

Apprentissage Automatique Numérique

Rapport de TP

Classifieur Bayésien

M1 ISI

Khalid Al-kassoum Houssam

But :

Le but de ce TP est de créer un classifieur bayésien capable de classer 3 espèces d'iris dans leur classe respective selon certains critères. Dans ce TP nous disposons respectivement de 4 caractères (longueur pétale et sépale, largeur pétale et sépale), et nous faisons la classification en utilisant deux caractères à la fois.

1- Echantillonnage :

Dans cette partie du TP nous divisons l'ensemble des échantillons en 3 groupes (échantillon d'apprentissage, de développement et de test) après avoir pris les soins de mélanger les échantillons. L'échantillon d'apprentissage nous sert pour l'apprentissage des paramètres du classifieur bayésien, celui de développement nous permet la détermination des paramètres optimaux du classifieur, puis enfin celui de test nous permet de tester les performances de notre classifieur.

2- Apprentissage :

Ici nous déterminons les paramètres du classifieur, pour cela nous procédons au calcul des espérances (moyennes), les variances, les matrices de variance-covariance et les probabilités a priori. On obtient alors des paramètres pour une loi gaussienne ce qui nous permettra par la suite de calculer la probabilité d'appartenance d'un élément à une classe.

Pour classer un élément dans une classe on choisit la classe dont la probabilité d'appartenance pour cet élément est la plus grande.

Sur nos données d'apprentissage on obtient :

Tableau des moyennes

ESPERANCE CLASSE	LONG PETALE	LARG PETALE	LONG SEPALE	LARG SEPALE
0	4.9719	3.3906	1.4625	0.2313
1	5.8297	2.7541	4.1676	1.2892
2	6.5452	3.0161	5.5129	2.0290

Tableau des variances

VARIANCE CLASSE	LONG PETALE	LARG PETALE	LONG SEPALE	LARG SEPALE
0	0.1026	0.1196	0.0373	0.0090
1	0.2378	0.0987	0.1979	0.0329
2	0.2767	0.0878	0.2508	0.0704

Tableau des covariances

COVAR CLASSE	LON PET LAR PET	LON PET LON SEP	LON PET LAR SEP	LAR PET LON SEP	LAR PET LAR SEP	LON SEP LAR SEP
0	4.857e-16	1.318e-16	4.510e-17	5.065e-16	4.475e-16	7.285e-17
1	1.080e-16	5.761e-16	1.800e-17	3.000e-16	-2.280e-16	2.400e-16
2	2.721e-16	1.317e-15	9.025e-16	1.074e-15	6.589e-16	1.704e-15

Les probabilités a priori sont :

prob a priori classe 0 : 0.32

prob a priori classe 1 : 0.37

prob a priori classe 2 : 0.31

3- Développement :

Dans cette partie on a pour but de trouver les deux caractéristiques qui permettent de mieux dissocier les classes .

Les caractères sont respectivement :

1 = longueur pétale

2 = largeur pétale

3 = longueur sépale

4 = largeur sépale

en classifiant sur les caractères 1 et 2 ont a :

bonne classification 17

mauvaise classification 13

en classifiant sur les caractères 1 et 3 ont a :

bonne classification 26

mauvaise classification 4

en classifiant sur les caractères 1 et 4 ont a :

bonne classification 26

mauvaise classification 4

en classifiant sur les caractères 2 et 3 ont a :

bonne classification 24

mauvaise classification 6

en classifiant sur les caractères 2 et 4 ont a :

bonne classification 23

mauvaise classification 7

en classifiant sur les caractères 3 et 4 ont a :

bonne classification 27

mauvaise classification 3

En comparant ces différentes versions du classifieur sur les données de développement, on remarque que la meilleure version est celle qui classe sur les caractères 3 et 4.

4 - Evaluation du système :

Nous prenons comme meilleur système celui qui classe selon la Longueur du sépale et la largeur du sépale.

On obtient la matrice de confusion suivante sur les données de test :

Matrice de confusion

CLASSE	0	1	2
0	6	0	0
1	0	6	3
2	0	0	5

bonne classification 17
mauvaise classification 3

Les trois erreurs portent sur des éléments de classe 1 et qui ont été identifiés comme éléments de classe 2