

Examen de Scoring Appliqué à la Détection du Risque

Durée : deux heures. L'utilisation d'un poste informatique, de logiciels de statistique, du cours et d'internet sont autorisés.

**Exercice 1.**

On considère les données `finance` de `Rmixmod`. On veut estimer la classe (`healthy/bankruptcy`) des firmes de 2003 en s'appuyant sur l'AFD à un facteur des firmes de 2002 basée sur la métrique euclidienne.

1. La première firme de 2003 :
  - (a) A quelle distance se trouve-t-elle du centre de la classe `healthy` sur l'axe factoriel ?
  - (b) A quelle classe de risque l'affecteriez-vous si vous ne connaissiez pas sa classe ?
2. Quel est le taux d'erreur de classement des firmes de 2003 ?
3. Déterminez et interprétez le taux de faux positifs.

**Exercice 2.**

Table 1 donne le flux et le statut (0/1 pour low/high risk) de dix clients bancaires.

<i>client</i>	1	2	3	4	5	6	7	8	9	10
<i>flux</i>	5,1	5,2	5,0	5,3	4,9	5,2	5,8	5,1	5,2	5,5
<i>statut</i>	1	1	1	1	0	0	0	0	0	0

Table 1: Flux et statut (0: low risk; 1: high risk) de dix clients bancaires

On souhaite pouvoir affecter tout nouveau client à l'une des deux classes de risque. Pour cela on suppose : (i) dans chaque classe de risque le flux est gaussien et (ii) les couples : flux  $\times$  classe sont mutuellement indépendants.

Affecter à la classe 1 un client de la classe 0 coûte 80 euros ; l'erreur inverse coûte 100 euros. Un client bien classé ne coûte rien.

1. Décrivez précisément le modèle des couples flux  $\times$  classe.
2. Quelle est la valeur de *BIC* du modèle estimé sur les données de Table 1 ?
3. Avec quelle probabilité M. Li dont le flux vaut 5,2 appartient-il à la classe 0 ?

*Règle de classement 1 : on affecte les nouveaux clients par Maximum a Posteriori.*

4. A quelle classe affecte-t-on M. Li ?
5. Quelle(s) valeur(s) du flux constitue(nt) la frontière de classement ?

*Règle de classement 2 : on affecte à la classe 1 les clients dont le flux est inférieur à 5,1.*

6. Quel est le risque d'une erreur : (a) de première espèce ? (b) de seconde espèce ?
7. Quelle est l'erreur dont le coût attendu est le plus élevé ?
8. Quel est le coût attendu du classement dans la règle 2 ?

*Règle de classement 3 : la valeur  $\gamma$  du flux séparant les deux classes minimise le coût attendu du classement.*

9. Quelle est la valeur de  $\gamma$  ?
10. Quel est le coût minimal attendu du classement ?

**Exercice 3.**

On considère les données `http://alexandre.lourme.free.fr/scoring_data_banknote`. Par un modèle de régression logistique, on souhaite estimer la classe (`genuine/counterfeit`) d'un billet présentant les caractéristiques suivantes : `Length:214.9, Left:130.1, Right:129.8`.

1. Décrivez précisément le modèle de régression logistique dans ce contexte.

2. Décrivez le score et la règle d'affectation.
  3. Déterminez les paramètres du modèle.
  4. A quelle classe le nouveau billet doit-il être affecté ?
  5. Estimez l'erreur de classement par Leave-One-Out.
- 
-