

## 1 Exercices

1. Calculer la règle de discrimination du maximum de vraisemblance sur la base d'observations unidimensionnelle de distribution exponentielle.
2. Montrer que le discriminant linéaire de Fisher est équivalent à la règle du maximum de vraisemblance pour des distributions multivariées normales avec des matrices de covariances égales ( $J=2$ ).
3. On suppose que les observations proviennent de trois populations distinctes,  $\Pi_1, \Pi_2$  et  $\Pi_3$ , caractérisées par leur distributions binomiales:

$\Pi_1 : X \sim Bi(10, 0.2)$  avec la probabilité à priori  $\pi_1 = 0.5$ ;

$\Pi_2 : X \sim Bi(10, 0.3)$  avec la probabilité à priori  $\pi_2 = 0.3$ ;

$\Pi_3 : X \sim Bi(10, 0.5)$  avec la probabilité à priori  $\pi_3 = 0.2$ .

Utiliser la méthode de Bayes pour déterminer les règles discriminantes  $R_1, R_2$  et  $R_3$

4. Soit le jeux de données suivant :

	t=1			t=2		
$x_1$	2	4	3	5	4	4
$x_2$	12	10	8	7	9	5

$x_1$  et  $x_2$  suivent une distribution normale bivariée au sein de chaque groupe, où la matrice de covariance matrix est supposée être la même dans les deux groupes.

- Estimer la moyenne des groupes, la matrice de covariance, et la probabilité à priori à partir de cette ensemble de données d'apprentissage.
- Estimer les fonctions discriminantes linéaires  $a_1(x_1, x_2)$  and  $a_2(x_1, x_2)$  pour les groupes 1 et 2 respectivement.

- Donner une règle de classification linéaire pour ce problème et construisez une matrice de confusion en appliquant cette règle à l'ensemble d'apprentissage. Quelle est le taux d'erreur ?
- Dessiner la frontière entre les zones qui (selon la fonction de classification calculer précédemment) appartiennent au groupe 1 et 2 respectivement, dans un "scatterplot" des données (diagramme de dispersion). Pouvez-vous trouver une ligne droite qui a le plus faible taux d'erreur?

Rappel: Matrice de confusion

Pour évaluer la qualité de l'apprentissage on confronte la prédiction du modèle, les classes obtenues (colonnes), avec les valeurs réelles de l'échantillon, les classes désirées (lignes). Sur la diagonale, on trouve donc les valeurs bien classées, hors de la diagonale les éléments mal classés ; la somme des valeurs sur une ligne donne le nombre d'exemplaires théoriques de la catégorie. De ce tableau croisé il est alors possible d'extraire des indicateurs synthétiques, le plus connu étant le taux d'erreur ou taux de mauvais classement. Ce tableau est souvent exprimé en pourcentages, ce sont alors des taux.

		Valeur Prédiction		
		p	n	total
valeur actuelle	p'	Vrai Positif	Faux Négatif	P'
	n'	Faux Positif	Vrai Négatif	N'
total		P	N	