TD 1 : Eléments de correction

Exercice 1

1. On suppose que les $(x_i, y_i), i = 1, \dots, n$ sont des réalisations de variables aléatoires i.i.d. $(X_i, Y_i), i = 1, \dots, n$. On s'intéresse alors à la loi conditionnelle de Y sachant X.

2. Une règle de prévision est une fonction $g : \mathcal{X} \to \mathcal{Y}$.

3. La règle de Bayes est définie par

$$g^\star(x) = \begin{cases} 1 & \text{si } \mathbf{P}(Y = 1|X = x) \geq 0.5 \\ 0 & \text{sinon.} \end{cases}$$

   L'erreur de Bayes est définie par $L^\star = \mathbf{P}(g^\star(X) \neq Y)$.

4. On a

$$\begin{aligned}\mathbf{P}(g(X) \neq Y|X = x) &= 1 - (\mathbf{P}(g(X) = Y, g(X) = 1|X = x) + \mathbf{P}(g(X) = Y, g(X) = -1|X = x)) \\ &= 1 - \left(\mathbf{1}_{g(x)=1}\mathbf{P}(Y = 1|X = x) + \mathbf{1}_{g(x)=-1}\mathbf{P}(Y = -1|X = x)\right) \\ &= 1 - (\mathbf{1}_{g(x)=1}\eta(x) + \mathbf{1}_{g(x)=-1}(1 - \eta(x))).\end{aligned}$$

5. On déduit

$$\begin{aligned}\mathbf{P}(g(X) \neq Y|X = x) - &\mathbf{P}(g^\star(X) \neq Y|X = x) \\ &= \eta(x)\left(\mathbf{1}_{g^\star(x)=1} - \mathbf{1}_{g(x)=1}\right) + (1 - \eta(x))\left(\mathbf{1}_{g^\star(x)=-1} - \mathbf{1}_{g(x)=-1}\right) \\ &= (2\eta(x) - 1)\left(\mathbf{1}_{g^\star(x)=1} - \mathbf{1}_{g(x)=1}\right) \\ &\geq 0\end{aligned}$$

   par définition de $g^\star$.

6. En intégrant par rapport à μdx, où μ désigne la loi de X on obtient

$$\mathbf{P}(g(X) \neq Y) \geq \mathbf{P}(g^\star(X) \neq Y)$$

   pour toute règle g.

7. Si $x \leq 0$, on a

$$\mathbf{P}(Y = 1|X = x) = \mathbf{P}(U \leq 2|X = x) = \mathbf{P}(U \leq 2) = \frac{1}{5}.$$

   De même si $x > 0$,

$$\mathbf{P}(Y = 1|X = x) = \mathbf{P}(U > 1|X = x) = \mathbf{P}(U > 1) = \frac{9}{10}.$$

   On a donc

$$g^\star(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 & \text{si } x > 0. \end{cases}$$

L'erreur de Bayes vaut

$$L^\star = \mathbf{P}(g^\star(X) \neq Y | X \leq 0)\mathbf{P}(X \leq 0) + \mathbf{P}(g^\star(X) \neq Y | X > 0)\mathbf{P}(X > 0).$$

Or

$$\mathbf{P}(g^\star(X) \neq Y | X \leq 0) = \mathbf{P}(Y \neq -1 | X \leq 0) = \mathbf{P}(U \leq -2 | X \leq 0) = \frac{1}{5}$$

et

$$\mathbf{P}(g^\star(X) \neq Y | X > 0) = \mathbf{P}(Y \neq 1 | X > 0) = \mathbf{P}(U < 1 | X \leq 0) = \frac{1}{10}.$$

On obtient

$$L^\star = \frac{1}{5}\frac{1}{2} + \frac{1}{10}\frac{1}{2} = \frac{3}{20}.$$

Exercice 2

1.

2. On désigne par $\mathcal{I}_k, k = 0, 1$ les indices des observations dans le groupe k. On a alors

$$L_{X|Y}(x_1, \ldots, x_n) = \prod_{i \in \mathcal{I}_0} f_0(x_i) \prod_{i \in \mathcal{I}_1} f_1(x_i).$$

On obtient ainsi

$$\mathcal{L}_{X|Y}(x_1, \ldots, x_n) = -(n_0 + n_1)\log(\sqrt{2\pi}) - (n_0 + n_1)\log\sqrt{\sigma^2} - \frac{1}{2}\sum_{i \in \mathcal{I}_0} \frac{(x_i - \mu_0)^2}{\sigma^2} - \frac{1}{2}\sum_{i \in \mathcal{I}_1} \frac{(x_i - \mu_1)^2}{\sigma^2}.$$

En dérivant par rapport à μ_0 et μ_1 il est facile de voir que les EMV de ces paramètres sont donnés par

$$\widehat{\mu}_0 = \frac{1}{n_0}\sum_{i \in \mathcal{I}_0} x_i \quad \text{et} \quad \widehat{\mu}_1 = \frac{1}{n_1}\sum_{i \in \mathcal{I}_1} x_i.$$

On dérive maintenant par rapport à σ^2 :

$$\frac{\partial \mathcal{L}_{X|Y}(x_1, \ldots, x_n)}{\partial \sigma^2} = -\frac{n_0 + n_1}{\sigma^2} + \sum_{i \in \mathcal{I}_0} \frac{(x_i - \mu_0)^2}{2\sigma^4} + \sum_{i \in \mathcal{I}_1} \frac{(x_i - \mu_1)^2}{2\sigma^4}.$$

On déduit que l'EMV de σ^2 est donné par

$$\widehat{\sigma}^2 = \frac{1}{n}\left(\sum_{i \in \mathcal{I}_0}(x_i - \widehat{\mu}_0)^2 + \sum_{i \in \mathcal{I}_1}(x_i - \widehat{\mu}_1)^2\right).$$

3. Ces estimateurs correspondent à ceux proposés dans le cours, à part pour $\widehat{\sigma}^2$ où on divise par $n-2$ au lieu de n pour débiaiser.

Exercice 3

1.

2. Sous le modèle LDA, maximiser $\mathbf{P}(Y = k | X = x)$ est équivalent à maximiser $\pi_k f_k(x)$ où $f_k(x)$ est la densité d'un vecteur gaussien $\mathcal{N}(\mu_k, \Sigma)$. Par conséquent cela revient à maximiser

$$-\frac{1}{2}(x - \mu_k)^t \Sigma^{-1}(x - \mu_k) + \log(\pi_k),$$

ou encore

$$x^t \Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^t \Sigma^{-1}\mu_k + \log \pi_k.$$

Exercice 4

1. On a
$$V=\frac{1}{n}\sum_{i=1}^{n}(X_i-g)(X_i-g)^t,\ B=\frac{1}{n}\sum_{k=1}^{K}n_k(g_k-g)(g_k-g)^t,$$
$$W=\frac{1}{n}\sum_{k=1}^{K}\sum_{i:Y_i=k}(X_i-g_k)(X_i-g_k)^t.$$

2. Il est facile de voir que $\pi_a(u)=\frac{\langle u,a\rangle}{\|a\|^2}a$.

3. On a ainsi
$$V(a)=\frac{1}{n}\sum_{i=1}^{n}\|\pi_a(X_i-g)\|^2=\frac{1}{n}\sum_{i=1}^{n}a^t(X_i-g)^t(X_i-g)a=a^tVa.$$
De même $B(a)=a^tBa$ et $W(a)=a^tWa$.

4. On écrit le Lagrangien
$$L(a,\lambda)=B(a)+\lambda W(a)=a^t(B+\lambda W)a,$$
puis on le dérive par rapport à a :
$$\frac{\partial L(a,\lambda)}{\partial a}=2(B+\lambda W)a.$$
Par conséquent la solution $a^\star$ vérifie $W^{-1}Ba^\star=\lambda a^\star$, c'est donc un vecteur propre de $W^{-1}B$. De plus, il est facile de voir que $J(a^\star)=\lambda$, par conséquent $a^\star$ est un vecteur propre associé à la plus grande valeur propre de $W^{-1}B$.

5. $W^{-1}B$ possédant au plus une valeur propre non nulle, il suffit de vérifier que $W^{-1}(g_1-g_0)$ est vecteur propre de $W^{-1}B$. On a
$$W^{-1}BW^{-1}(g_1-g_0)=\frac{n_0n_1}{n^2}W^{-1}(g_1-g_0)(g_1-g_0)^tW^{-1}(g_1-g_0)=\lambda W^{-1}(g_1-g_0)$$
avec
$$\lambda=\frac{n_0n_1}{n^2}(g_1-g_0)^tW^{-1}(g_1-g_0).$$

6. Le nouvel x est affecté au groupe 1 si $\|\pi_{a^\star}(x-g_1)\|\leq\|\pi_{a^\star}(x-g_0)\|$. Or
$$\pi_{a^\star}(x-g_1)=\frac{(x-g_1)^ta_1^\star}{\|a_1^\star\|^2}a_1^\star\quad\text{et}\quad\pi_{a^\star}(x-g_0)=\frac{(x-g_0)^ta_1^\star}{\|a_1^\star\|^2}a_1^\star$$
où $a_1^\star=W^{-1}(g_1-g_0)$. Par conséquent, x est affecté au groupe 1 si
$$((x-g_1)^ta_1^\star)^2\leq((x-g_0)^ta_1^\star)^2\iff-2x^ta_1^\star\times g_1^ta_1^\star+(g_1^ta_1^\star)^2\leq-2x^ta_1^\star\times g_0^ta_1^\star+(g_0^ta_1^\star)^2$$
ou encore
$$2x^ta_1^\star(g_0^ta_1^\star-g_1^ta_1^\star)\leq(g_0^ta_1^\star)^2-(g_1^ta_1^\star)^2=(g_0^ta_1^\star+g_1^ta_1^\star)(g_0^ta_1^\star-g_1^ta_1^\star).$$
Comme $g_0^ta_1^\star-g_1^ta_1^\star=-(g_1-g_0)^tW^{-1}(g_1-g_0)\leq 0$ on déduit que x est affecté au groupe 1 si
$$2x^ta_1^\star\geq g_0^ta_1^\star+g_1^ta_1^\star\iff x^tW^{-1}(g_1-g_0)\geq\frac{1}{2}(g_1+g_0)^tW^{-1}(g_1-g_0).$$

7. Il est facile de voir que $d(x,g_1)\leq d(x,g_0)$ si et seulement si
$$x^tW^{-1}g_1-\frac{1}{2}g_1^tW^{-1}g_1\leq x^tW^{-1}g_0-\frac{1}{2}g_0^tW^{-1}g_0.$$
D'où le résultat.

8. Si on se place dans le modèle Gaussien d'analyse discriminante linéaire, x est affecté au groupe 1 si $\mathbf{P}(Y=1|X=x) \geq \mathbf{P}(Y=0|X=x)$, c'est-à-dire

$$\log(\pi_1) - \frac{1}{2}(x-\mu_1)^t\Sigma^{-1}(x-\mu_1) \geq \log(\pi_0) - \frac{1}{2}(x-\mu_0)^t\Sigma^{-1}(x-\mu_0),$$

ou encore

$$x^t\Sigma^{-1}(\mu_1-\mu_0) \geq \frac{1}{2}(\mu_1+\mu_0)^t\Sigma^{-1}(\mu_1-\mu_0) + \log\frac{\pi_0}{\pi_1}.$$

On conclut en remarquant que si

- μ_0 et μ_1 sont estimés par g_0 et g_1 (ce qui est le cas généralement) ;
- Σ est, à une constante multiplicative près, estimé par W (ce qui est le cas généralement) ;
- les probabilités a priori π_0 et π_1 sont égales,

les 2 règles coïncident.

Exercice 5

On désigne par $G^\star$ le centre de gravité du nuage de points. Vu que les effectifs sont égaux, ce centre de gravité est donné par la moyenne des centres des groupes 0 et 1. On note $a^\star = \text{vect}(\alpha_1, \alpha_2)$ le premier axe discriminant. Par définition, la frontière de l'analyse discriminante linéaire est donnée par l'ensemble des points M de coordonnées (x, y) tels que

$$\langle \overrightarrow{G^\star M}, a^\star \rangle = 0 \quad \iff \quad y = -\frac{\alpha_1}{\alpha_2}x + \frac{\alpha_1}{\alpha_2}G_x^\star + G_y^\star.$$

L'application numérique donne $y = 1.35x - 0.86$.

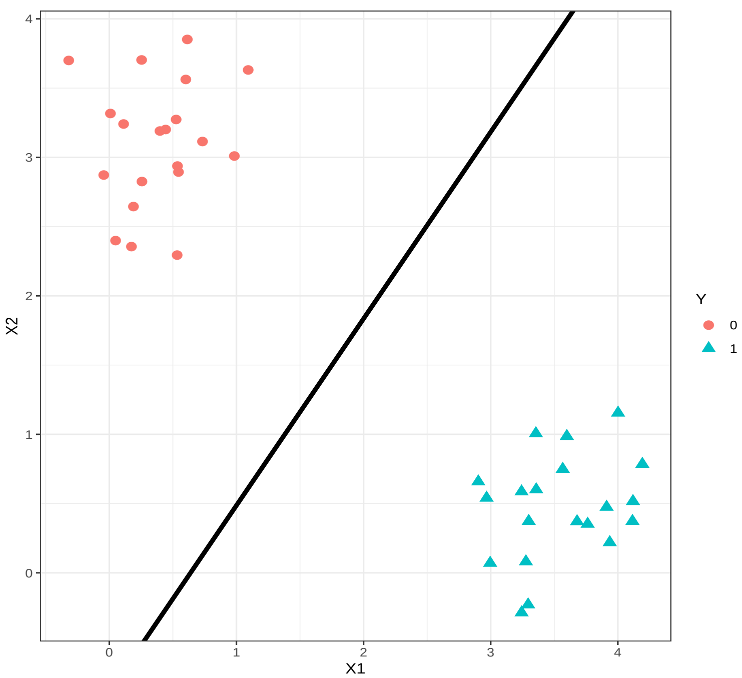

Figure 1: Frontière LDA.