Regression and Classification, EMINES. *Examination, 2024. 2 hours*

Recall the notation $[a:b] = \{a, a+1, \dots, b\}$ for any integers $a < b$.

Exercise 1 (**Weighted optimal Bayes classifier.**) Consider two random variables (X, Y) taking values on $\mathbb{R}^d \times \{0,1\}$. Let $h : \mathbb{R}^d \to \{0,1\}$ be a classifier. We are first interested in minimizing (with respect to h) the weighted misclassification probability

$$M_2(h) = \alpha \mathbb{P}(Y = 0, h(X) = 1) + (1-\alpha)\mathbb{P}(Y = 1, h(X) = 0)$$

where $\alpha \in [0,1]$ is given. For example, the random variable Y may represent the illness of a patient, and in this situation, a misclassification error when the patient is ill ($Y = 1$) may be much more severe than when the patient is in good health ($Y = 0$). Hence, the coefficient α weights the importance we place on these two types of errors. Finally, we aim to find the optimal classifier in this context, that is, we aim to solve the minimization problem

$$h_2^* = \mathrm{argmin}_{h \in \mathcal{H}} M_2(h)$$

where $\mathcal{H}$ is the set of all measurable functions $h : \mathbb{R}^d \to \{0,1\}$.

1. Show that
$$M_2(h) = \mathbb{E}\left[\alpha \mathbb{P}(Y = 0|X)\mathbf{1}_{\{h(X)=1\}} + (1-\alpha)\mathbb{P}(Y = 1|X)\mathbf{1}_{\{h(X)=0\}}\right].$$
2. Deduce that $M_2(h) \geqslant \mathbb{E}\left[\min\left(\alpha \mathbb{P}(Y = 0|X), (1-\alpha)\mathbb{P}(Y = 1|X)\right)\right]$.
3. Deduce that in this context, the optimal classifier h_2^* writes:
$$h_2^*(X) = \begin{cases} 1 & \text{if } \mathbb{P}(Y = 1|X) \geqslant \delta \\ 0 & \text{otherwise} \end{cases}$$
where δ should be expressed with respect to α.

We now turn to the more general case where Y may take d values instead of only two values. Returning to the example of a patient, the different values of Y may represent different states of illness for the patient. More precisely, we consider two random variables (X, Y) taking values on $\mathbb{R}^d \times [0:(d-1)]$. Let $h : \mathbb{R}^d \to [0:(d-1)]$ be a classifier. We are interested in minimizing the weighted misclassification probability

$$M(h) = \sum_{j=0}^{d-1} \alpha_j \mathbb{P}(Y = j, h(X) \neq j)$$

where $(\alpha_j)_{j \in [0:d-1]}$ are non-negative coefficients satisfying $\sum_{j=0}^{d-1} \alpha_j = 1$. The minimization problem hence writes

$$h^* = \mathrm{argmin}_{h \in \mathcal{H}} M(h)$$

where $\mathcal{H}$ is the set of all measurable functions $h : \mathbb{R}^d \to [0:d-1]$.

4. Show that
$$M(h) = \sum_{i=0}^{d-1}\sum_{j=0}^{d-1} \mathbb{P}(h(X) = i, Y = j)\beta_{i,j}$$
where $(\beta_{i,j})_{0 \leqslant i,j \leqslant d-1}$ should be expressed in terms of $(\alpha_j)_{0 \leqslant j \leqslant d-1}$.
5. Deduce that for any classifier h, $M(h) \geqslant \mathbb{E}\left[\min_{i \in [0:d-1]}\left(\sum_{j \neq i} \alpha_j \mathbb{P}(Y = j|X)\right)\right]$.
6. Deduce the expression of the optimal classifier h^*.

Exercise 2 Let $Y_n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ and $\mathbf{X}_n = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} \in M_{n,p}(\mathbb{R})$. Define $\hat{\beta}_n = \mathrm{argmin}_{\beta \in \mathbb{R}^p} \|\mathsf{Y}_n - \beta \mathbf{X}_n\|^2$.

Define $\hat{y}_n = \mathbf{x}_n' \hat{\beta}_n$ and $\hat{y}_n^- = \mathbf{x}_n' \hat{\beta}_{n-1}$. Define $H = \mathbf{X}_n(\mathbf{X}_n^T \mathbf{X}_n)^{-1}\mathbf{X}_n^T$, the orthogonal projection matrix on $\mathcal{I}(\mathbf{X}_n)$ (also called the hat matrix). In this exercise, we want to show that

$$y_n - \hat{y}_n^- = \frac{y_n - \hat{y}_n}{1 - h_{n,n}}$$

where $H = (h_{k,\ell})_{1 \leqslant k,\ell \leqslant n}$.

1. Recall without proof the explicit expression of $\hat{\beta}_n$ in terms of $\mathbf{X}_n$ and Y_n.
2. Define $\tilde{\mathsf{Y}}_n = (\tilde{y}_i)_{1\leqslant i\leqslant n} \in \mathbb{R}^n$ such that $\tilde{y}_i = y_i$ for $i \in \{1,\dots,n-1\}$ and $\tilde{y}_n = \hat{y}_n^-$. Show that for any $\beta \in \mathbb{R}^p$,
$$\sum_{j=1}^{n}(\tilde{y}_j - \mathbf{x}_j'\beta)^2 \geqslant \sum_{j=1}^{n-1}(y_j - \mathbf{x}_j'\beta)^2 \geqslant \sum_{j=1}^{n-1}(y_j - \mathbf{x}_j'\hat{\beta}_{n-1})^2$$
3. Deduce that $\hat{\beta}_{n-1} = \mathrm{argmin}_{\beta\in\mathbb{R}^p}\|\tilde{\mathsf{Y}}_n - \mathbf{X}_n\beta\|^2$.
4. Deduce an expression of $\hat{\beta}_{n-1}$ in terms of $\tilde{\mathsf{Y}}_n$ and $\mathbf{X}_n$.
5. Deduce that $\hat{y}_n^- = \sum_{j=1}^n h_{n,j}\tilde{y}_j$
6. Show that $\hat{y}_n^- = \hat{y}_n - h_{n,n}y_n + h_{n,n}\hat{y}_n^-$.
7. Conclude.