}\{g(X)=-1\} + \omega\mathbb{I}\{Y=-1\}\mathbb{I}\{g(X)=+1\}\big)$$

2. Classification with mass constraint - set $u \in (0,1)$

$$\min_g \mathbb{P}(Y \neq g(X)) \quad \text{subject to} \quad \mathbb{P}(g(X)=1)=u$$

3. Classification with reject option - set $\gamma \in (0,1/2)$

$$L_d^R(g) = \mathbb{P}(Y \neq g(X)\ ,\ g(X) \neq \circledR) + \gamma\mathbb{P}(g(X)=\circledR)$$

Exercise 6 - Consider (X,Y) a random pair that models classification data with labels in $\{0,1\}$. Define the following classification error

$$L_\omega(g) = \mathbb{E}\big(2\omega(Y)\cdot\mathbb{I}\{Y \neq g(X)\}\big)$$

where $\omega(0)+\omega(1)=1$.

Consider the unit square in $\mathbb{R}^2$.

1. Plot the curves defined by $g \mapsto (\mathbb{P}\{g(X)=1 \mid Y=0\}, \mathbb{P}\{g(X)=1 \mid Y=1\})$ when g varies such that $L_\omega(g)=C$ with C fixed, for different values of C.
2. Same question but assuming now that $\mathbb{P}\{g(X)=1\}=C$ with C fixed.

Exercise 7 - We consider the model for classification data where X is a random vector on $\mathbb{R}^d$ and Y is a random variable taking values in $\{-1,+1\}$. We denote $\eta(x)=\mathbb{P}\{Y=+1 \mid X=x\}$ the posterior probability. We consider the following problems for which the question is to compute the optimal decision rule g^* or f^* - please also provide the main proof arguments.

1. Criterion to minimize : $R(g)=\mathbb{E}\big((Y-g(X))^2\big)$ where $g\ :\ \mathbb{R}^d \to \{-1,+1\}$
2. Criterion to minimize : $R(f)=\mathbb{E}\big((Y-f(X))^2\big)$ where $f\ :\ \mathbb{R}^d \to \mathbb{R}$
3. Criterion to minimize : $A(f)=\mathbb{E}\exp(-Yf(X)))$ where $f:\mathbb{R}^d \to \mathbb{R}\cup\{-\infty,+\infty\}$.
4. Criterion to minimize : $A(f)=\mathbb{E}\big(\log(1+\exp(-Yf(X)))\big)$ where $f:\mathbb{R}^d \to \mathbb{R}\cup\{-\infty,+\infty\}$.
5. Criterion to minimize : $A_2(f)=\mathbb{E}\big(\max\{0,1-Yf(X)\}\big)$ where $f\ :\ \mathbb{R}^d \to \mathbb{R}$.

Explain why such criteria are relevant for the binary classification problem.

Exercice 8 - Consider IID random pairs (X, Y) and (X', Y') over $\mathbb{R}^d \times \mathcal{Y}$. Set the following posterior probabilities :

$$\forall x, x' \in \mathbb{R}^d, \quad \begin{aligned} \rho_+(x, x') &= \mathbb{P}\{Y - Y' > 0 \mid X = x, X' = x'\} \\ \rho_-(x, x') &= \mathbb{P}\{Y - Y' < 0 \mid X = x, X' = x'\} \end{aligned}$$

and for any preference rule $\pi : \mathbb{R}^d \times \mathbb{R}^d \to \{-1, 0, 1\}$, consider the pairwise error measure

$$L(\pi) = \mathbb{P}\left\{(Y - Y') \cdot \pi(X, X') < 0\right\} \ .$$

1. Find the Bayes rule π^* and the Bayes error $L^* = L(\pi^*)$ for this problem, as well as the excess of risk $L(\pi) - L^*$ for any preference rule π (will involve ρ_+ and ρ_-).
2. Assume $\mathcal{Y} = \{-1, +1\}$ and denote by $\eta(x) = \mathbb{P}\{Y = +1 \mid X = x\}$. Provide the expressions for $\rho_+(x, x')$ and $\rho_-(x, x')$ and discuss how the behavior of η could lead to difficult situations for the learning process to be efficient.
3. Assume now that $\mathcal{Y} = \mathbb{R}$ and that $Y = m(X) + \sigma(X) \cdot N$ where m and σ are P_X-measurable functions, N is a random noise variable with normal distribution $\mathcal{N}(0, 1)$, while N and X are independent random variables. Provide the expressions for $\rho_+(x, x')$ and $\rho_-(x, x')$ in this case and discuss the relation between properties of the model and the learning process.