

Apprentissage supervisé

Aucun document n'est autorisé.

Durée 2 heures

Exercice 1 : Questions de cours

1. Décrivez ce qu'est un modèle de type « régression logistique »
2. Expliquez le principe de la méthode des k plus proches voisins ? Illustrez son principe avec un petit dessin ? Dans le cas d'égalité des probabilités d'appartenance aux différentes classes, quelles stratégies proposeriez-vous pour enlever l'ambiguïté.
3. Expliquez comment estimer les performances d'un modèle de classification (cross-validation, etc) et en quoi cela est important.
4. Dérivez et explicitez la méthode permettant d'obtenir les axes factoriels d'une analyse discriminante. Interprétez ces résultats.

Exercice 2 : QCM

Dans une question à deux réponses, une mauvaise réponse est pénalisée : si la question vaut 1, une réponse fausse implique -0.5.

1. L'AFD et une méthode	A – de classification non supervisée B – de prédiction C – de visualisation	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
2. L'AFD admet-elle une interprétation probabiliste	A – Vrai B – Faux	<input type="checkbox"/> <input type="checkbox"/>
3. Un axe discriminant est une combinaison linéaire	A – des variables initiales B – des valeurs propres C – des vecteurs propres	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
4. Dans une AFD, la variance d'un axe discriminant est égale à	A – 1 B – la valeur propre associée C – 0	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
5. Les axes discriminants sont	A – corrélées deux à deux B – non corrélées deux à deux C – ni l'un ni l'autre	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
6. Dans l'analyse discriminante quadratique toutes les classes ont	A – même matrice de variance-covariance B – différentes matrices de variance-covariance C – ni l'un ni l'autre	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
7. Dans une AFD les corrélations entre les variables et les axes discriminants sont	A – uniquement positives B – uniquement négatives C – positives ou négatives	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
8. la matrice à diagonaliser dans une AFD est	A – $S^{-1}W S_B$ B – $S^{-1}T S_B$ C – autre	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

9. La performance de la méthode des k plus proches voisins dépend-t-elle de	A – nombre de voisins B – distance utilisée C – ni l'un ni l'autre	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
10. La régression linéaire n'est pas adaptée au problème de classification car la variable à prédire Y prend ses valeurs	A – dans R B – dans [0,1] C – dans C	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
11. L'AFD est une ACP sur le nuage des points des K centres pondérés par les effectifs des classes	A – vrai B – faux	<input type="checkbox"/> <input type="checkbox"/>
12. En AFD, combien faut-il d'axes pour discriminer K classes	A – K B – K-1 C – K+1	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
13. La régression logistique permet de prédire une variable	A – catégorielle B – continue C – binaire	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
14. Dans le cadre de la régression logistique les variables explicatives peuvent être	A – quantitatives B – qualitatives C – quantitatives et qualitatives	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
15. En apprentissage supervisé, si la variable à expliquer Y est catégorielle, on parlera d'un problème de	A – classification B – régression C – ni l'un ni l'autre	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
16. L'AFD consiste à projeter le nuage de points sur l'axe qui maximise	A – l'inertie inter classe B – l'inertie intra classe C – le rapport inertie inter/ inertie intra	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Exercice 3 : Interprétation de code et sorties R

Considérons la table discrim suivante croisant un ensemble de 10 individus décrits par 4 variables quantitatives :

age	revenu	patrimoine	emprunt	groupe
45	250	1300	600	3
47	160	1150	450	3
38	165	850	370	1
36	175	770	250	1
29	99	450	400	1
39	170	1400	120	3
27	120	1400	160	2
51	160	1300	320	3
32	155	1500	350	2
35	170	1400	180	2

a) Expliquer la fonction de chacune des lignes de code ci-dessous

1. `data1=read.table("C:/Users/pc/Documents/Cours/Appsup/Course/Course/TD3/TP3_LDA/discrim.txt",header=T)`
2. `attach(data1)`
3. `library(MASS)`
4. `lda.fit=lda(groupe~age+revenu+patrimoine+emprunt,data1) #data1=Données`

5. names(lda.fit)
6. lda.fit\$means
7. plot(lda.fit, col = as.numeric(data1[,5]))
8. lda.fit\$scaling
9. D=as.matrix(data1[,1:4])%*%lda.fit\$scaling
10. cor(data1[,1:4],D)
11. lda.ghat <- predict(lda.fit)\$class
12. sum(lda.ghat != groupe)
13. mean(lda.ghat != groupe)
14. table(lda.ghat, groupe)

b) Interpréter les résultats de l'analyse discriminante linéaire appliquée sur la table discrim, préciser en particulier la signification des valeurs et formules soulignées ?

```
> lda.fit
```

```
Call:
```

```
lda(groupe ~ age + revenu + patrimoine + emprunt, data = data1)
```

```
Prior probabilities of groups:
```

```
  1  2  3
0.3 0.3 0.4
```

```
Group means:
```

	age	revenu	patrimoine	emprunt
1	34.33333	<u>146.3333</u>	690.000	340.0
2	31.33333	148.3333	1433.333	230.0
3	45.50000	185.0000	<u>1287.500</u>	372.5

```
Coefficients of linear discriminants:
```

	LD1	LD2
age	<u>0.048261265</u>	<u>-2.169801e-01</u>
revenu	0.025594479	5.976213e-04
patrimoine	-0.011352863	1.724238e-04
emprunt	-0.005286007	-5.039816e-05

```
Proportion of trace:
```

```
  LD1  LD2
0.8451 0.1549
```

```
> D=as.matrix(data1[,1:4])%*%lda.fit$scaling
```

```
> D
```

	LD1	LD2
[1,]	<u>-9.359949</u>	<u>-9.420789</u>
[2,]	-9.071099	-9.926839
[3,]	-5.548739	-8.018725
[4,]	-3.846767	-7.586535
[5,]	-3.289761	-6.175828
[6,]	-10.295078	-8.125285
[7,]	-12.365377	-5.553420
[8,]	-9.893802	-10.762344
[9,]	<u>-13.367892</u>	<u>-6.609737</u>
[10,]	-10.805283	-7.260388

```
> cor(data1[,1:4],D)
```

	LD1	LD2
age	-0.03673802	<u>-0.9991592</u>
revenu	-0.13640214	-0.5627526
patrimoine	<u>-0.96159782</u>	-0.1540505
emprunt	0.23596443	-0.4321157

```

> sum(lda.ghat != groupe)
[1] 0

> mean(lda.ghat != groupe)
[1] 0

> table(lda.ghat, groupe)
      groupe
lda.ghat 1 2 3
1      3 0 0
2      0 3 0
3      0 0 4

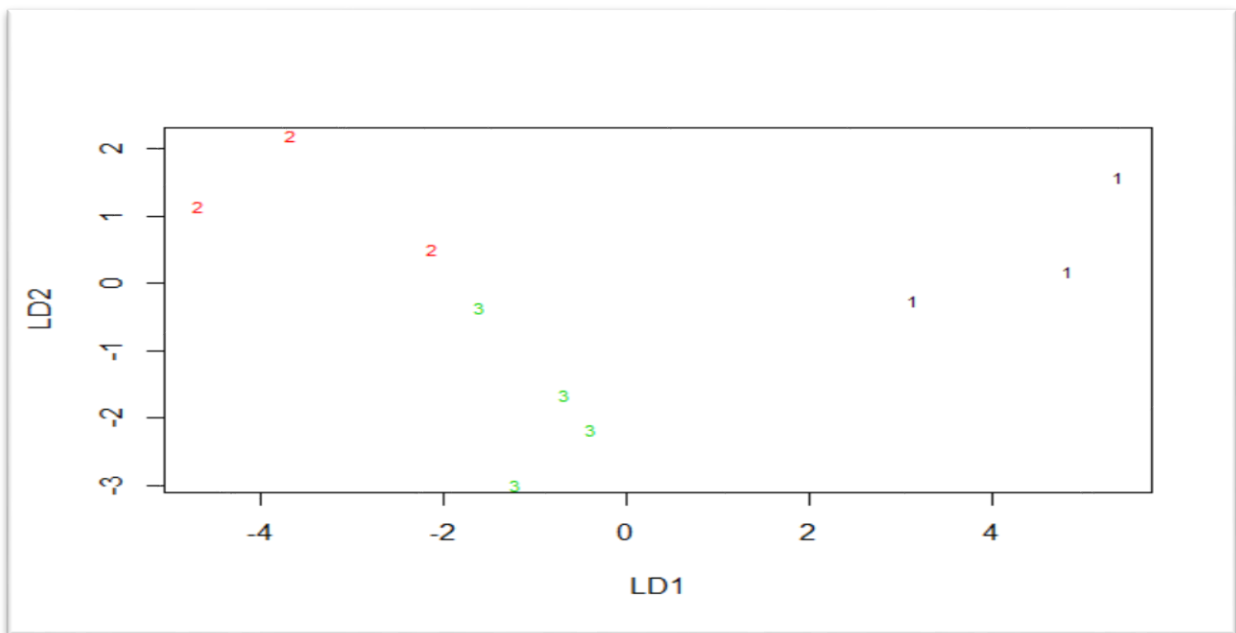
```

c) Interpréter la projection des individus de la table discrim sur le premier plan discriminant

```

> plot(lda.fit, col = as.numeric(data1[,5]))

```



Exercice 4

- Expliquer la règle géométrique de Fischer (AFD), donnez l'expression de la distance qui permet de faire l'affectation d'une nouvelle donnée x
- Expliquez l'hypothèse naïve sur laquelle repose le classifieur bayésien naïf