

Correction examen
final 17/12/2015
HPPA 103

(1)

Exercice 1

Voir cours

Exercice 2

$$1) P(Y=i|X=x) = \frac{f_i(x)\pi_i}{f_0(x)\pi_0 + f_1(x)\pi_1}$$

$i=0,1$

$$\pi_i = P(Y=i)$$

$f_i(x)$ densité de $X|Y=i$

2) Comme tout élément a priori, le prédicteur optimal est le prédicteur de Bayes

$$g^\#(x) = \begin{cases} 0 & \text{si } P(Y=0|X=x) \geq \frac{1}{2} \\ 1 & \text{sinon} \end{cases}$$

Exercice 3

(2)

1) model.kmm ← kmm3(y ~ .,
data = données, k = 3)
kmm3 est une fonction de la
bibliothèque `corset`

2) res.kmm ← twin(y ~ .,
data = données, method = "kmm",
metric = "Accuracy", k = 3,
trialcontrol = twinControl(
method = "repeated cv", number = 5,
repeats = 100))

Plus en œuvre d'une méthode de
validation croisée à 10
ensembles répétée 100 fois

3) model.corset ← report(y ~ .,
data = données)

Exercice 4

(3)

Correction rapide, deux
éléments principaux :

- 1] comparaison par validation
croisée $n = 10$ ensembles
répétés 100 fois des
techniques de la régression
logistique et de l'analyse
discriminatoire linéaire
 \Rightarrow la régression logistique
obtient de meilleurs résultats
- 2] Plin en œuvre de la stratégie
des forêts aléatoires
 \Rightarrow meilleurs résultats que régression
logistique.