# Machine Learning for EDS

Tutorial week 2

2023/2024

Generically in the following exercises, consider $\mathcal{X}$ to be a measurable space of features, $\mathcal{Y}$ a measurable space of outputs, $P$ a distribution over $\mathcal{X} \times \mathcal{Y}$, $(X, Y) \sim P$, and $\eta(X) = \mathbb{E}[Y|X]$.

**Problem 1** Consider two 6-face dice: die 1 and die 2. On die 1, 3 faces are blue, and 3 faces are red. On die 2, 5 faces are blue, and the remaining one is red. First choose a die at random, say die 1 with probability $p \in [0, 1]$. The selected die is rolled and the color appearing on the upward face is recorded. Denote by $X \in \{1, 2\}$ the selected die, and $Y \in \{0, 1\}$ the outcome of rolling the die, say $Y = 1$ for blue, $Y = 0$ for red. Let us study the problem of predicting the outcome $Y$ given $X$.

1. Is this a regression or a classification problem? Why?
2. Show that $\eta(X) = \mathbb{P}(Y = 1|X)$.
3. For the 0-1 cost function, what is the general form (expressed in terms of $\eta$) of Bayes predictors and Bayes risk in this framework?
4. Derive $\eta$ and Bayes predictors explicitly, and then show that Bayes risk is equal to $\mathcal{R}_P^* = \frac{1}{6} + \frac{p}{3}$.
5. Intuitively, why is Bayes risk an increasing function of $p$?

**Problem 2** Consider the regression framework with $\mathcal{Y} = \mathbb{R}$. Let $F_{Y|X}(\cdot|x)$ be the conditional distribution function of $Y|X = x$. In this exercise, we assume the existence of the conditional PDF $f_{Y|X=x}(\cdot|x)$, which may be relaxed.

1. For the absolute value cost $c(y, y') = |y - y'|$ for $y, y' \in \mathcal{Y}$, let us show that $f^* : x \mapsto \mathrm{med}(Y|X = x)$ is a Bayes predictor, where $\mathrm{med}(Y|X = x)$ stands for the median of the distribution of $Y$ given $X = x$:
   (a) Fix $x \in \mathcal{X}$, and consider the function $r_x : a \mapsto r_x(a) = \mathbb{E}\Big[|a - Y|\Big|X = x\Big]$. Show that for any $a \in \mathbb{R}$, $r_x'(a) = 2\mathbb{P}(Y \leq a|X = x) - 1$ and $r_x''(a) = 2f_{Y|X=x}(a)$.

   *Hint: Recall that for* $\varphi(x) = \int_{A(x)}^{B(x)} g(x, t)dt$ *(assuming all functions are differentiable),*
   $$\varphi'(x) = \int_{A(x)}^{B(x)} \frac{\partial g(x, t)}{\partial x} dt + B'(x)g(x, B(x)) - A'(x)g(x, A(x))$$

*Note: This rule cannot be applied immediately to improper integrals. To be able to apply the rule, you could split the integrals $\int_a^\infty y f_{Y|X}(y|x) = \int_a^M y f_{Y|X}(y|x) + \int_M^\infty y f_{Y|X}(y|x)$.*

(b) Show that $r_x$ admits a global minimum at $a_x^* = \mathrm{med}(Y|X=x)$.

(c) Deduce that $\mathbb{E}\Big[|f^*(X) - Y|\Big] \le \mathbb{E}\Big[|f(X) - Y|\Big]$ for any predictor $f \in \mathcal{F}$ and conclude that $f^* : x \mapsto \mathrm{med}(Y|X=x)$ is a Bayes predictor.

2. Derive a Bayes predictor when $c$ is the asymmetric absolute value cost:
$c(y, y') = c_-(y' - y)\mathbb{1}_{y<y'} + c_+(y - y')\mathbb{1}_{y \ge y'}$ for $y, y' \in \mathcal{Y}$, with $c_-, c_+ \ge 0$, $c_- + c_+ > 0$.
How does it relate to the Bayes predictor of the previous question?

**Problem 3** Let $\mu_0, \mu_1 \in \mathbb{R}^d$ two vectors and $\Sigma_0, \Sigma_1$ two symmetric positive definite matrices of size $d$. Let $Y$ be a random variable such that $\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = p \in [0, 1]$, and let $X$ be a random variable such that $X|Y = i \sim \mathcal{N}(\mu_i, \Sigma_i)$, for $i \in \{0, 1\}$. We consider the classification problem of predicting $Y$ given $X$ with the 0-1 cost.

1. Recall the general form of Bayes classifiers in this framework (expressed in terms of $\eta$).

2. Show that $\eta(x) > 1/2 \iff \dfrac{\mathbb{P}(Y = 1|X = x)}{\mathbb{P}(Y = 0|X = x)} > 1$.

3. For $i = 0, 1$, derive an expression of $\mathbb{P}(Y = i|X = x)$ in terms of $N_i(x)$ (the pdf of a multivariate normal with mean $\mu_i$ and variance $\Sigma_i$), $f_X(x)$ and $p$.

4. Show that a Bayes classifier can be expressed as
$$f^* : x \in \mathbb{R}^d \mapsto \mathbb{1}_{x^\top A x + b^\top x + c > 0},$$
for some matrix $A$, vector $b$ and constant $c$, depending on $\mu_0, \mu_1, \Sigma_0, \Sigma_1$ and $p$.
*Hint: Use the pdf of the multivariate normal distribution.*

5. What is the geometric form of the decision boundary for this classifier? What if $\Sigma_0 = \Sigma_1$?

**Problem 4** Consider the binary classification framework with $\mathcal{Y} = \{0, 1\}$. Let $c$ the asymmetric 0-1 cost function: $c(y, y') = w_{y'}\mathbb{1}_{y \ne y'}$ , for $y, y' \in \{0, 1\}$, with $w_0, w_1 \ge 0$, $w_0 + w_1 > 0$. Show that $f^* : x \in \mathcal{X} \mapsto \mathbb{1}_{\eta(x) > w_0/(w_0+w_1)}$ is a Bayes predictor and derive Bayes risk, the expression of the excess risk and the set of all Bayes predictors.

1. Define the risk conditional on $X$, $r_X(f) := \mathbb{E}\big[c(f(X), Y)\big|X\big]$, for any classifier $f \in \mathcal{F}$. Show that
$$r_X(f) = w_0 f(X)(1 - \eta(X)) + w_1(1 - f(X))\eta(X).$$

2. Deduce that $r_X(f) \geq \min\Big(w_0(1-\eta(X)), w_1\eta(X)\Big)$, and show that
$$r_X(f^*) = \min\Big(w_0(1-\eta(X)), w_1\eta(X)\Big).$$

3. Show that $f^*$ is a Bayes predictor and show that Bayes risk reads
$$\mathcal{R}_P^* = \mathbb{E}\Big[\min\Big(w_0(1-\eta(X)), w_1\eta(X)\Big)\Big].$$

4. Recall the definition of the excess risk of a classifier $f \in \mathcal{F}$ in terms of $\mathcal{R}_P^c(f)$ and $\mathcal{R}_P^*$.

5. Show that
$$\ell(f^*, f) = (w_0 + w_1)\mathbb{E}\Big[\mathbb{1}_{f(X)\neq f^*(X)}\Big|\frac{w_0}{w_0+w_1} - \eta(X)\Big|\Big],$$
and deduce that all Bayes classifiers are almost surely equal to $f^*$.