Hicham Janati hicham.janati@telecom-paris.fr

Exercice 1 (Conditional distributions).
Consider a binary classification problem where the target variable Y takes values in $\{0,1\}$ and the observed data X is a random variable taking values in a subset of $\mathbb{R}^d$. We assume that the following elements are known :

1. The constant $p = P(Y = 1)$
2. The conditional distribution $P_1 = P(X|Y = 1)$
3. The conditional distribution $P_0 = P(X|Y = 0)$

Write down the posterior distribution $\eta(X) = P(Y = 1|X)$ as a function of p, P_0 and P_1.

Exercice 2 (Bayes risk).
Consider a binary classification problem where the target variable Y takes values in $\{0,1\}$ and the observed data X is a random variable taking values in a subset of $\mathbb{R}^d$ denoted by $\mathcal{X}$. Let $\eta(X) = P(Y = 1|X)$.

Show that the Bayes classifier defined by

$$h^\star = \arg\min_h P(Y \neq h(X))$$

verifies :

$$P(Y \neq h^\star(X)) = \int_{\mathcal{X}} \min(\eta(x), 1 - \eta(x)) \mathrm{d}P_X(x)$$

Exercice 3 (Bayes Risk - Weighted).
Let $w : y \in \{0,1\} \to [0,1]$ be a function such that $w(0) + w(1) = 1$. Consider the same setting of the previous exercise but with a modified loss :

$$L_w(g) = \mathbb{E}\left(2w(Y)\mathbb{1}_{\{Y \neq g(X)\}}\right)$$

Find the Bayes Classifier for the weighted loss L_w as a function of $w_0 = w(0)$ and $w_1 = w(1)$.

Exercice 4 (Sum of exponentials).
Consider three independent random variables T, U, V following a standard Exponential distribution $\mathcal{E}(1)$. Define $Y = \mathbb{1}_{\{T+U+V \geq \theta\}}$ where $\theta \geq 0$.

1. Compute the Bayes classifier for $X = (U, T)$ i.e V is not observed.
2. Compute the Bayes classifier for $X = U$ i.e both V and T are not observed.
3. Propose a classifier if no data are observed.

Exercice 5 (Concentration bounds).
Consider a binary classification problem where the data X is a random variable taking values in a subset of $\mathbb{R}^d$ and the target Y takes values in $\{-1,1\}$. Let $\mathcal{G}$ be a finite set of classifiers $g : \mathcal{X} \to \{-1,1\}$. We assume that there exists $g^\star \in \mathcal{G}$ such that the 0-1 loss is optimal i.e $L^\star = L(g^\star) = P(Y \neq g^\star(X)) = 0$. Let $n \in \mathbb{N}^\star$, the observed data are i.i.d samples $(X_1, Y_1), \ldots, (X_n, Y_n)$ following the distribution of (X, Y). Consider the empirical 0-1 loss :

$$L_n(g) = \min_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Y_i \neq g(X_i)\}}.$$

We denote the empirical classifier :

$$\hat{g}_n = \arg\min_{g \in \mathcal{G}} L_n(g).$$

1. Show that $\mathbb{P}(\min_{g \in \mathcal{G}} L_n(g) = 0) = 1$.
2. Show that for all $\varepsilon \in [0,1]$:
$$\mathbb{P}(L(\hat{g}_n) > \varepsilon) \leq |\mathcal{G}|(1-\varepsilon)^n$$
3. Deduce that for all $\varepsilon > 0$:
$$\mathbb{P}(L(\hat{g}_n) > \varepsilon) \leq |\mathcal{G}|e^{-n\varepsilon}$$
4. Show that :
$$\mathbb{E}(L(\hat{g}_n)) \leq \frac{\log(e|\mathcal{G})}{n}$$

Hicham Janati hicham.janati@telecom-paris.fr

Exercice 6 (Convergence of regressors).
Consider a binary classification problem where $Y \in \{-1, 1\}$. Given the posterior $\eta^\star(x) = P(Y = 1 \mid X = x)$, the Bayes classifier is given by $g^\star(x) = 2\mathbb{1}_{\{\eta^\star(x) \geq \frac{1}{2}\}} - 1$. Let $(\eta_n)_{n \in \mathbb{N}^\star}$ be a sequence of functions $\mathcal{X} \to]0, 1[$ and define the classifier $g_n : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta_n(x) \geq \frac{1}{2}\}} - 1$. The Bayes risk is denoted by $L^\star = L(g^\star)$.

1. Show that :
$$L(g_n) - L^\star = 2\mathbb{E}\left[\left|\eta^\star(X) - \frac{1}{2}\right| |\psi(X)|\right]$$
where $\psi(X) = \mathbb{1}_{\{\eta_n(X) < \frac{1}{2}\}} - \mathbb{1}_{\{\eta(X) < \frac{1}{2}\}}$.
2. Show that
$$L(g_n) - L^\star \leq 2\mathbb{E}\left[|\eta^\star(X) - \eta_n(X)| \, |\psi(X)|\right]$$
3. Assume there exists $\delta > 0$ such that $\forall x \, |\eta^\star(x) - \frac{1}{2}| > \delta$. Show that :
$$\mathbb{E}\left[|\psi(X)|\right] \leq \frac{L(g_n) - L^\star}{2\delta}$$
and conclude the upper bound :
$$L(g_n) - L^\star \leq \frac{2}{\delta}\mathbb{E}\left[(\eta^\star(X) - \eta_n(X))^2\right]$$
4. Assume that $L^\star = 0$. Show that for all $q \in \mathbb{N}^\star$:
$$L(g_n) - L^\star \leq 2^q \mathbb{E}\left[|\eta^\star(X) - \eta_n(X)|^q\right]$$
5. Let $\eta : \mathcal{X} \to [0, 1]$ and $g : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta(x) \geq \frac{1}{2}\}} - 1$. Show that for all $\alpha > 0$:
$$|L(g_n) - L(g)| \leq 2P\left(\left|\eta(X) - \frac{1}{2}\right| < \alpha\right) + 2P(|\eta(X) - \eta_n(X)| \geq \alpha)$$
6. Assume that $\lim_{n \to +\infty} \mathbb{E}\left[|\eta(X) - \eta_n(X)|\right] = 0$ and that $P(\eta(X) = \frac{1}{2}) = 0$. Show that :
$$\lim_{n \to +\infty} L(g_n) = L(g).$$
7. We assume that we can no longer observe Y. But we can observe a proxy target $Z \in \{-1, 1\}$ such that :
$$\mathbb{P}(Z = 1 | Y = -1, X) = P(Z = 1 | Y = -1) = a < \frac{1}{2}.$$
$$\mathbb{P}(Z = -1 | Y = 1, X) = P(Z = -1 | Y = 1) = b < \frac{1}{2}.$$
Let $\eta : x \in \mathcal{X} \mapsto \mathbb{P}(Z = 1 | X = x)$ and g its associated classifier. Show that :
$$L(g) \leq L^\star \left(1 + \frac{2|a - b|}{1 - 2\max(a, b)}\right).$$
Comment on the situation where $a = b$.

Exercice 7 (VC dimension).
Compute the VC dimension of the following sets :

1. Half-spaces of $\mathbb{R}^d$ i.e linear classifiers.
2. Boxes of the form $] - \infty, x_1] \times \cdots \times] - \infty, x_d]$ where $(x_1, \ldots, x_d) \in \mathbb{R}^d$.
3. Rectangles of $\mathbb{R}^d$.

Exercice 8 (VC dimension).
Provide an upper bound of the VC dimension of the following sets :

1. Closed balls of $\mathbb{R}^d$.

HICHAM JANATI hicham.janati@telecom-paris.fr

2. Any finite set of classifiers.

Exercice 9 (Sauer's Lemma).
Let A be a set class in $\mathbb{R}^d$ with VC dimension $V < \infty$. Show that :

1. for all $n \in \mathbb{N}^\star$ $\quad s_A(n) \leq (n+1)^V$
2. for all $n \geq V$ $\quad s_A(n) \leq \left(\frac{ne}{V}\right)^V$

Exercice 10 (Parametrized linear classifiers).
Consider a binary classification problem in $\mathbb{R}$ where the target variable Y takes values in $\{0, 1\}$. Let $\mathcal{G}$ be the set of all functions from $\mathbb{R} \to \{0, 1\}$. L denotes the usual 0-1 loss and $L^\star = \inf_g L(g)$. Consider the following family of classifiers denoted by $\mathcal{G}_\prime$:

$$g_{(x_0,y_0)} : x \in \mathbb{R} \mapsto \begin{cases} y_0 \text{ if } x \leq x_0. \\ 1 - y_0 \text{ if } x > x_0 \end{cases}$$

for $(x_0, y_0) \in \mathbb{R} \times \{0, 1\}$. For the sake of convenience, we denote $L(x_0, y_0) = L(g_{(x_0,y_0)})$ and $L_0 = \inf_{x_0,y_0} L(x_0, y_0)$.

1. What is the VC dimension of $\mathcal{G}_0$?
2. Write down the loss of an element in $\mathcal{G}_0$ as a function of $F_y(x) = \mathbb{P}(X \leq x|Y = y)$ and $p = \mathbb{P}(Y = 1)$.
3. Show that $L_0 \leq \frac{1}{2}$.
4. Using the result $\min(a, b) = \frac{a+b-|a-b|}{2}$, show that :
$$L_0 = \frac{1}{2} - \sup_x \left| pF_1(x) - (1-p)F_0(x) - p + \frac{1}{2} \right|$$
Provide a simple expression for $p = \frac{1}{2}$.
5. Show that $L_0 = \frac{1}{2}$ if and only if $L^\star = \frac{1}{2}$.
6. Prove the inequality of Chebychev-Cantelli i.e for any real random variable Z and $t \geq 0$:
$$\mathbb{P}(Z - \mathbb{E}(Z) \geq t) \leq \frac{\mathbb{V}(Z)}{\mathbb{V}(Z) + t^2}$$
7. Let m_y and σ_y^2 denote the mean and variance of the conditional distribution $X|Y = y$. Show that :
$$L_0 \leq \left(1 + \frac{(m_0 - m_1)^2}{(\sigma_0 + \sigma_1)^2}\right)^{-1}$$
Discuss the performance and limitations of the model $\mathcal{G}_0$.