STA 212: Méthodes de simulation statistiques **(à rendre le : 07/05/2020 à 23h)**

Devoir : partie théorique

Enseignant: Mohammed Sedki *pageweb:* masedki.github.io

Instructions : Cette partie traite quelques éléments de la théorie de la décision pour la classification binaire avec une généralisation à la classification multi-classes.

Exercice 1 : Modélisation probabiliste (4 points points)

Nous avons vu en cours un résultat d'optimalité de la règle de classification de Bayes en classification binaire pour la perte binaire. Rappelons que la règle de Bayes est donnée par

$$g^*(x) = \begin{cases} 1 & \text{si} \quad \eta(x) \geq \frac{1}{2} \\ 0 & \text{sinon,} \end{cases}$$

où $\eta(x) = \mathbb{P}(Y = 1|X = x)$. La dépendance de la loi conditionnelle de Y sachant $X = x$ rend cette règle inaccessible. Nous avons proposé en cours de mimer la règle de Bayes en approchant η par un estimateur $\widehat{\eta}$ à partir des données (régression logistique, k-plus-proches-voisins etc). La règle de classification estimée est donnée par

$$\widehat{g}_n(x) = \begin{cases} 1 & \text{si} \quad \widehat{\eta}(x) \geq \frac{1}{2} \\ 0 & \text{sinon,} \end{cases}$$

En régression logistique, l'estimateur $\widehat{\eta}$ de η est obtenue en remplaçant les paramètres de régression par l'estimateur de maximum de vraisemblance dans la fonction de lien logistique. Notons $\mathcal{R}(g)$ le risque associé à la fonction de perte binaire de la règle de classification g donné par

$$\mathcal{R}(g) = \mathbb{E}\Big[\mathbb{1}\left\{g(X) \neq Y\right\}\Big] = \mathbb{P}\big(g(X) \neq Y\big)$$

(a) Montrer que

$$\mathcal{R}\big(\widehat{g}_n\big) - \mathcal{R}\big(g^*\big) \leq 2\mathbb{E}\Big[\big|\widehat{\eta}(X) - \eta(X)\big|\Big].$$

(b) Interpréter le résultat précédent.

Exercice 2 : Classification multi-classes (6 points points)

Supposons maintenant que la variable réponse Y est à valeurs dans un ensemble fini dénombrable $\mathcal{Y} = \{1, 2, \ldots, K\}$ où K est le nombre de classes. On note g une règle de classification à valeurs dans $\mathcal{Y}$(une application de $\mathcal{X} \mapsto \mathcal{Y}$).

(a) Pour la perte binaire $\ell\big(g(x), y\big) = \mathbb{1}\{g(x) \neq y\}$, montrer que la règle de classification de Bayes est donnée par

$$g^*(x) = \underset{j\in\mathcal{Y}}{\operatorname{argmax}}\, \eta_j(x), \quad \text{où} \quad \eta_j(x) = \mathbb{P}\big(Y = j \mid X = x\big).$$

(b) Pour toute règle de classification $g : \mapsto \mathcal{Y}$, vérifier que

$$\mathcal{R}\big(g\big) - \mathcal{R}\big(g^*\big) = \mathbb{E}\Big[\max_{j\in\mathcal{Y}} \eta_j(X) - \eta_{g(X)}(X)\Big].$$

(c) Soit $\widehat{\eta}_j$ un estimateur de η_j à partir d'un jeu de données (n réalisations i.i.d du couple (X, Y)). On définit $\widehat{g}_n$ la règle de classification obtenue par substitution, *i.e.* $\widehat{g}_n(x) = \underset{j\in\mathcal{Y}}{\operatorname{argmax}}\, \widehat{\eta}_j(x)$. Montrer que

$$\mathcal{R}\big(\widehat{g}_n\big) - \mathcal{R}\big(g^*\big) \leq 2\mathbb{E}\Big[\max_{j\in\mathcal{Y}} \big|\widehat{\eta}_j(X) - \eta_j(X)\big|\Big].$$