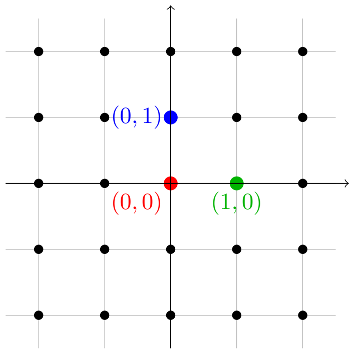
Exercice Sheet 1 – Fundamentals of supervised learning

Mathematics for Machine Learning

15 September 2025

For Exercises 1 and 2, we define the following partition of $\mathbb{R}^2$:

$$\mathcal{A} = ([k_1, k_1+1) \times [k_2, k_2+1))_{(k_1,k_2)\in\mathbb{Z}^2}.$$



It divides $\mathbb{R}^2$ into square cells of size 1, whose edges have integer coordinates.

Exercise 1. We consider a regression task $\mathbb{R}^2 \to \mathbb{R}$, with data $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2 \times \mathbb{R}$. Let $\mathcal{F}$ be the model containing all functions $f\colon \mathbb{R}^2 \to \mathbb{R}$ that are constant on each cell of the partition $\mathcal{A}$: for each $A \in \mathcal{A}$, there exists $f_A \in \mathbb{R}$ such that

$$\forall x \in A, \quad f(x) = f_A.$$

We want to determine all the empirical risk minimizers of this model for the quadratic cost.

1. Let $f \in \mathcal{F}$. Prove that the empirical risk of f for the quadratic cost reads

$$\hat{R}_n(f) = \frac{1}{2n} \sum_{A\in\mathcal{A}} \sum_{i/x_i\in A} (y_i - f_A)^2.$$

2. Let $1 \le i \le n$. Compute

$$\operatorname{argmin}_{f_A\in\mathbb{R}} \sum_{i/x_i\in A} (y_i - f_A)^2.$$

   Hint: the function to minimize is a polynomial of degree 2 in f_A.

3. Which are the empirical risk minimizers of $\mathcal{F}$ on the data $(x_1, y_1), \ldots, (x_n, y_n)$ for the quadratic cost? What happens in the cells with no data?

Exercise 2. We consider a classification task $\mathbb{R}^2 \to \{0, 1\}$, with data $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2 \times \{0, 1\}$. Let $\mathcal{F}$ be the model containing all functions $f\colon \mathbb{R}^2 \to \{0, 1\}$ that are constant on each cell of the partition $\mathcal{A}$: for each $A \in \mathcal{A}$, there exists $f_A \in \{0, 1\}$ such that

$$\forall x \in A, \quad f(x) = f_A.$$

We want to determine all the empirical risk minimizers of this model for the 0/1 cost.

1. Let $f \in \mathcal{F}$. Prove that the empirical risk of f for the 0/1 cost reads

$$\hat{R}_n(f) = \frac{1}{n} \sum_{A\in\mathcal{A}} \sum_{i/x_i\in A} \mathbf{1}_{y_i \neq f_A}.$$

2. Let $1 \leq i \leq n$. Prove that $\sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A} = \sharp\{i \mid x_i \in A, y_i = 1 - f_A\}$, where $\sharp$ denotes the number of elements in the set. Then, compute

$$\text{argmin}_{f_A \in \{0,1\}} \sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A}.$$

3. Which are the empirical risk minimizers of $\mathcal{F}$ on the data $(x_1, y_1), \ldots, (x_n, y_n)$ for the 0/1 cost? What happens in the cells with no data?

Exercise 3. We consider a regression task $\mathbb{R} \to \mathbb{R}$, with data $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$. Let $\mathcal{F}$ be the model containing all constant functions $f \colon \mathbb{R} \to \mathbb{R}$.

1. Let $f \in \mathcal{F}$. Denote $a \in \mathbb{R}$ such that $f(x) = a$ for all $x \in \mathbb{R}$. Write the empirical risk of f for the absolute value cost, using a and $y_1, \ldots, y_n$. We denote it $\varphi(a)$.

2. Assume $y_1 < \cdots < y_n$. We want to determine the empirical risk minimizers of $\mathcal{F}$ for the absolute value cost. Let $f \in \mathcal{F}$ equal to $a \in \mathbb{R}$.
   (a) Assume that $a \in [y_i, y_{i+1}]$. Prove that $\varphi(a)$ is equal to $\frac{1}{n}\sum_{j=1}^{i}(a - y_j) + \frac{1}{n}\sum_{j=i+1}^{n}(y_j - a)$.
   (b) Compute the derivative of φ with respect to a, on each interval $(-\infty, y_1], [y_1, y_2], \ldots, [y_{n-1}, y_n], [y_n, +\infty)$. *Hint: on $[y_i, y_{i+1}]$, you should obtain $\varphi'(a) = \frac{2i}{n} - 1$.* Write the variation table of φ, distinguishing the cases "n even" and "n odd".
   (c) What are the minimizers of the empirical risk φ?

Solutions

Exercise 2.

1. We have

$$\begin{aligned}
\hat{R}_n(f) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{y_i \neq f(x_i)} && \text{(by definition)} \\
&= \frac{1}{n}\sum_{A \in \mathcal{A}} \sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f(x_i)} && \text{(to sum over } x_1, \ldots, x_n\text{, we first sum on each } x_i \text{ inside one cell} \\
& && \text{of the partition } A\text{, then we sum over all cells } A \in \mathcal{A}) \\
&= \frac{1}{n}\sum_{A \in \mathcal{A}} \sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A} && (f(x_i) = f_A \text{ when } x_i \in A)
\end{aligned}$$

2. We have $\sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A} = \sharp\{i \mid x_i \in A, y_i \neq f_A\} = \sharp\{i \mid x_i \in A, y_i = 1 - f_A\}$. Denote this quantity $g(f_A)$. Let us determine $\text{argmin}_{f_A \in \{0,1\}} g(f_A)$. $g(0)$ is the number of data points in A that have a label equal to 1. $g(1)$ is the number of data points in A that have a label equal to 0.
   - If $g(0) > g(1)$, i.e., there are more positive labels than negative labels in A, then
   $$\text{argmin}_{f_A \in \{0,1\}} g(f_A) = 1.$$
   - If $g(1) > g(0)$, i.e., there are more negative labels than positive labels in A, then
   $$\text{argmin}_{f_A \in \{0,1\}} g(f_A) = 0.$$
   - If $g(0) = g(1)$, i.e., there are as many negative labels as positive labels in A, then
   $$\text{argmin}_{f_A \in \{0,1\}} g(f_A) = \{0, 1\},$$
   namely both 0 and 1 minimize the function g.

3. To construct an empirical risk minimizer f, we must choose, for each $A \in \mathcal{A}$ containing at least one data point, a value f_A that minimizes $\sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A}$. Indeed, according to question 1,

$$\hat{R}_n(f) = \frac{1}{n}\sum_{A \in \mathcal{A}} \sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A},$$

   so minimizing each term $\sum_{i/x_i \in A} \mathbf{1}_{y_i \neq f_A}$ leads to minimizing the whole empirical risk.
   We obtain, with question 2, that f is an empirical risk minimizer if it predicts the most frequent label in each cell A of the partition, and any value on the cells where there is no data.