TD 1 : Analyse discriminante linéaire

Exercice 1
On dispose de n observations $(x_1, y_1), \ldots, (x_n, y_n)$ telles que $x_i \in \mathcal{X}$ et $y_i \in \mathcal{Y}$ pour $i = 1, \ldots, n$. On souhaite expliquer les sorties y_i par les entrées x_i.

1. Donner une approche statistique permettant de répondre à ce problème.

2. Rappeler la définition d'une règle de prévision.

3. Rappeler la définition de la règle de Bayes $g^\star$ et de l'erreur de Bayes $L^\star$.

4. Soit g une règle de décision. Montrer que

$$\mathbf{P}(g(X) \neq Y | X = x) = 1 - (\mathbf{1}_{g(x)=1}\eta(x) + \mathbf{1}_{g(x)=-1}(1 - \eta(x)))$$

   où $\eta(x) = \mathbf{P}(Y = 1|X = x)$.

5. En déduire que pour tout $x \in \mathcal{X}$ et pour toute règle g

$$\mathbf{P}(g(X) \neq Y | X = x) - \mathbf{P}(g^\star(X) \neq Y | X = x) \geq 0.$$

6. Conclure.

7. On considère (X, Y) un couple aléatoire à valeurs dans $\mathbb{R} \times \{0, 1\}$ tel que

$$X \sim \mathcal{U}[-2, 2], \qquad U \sim \mathcal{U}[0, 10] \quad \text{et} \quad Y = \begin{cases} \mathbf{1}_{U \leq 2} & \text{si } X \leq 0 \\ \mathbf{1}_{U > 1} & \text{si } X > 0 \end{cases}$$

   où $\mathcal{U}[a, b]$ désigne la loi uniforme sur $[a, b]$. Les variables X et U sont supposées indépendantes. Calculer la règle de Bayes et l'erreur de Bayes.

Exercice 2
On cherche à expliquer une variable aléatoire Y à valeurs dans $\{0, 1\}$ par une variable aléatoire X à valeurs dans $\mathbb{R}$.

1. Quels sont les paramètres à estimer dans le modèle d'analyse discriminante linéaire.

2. Calculer la vraisemblance conditionnelle à Y et en déduire les estimateurs des paramètres des lois gaussiennes..

3. Comparer les estimateurs obtenus avec ceux du cours.

Exercice 3
On cherche à expliquer une variable aléatoire Y à valeurs dans $\{0, 1\}$ par une variable aléatoire X à valeurs dans $\mathbb{R}^p$.

1. Rappeler le modèle d'analyse discriminante linéaire.

2. Soit $x \in \mathbb{R}^p$ un nouvel individu. Montrer que la règle qui consiste à affecter x dans le groupe qui maximise $\mathbf{P}(Y = k|X = x)$ est équivalente à la règle qui consiste à affecter x dans le groupe qui maximise les fonctions linéaires discriminantes (on prendra soin de rappeler la définition des fonctions linéaires discriminantes).

Exercice 4 (Approche géométrique de LDA)
On considère un n-échantillon i.i.d. $(X_1, Y_1), \ldots, (X_n, Y_n)$ où X_i est une variable aléatoire à valeurs dans $\mathbb{R}^2$ et Y_i dans $\{0, 1\}$. On cherche une droite vectorielle a telle que les projections de chaque groupe sur a soient séparées "au mieux". Dit autrement, on cherche a telle que

- la distance entre les centres de gravité

$$g_0 = \frac{1}{\mathrm{card}\{i : Y_i = 0\}} \sum_{i:Y_i=0} X_i \quad \text{et} \quad g_1 = \frac{1}{\mathrm{card}\{i : Y_i = 1\}} \sum_{i:Y_i=1} X_i$$

  projetés sur a soit maximale (cette distance est appelée distance interclasse) ;

- la distance entre les pojections des individus et leur centre de gravité soit minimale (distance interclasse).

Pour un vecteur u de $\mathbb{R}^2$, on désigne par $\pi_a(u)$ son projeté sur la droite engendrée par a. Sans perte de généralité on supposera que a est de norme 1.

1. Rappeler les définitions des variances totale V, intra W et inter B des observations $(X_1, Y_1), \ldots,$ (X_n, Y_n).

2. Pour u fixé dans $\mathbb{R}^2$, exprimer $\pi_a(u)$ en fonction de u et a et en déduire que $\|\pi_a(u)\|^2 = a^t u u^t a$.

3. Exprimer les variances totale $V(a)$, intra $W(a)$ et inter $B(a)$ projetées sur a en fonction des variances calculées à la question 1.

On cherche maintenant à maximiser

$$J(a) = \frac{B(a)}{W(a)}$$

ou encore à

$$\text{maximiser } B(a) \quad \text{sous la contrainte} \quad W(a) = 1. \tag{1}$$

La méthode des multiplicateurs de Lagrange permet de résoudre un tel problème. La solution du problème de maximisation d'une fonction $f(x)$ sujette à $h(x) = 0$ s'obtient en résolvant l'équation

$$\frac{\partial L(x, \lambda)}{\partial x} = 0, \quad \text{où} \quad L(x, \lambda) = f(x) + \lambda h(x).$$

4. Montrer que la solution du problème (1) est un vecteur propre de $W^{-1}B$ associé à la plus grande valeur propre de $W^{-1}B$. On note $a^\star$ cette solution.

5. Montrer que $a^\star$ est colinéaire à $W^{-1}(g_1 - g_0)$. On pourra admettre que, dans le cas de 2 groupes, on a

$$B = \frac{n_0 n_1}{n^2}(g_0 - g_1)(g_0 - g_1)^t.$$

6. On considère la règle géométrique d'affectation qui consiste à classer un nouvel individu $x \in \mathbb{R}^p$ au groupe 1 si son projeté sur $a^\star$ est plus proche de $\pi_{a^\star}(g_1)$ que de $\pi_{a^\star}(g_0)$. Montrer que x sera affecté au groupe 1 si

$$S(x) = x^t W^{-1}(g_1 - g_0) > s$$

   où on exprimera s en fonction de g_0, g_1 et W.

7. Montrer que cette règle est équivalente à choisir le groupe qui minimise la distance de Mahalanobis

$$d(x, g_k) = (x - g_k)^t W^{-1}(x - g_k), \quad k = 0, 1.$$

8. On revient maintenant à l'approche probabiliste de l'analyse discriminante linéaire vue en cours et on considère la règle d'affectation qui consiste à décider "groupe 1" si $\mathbf{P}(Y = 1|X = x) \geq 0.5$. Montrer que dans ce cas, un nouvel individu x est affecter au groupe 1 si :

$$S(x) = x^t\Sigma^{-1}(\mu_1 - \mu_0) > \frac{1}{2}(\mu_1 + \mu_0)^t\Sigma^{-1}(\mu_1 - \mu_0) - \log\left(\frac{\pi_1}{\pi_0}\right).$$

Conclure.

Exercice 5

On dispose de $n = 20$ observation $(x_i, y_i), i = 1, \ldots, n$ où $x_i \in \mathbb{R}^2$ et $y_i \in \{0, 1\}$. On cherche à expliquer les y_i par les x_i à l'aide d'une analyse discriminante linéaire. Les données sont représentées sur la figure 1.

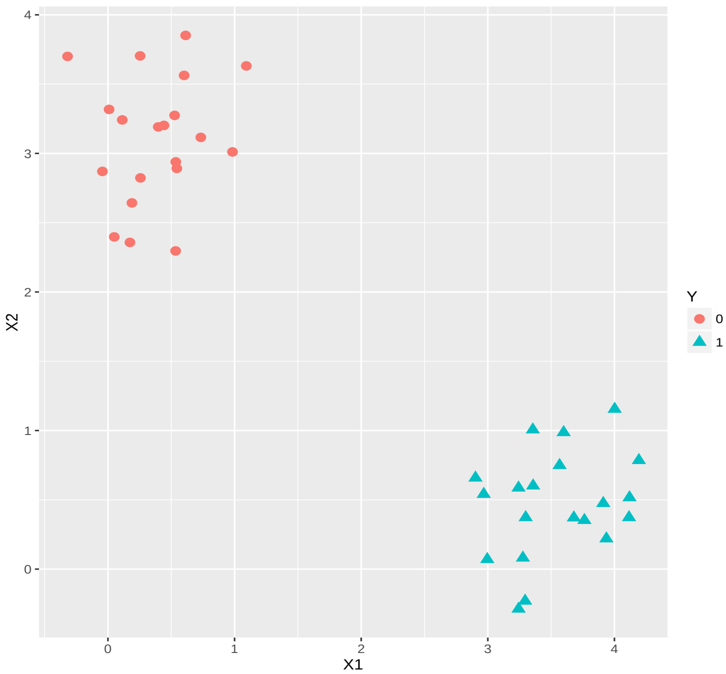

Figure 1: Le nuage de points.

On lance sur R

```
> mod <- lda(Y~.,data=D)
> mod
Call:
lda(Y ~ ., data = D)

Prior probabilities of groups:
  0   1
0.5 0.5

Group means:
         X1        X2
0 0.3850758 3.1009709
```

```
1 3.5410917 0.4692031

Coefficients of linear discriminants:
         LD1
X1  2.284995
X2 -1.694860
```

On considère la règle d'affectation géométrique. Calculer la frontière de cette règle.