Exercises sheet

Exercise 1: Multi-label Classification

Let us consider a discrete random variables $Y : \Omega \to \mathcal{Y} = \{a_1, \dots, a_K\}$ and a continuous random variable $X : \Omega \to \mathcal{X} = \mathbb{R}^d$.

1. Recall that X has a density given by

$$p_X(x) = \sum_{k=1}^{K} f(x, a_k),$$

for some $f : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}$, measurable.

2. Show that the Bayes classifier is

$$g^*(x) \in \arg\max_{a \in \mathcal{Y}} f(x, a).$$

3. In the case of binary classification where $K = 2$ and $a_1 = 0$, $a_2 = 1$, find the Bayes classifier.

Exercise 2: Non symmetric classification

We consider the binary classification problem where $Y \sim \mathrm{B}(p)$ and

$$X \mid Y = 0 \sim \mathcal{U}([0, 1/2]),$$
$$X \mid Y = 1 \sim \mathcal{U}([0, 1]).$$

1. Determine the cumulative distribution function (CDF) of X and its density p_X.

2. For any $x \in [0, 1]$, compute $\mathbb{E}[Y \mathbb{1}_{X \leq x}]$.

3. Show that, for any $x \in [0, 1]$,

$$\mathbb{E}[Y 1_{X \leq x}] = \int_0^x \eta^*(u) p_X(u) du,$$

where $\eta^*_P(x) = \mathbb{E}_P[Y \mid X = x]$ is the regression function.

4. Determine the conditional law of Y given $X = x$ and find the form of the Bayes classifier.

Exercise 3: Least Squares, Ridge, and Lasso in Dimension 1

Given two random variables $x : \Omega \to \mathbb{R}$ and $\varepsilon : \Omega \to \mathbb{R}$ we consider the random variable:

$$y = \beta^* x + \epsilon \tag{1}$$

for a given $\beta^* \in \mathbb{R}$ that we will try to estimate. The goal of this exercise is to compare, given a data set $((x_1, y_1), \dots, (x_1, y_1)) \in (\mathbb{R}^2)^n$, the least squares estimator:

$$\hat{\beta}^{(MC)} \in \arg\min_{\beta \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta x_i)^2,$$

with the ridge estimator

$$\hat{\beta}_\lambda^{(R)} \in \arg\min_{\beta \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta x_i)^2 + \gamma \|\beta^{(R)}\|^2,$$

and the Lasso estimator

$$\hat{\beta}_\lambda^{(L)} \in \arg\min_{\beta \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta x_i)^2 + \gamma \|\beta^{(R)}\|_1.$$

1. Write the expression of the least squares estimator $\hat{\beta}^{(MC)}$ in terms of $\{(x_i, y_i), i = 1, \ldots, n\}$. Compute the bias and variance of this estimator.

2. Write the minimization problem that the ridge estimator must solve in this framework and compute its bias, variance, and quadratic risk.

3. Give an expression for the point $x^* \in \mathbb{R}$ where the minimum of the following function is reached:

$$f(x) = a|x| + bx^2 + cx, \quad x \in \mathbb{R}, \text{ with } a, b > 0 \text{ and } c \in \mathbb{R}.$$

   Show that:

$$x^* = -\frac{c}{2b} \left(1 - \frac{a}{|c|}\right)_+ .$$

4. Compute the solution to the minimization problem for the Lasso estimator.

Exercise 4: Properties of the Ridge Estimator

We consider here the same model (1) as in the previous exercice but this time $X : \Omega \to \mathbb{R}^d$ and $\beta^* \in \mathbb{R}^d$. The Ridge estimator, given regularizing coefficient $\gamma > 0$, expresses:

$$\hat{\beta}_\gamma^{(R)} = \frac{1}{n} (X^\top X + \gamma I)^{-1} X^\top Y.$$

1. Show that the estimator:

$$\hat{\beta}^{(R')} \equiv \arg\min_{\beta \in \mathbb{R}^d, \|\beta\| \leq M_\gamma} \left(\sum_{i=1}^{n} (Y_i - X_i \beta)^2 \right),$$

   with $M_\gamma = \frac{1}{n} \|QX^\top Y\|$ and $Q \equiv (X^\top X + \gamma I)^{-1}$ is equal to $\hat{\beta}^{(R)}$.

2. Express the squared norm of the bias of $\hat{\beta}_\gamma^{(R)}$ in terms of the eigenvalues $\lambda_1, \ldots, \lambda_d$ (with multiplicities) of $X^\top X$:

$$B_\gamma^{(R)} := \|\mathbb{E}[\hat{\beta}_\gamma^{(R)}] - \beta^*\|^2.$$

3. Express the variance:

$$V_\gamma^{(R)} = \mathbb{E} \left[\|\hat{\beta}_\gamma^{(R)} - \mathbb{E}[\hat{\beta}_\gamma^{(R)}]\|^2 \right],$$

   in terms of the noise variance σ^2 and the eigenvalues $\lambda_1, \ldots, \lambda_d$.

Exercise 5: Square of the resolvent

This is a difficult problem (too difficult for a final exam exercise). Let us consider again the model (1):

$$y = x^T \beta^* + \varepsilon,$$

with $x : \Omega \to \mathbb{R}^p$, $\varepsilon : \Omega \to \mathbb{R}$, two independent variables and $\beta^* \in \mathbb{R}^p$ a deterministic vector.

1. Given a train data set $X = ((x_1, y_1), \dots, (x_n, y_n))$ and a test data (x, y), express the train MSE and the test MSE for the estimation of Y with the Ridge regression as a function of $X = (x_1, \dots, x_n) \in \mathbb{R}^{p \times p}$, x and β^*. For that, introduce the resolvent matrices $Q \equiv (\gamma I_p + \frac{1}{n} X X^T)^{-1}$ and $Q \equiv (\gamma I_n + \frac{1}{n} X^T X)^{-1}$

2. We will now try to estimate $\mathbb{E}[Q^2]$. Recall from the course the notation:

$$\forall i \in [n] : \qquad \Lambda_i \equiv 1 - \frac{1}{n} x_i^T Q_{-i} x_i$$

   where $Q_{-i} = (\gamma I_p + \frac{1}{n} X_{-i} X_{-i}^T)^{-1}$ and $X_{-i} = (x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \in \mathcal{M}_{p,n}$ and the identities:

$$Q = Q_{-i} + \frac{1}{n} \frac{Q_{-i} x_i x_i^T Q_{-i}}{\Lambda_i} \qquad \text{and} \qquad Q x_i = \frac{Q_{-i} x_i}{\Lambda_i}$$

   We further introduced the deterministic matrix:

$$\forall \Delta \in \mathbb{R} : \qquad \tilde{Q}^{\Delta} = \left(\gamma I_p + \frac{\Sigma}{\Delta} \right)^{-1} \qquad \text{with } \Sigma = \mathbb{E}[x_i x_i^T]$$

   and the scalar $\tilde{\Lambda} \in \mathbb{R}$ solution to:

$$\tilde{\Lambda} = \frac{1}{n} \mathrm{Tr}(\Sigma \tilde{Q}^{\tilde{\Lambda}}),$$

   To be able to set concentration results, we assume, as in the course that the matrix X has independent columns and that it is a λ-Lipschitz transformation of a Gaussian vector $Z \sim \mathcal{N}(0, I_q)$. Assuming that $\frac{p}{n}$ and $\|\Sigma\|$ are both bounded with a certain constant independent of p, n, q, first bound without justifications the following probabilities (employing some constants $C, c > 0$ independent with n, p, q):

   - $\mathbb{P}\left(\left| \Lambda_i - \tilde{\Lambda} \right| \geq t \right)$
   - $\mathbb{P}\left(\left| u^T Q_{-i} x_i \right| \geq t \right)$
   - $\mathbb{P}\left(\left| x_i^T Q_{-i} \Sigma \tilde{Q} u \right| \geq t \right)$
   - $\mathbb{P}\left(\left| \frac{1}{n} x_i^T Q_{-i} x_i - \frac{1}{n} \mathrm{Tr}(\mathbb{E}[Q_{-i}] \Sigma) \right| \geq t \right)$

3. Given a deterministic vector $u \in \mathbb{R}^p$ and a deterministic matrix $A \in \mathbb{R}^{p \times p}$, such that $\|u\| \leq 1$, $\|A\| \leq O(1)$, estimate:

$$u^T Q A (Q - \tilde{Q}^{\tilde{\Lambda}}) u,$$

   and deduce that:

$$\mathbb{E}[u^T Q A Q u] = u^T \tilde{Q}^{\tilde{\Lambda}} A \tilde{Q}^{\tilde{\Lambda}} u - \frac{\mathrm{Tr}(\Sigma \tilde{Q} A \tilde{Q})}{\tilde{\Lambda}^2 n} u^T \mathbb{E}[Q \Sigma Q] u + O\left(\frac{1}{\sqrt{n}} \right)$$

4. Playing on the value of A, give an estimate of $\mathbb{E}[Q^2]$ and $\mathbb{E}[Q \Sigma Q]$ and deduce an estimation of the train and test MSE of the Ridge regression.