

TP EX21 : Décision Bayésienne et Maximum de Vraisemblance

Pour une réaliser cet exercice, il faudra utiliser le code source compagnon de cet exercice.

Soit un ensemble de données d'apprentissage labellisées contenues dans le fichier **temperatureapp.txt**. On dispose pour un ensemble d'individus de la température du corps et de leur état de santé. La première colonne contient les températures et la seconde la classe correspondante (Classe 1 : Les individus bien portant; Classe 2 : Les individus malades). Cet ensemble d'apprentissage comprend 450 individus de la classe 1 et 50 individus de la classe 2.

On cherche à réaliser à partir de ces données un classifieur Bayésien qui à partir de la température du corps, détermine automatiquement l'état de santé d'un individu.

Dans la suite, on notera les densités de probabilité avec un petit p et les probabilités avec un grand P .

Le code source associé à l'exercice réalise les différentes actions suivantes :

- a) Lit les données d'apprentissage et trace les histogrammes des températures des individus de la classe 1 et de la classe 2
- b) Estime les probabilités a priori des deux classes.
- c) Estime par Maximum de Vraisemblance et trace $p(x/C1)$ et $p(x/C2)$ en faisant l'hypothèse que ces densités de probabilité conditionnelles sont des gaussiennes.
- d) Trace $P(x/C1)P(C1)$ et $P(x/C2)P(C2)$ en fonction de x .
- e) Utilise le classifieur Bayésien pour estimer l'état de santé des 100 individus dont la température est la 1^{ère} colonne du fichier **temperaturetest.txt**. (La classe réelle correspondant au vrai état de santé de l'individu est fournie dans la deuxième colonne du fichier.)
- f) Calcule la matrice de confusion en comparant l'état de santé réel des individus (2^{ème} colonne du fichier **temperaturetest.txt**), avec l'état de santé obtenu par le classifieur.

Après avoir compris et exécuté le code source accompagnant cet énoncé. Reproduire les résultats attendus lors d'une décision Bayésienne simple et répondre aux questions suivantes.

1°) Expliquer le principe de fonctionnement, la nature et les différentes étapes d'un algorithme de décision Bayésienne et de l'estimation paramétrique par une méthode de maximum de vraisemblance.

2°) Donner les valeurs moyennes et les variances de chacune des classes. Expliquer ce qu'elles nous apprennent.

3°) Donner les probabilités à priori de chacune des classes. Qu'implique cet a priori ? (Une classe est-elle privilégiée par cet a priori.)

4°) Quel est le seuil de décision de la règle de décision Bayésienne ? Expliquer comment il est déterminé.

5°) Analyser les résultats fournis par la matrice de confusion.

6°) Si on fait l'hypothèse que les lois conditionnelles $p(x/Ci)$ sont des lois uniformes :

$$p(x/Ci) = 1/q \text{ pour } a < x < a+q$$

$$p(x/Ci) = 0 \text{ ailleurs}$$

Pour un ensemble de données x_i , tirées indépendamment, suivant la loi $p(x/Ci)$, l'estimation par le Maximum de Vraisemblance des paramètres q et a de $p(x/Ci)$ revient à estimer : a par $\min(x_i)$ et $q = (\max(x_i) - \min(x_i))$.

Modifier le programme et appliquer la décision avec une hypothèse de loi uniforme pour les densités de probabilité conditionnelles $p(x/C1)$ et $p(x/C2)$

Quel est alors le seuil de décision ?

Quelle est l'hypothèse la plus judicieuse ? Expliquer pourquoi.

Donner et analyser la matrice de confusion.