Université Lille1 Sciences et Technologies | IEEA Informatique Electronique Electrotechnique Automatique

Université des Sciences et Technologies de Lille

Sujet de rattrapage

Reconnaissance de Formes
Master 1
2010–2011

Résumé

La durée de cet exam est de 2h. Les documents de cours, TD et de TP sont autorisés. Les 2 parties sont totalement indépendantes, et doivent être rédigés sur deux copies séparées.

1 Partie 1

1.1 Elements de statistiques

1. On jette une piece de monnaie truquée et on observe une variable aléatoire X qui vaut 1 si on observe pile et 0 si on observe face. Donner la loi standard de probabilité de la variable aléatoire X si la probabilité d'avoir pile vaut p.
2. On jette 100 fois cette pièce et on observe un échantillon $x_1, \ldots, x_{100}$. Ecrire la vraisemblance de cet échantillon.
3. Donner l'estimateur du maximum de vraisemblance pour le paramètre p.
4. Rappeler ce qu'est le classifieur de Bayes entre deux distributions.
5. On sait que p vaut soit 0.2, soit 0.6. Construire le classifieur de Bayes sur l'estimateur du maximum de vraisemblance pour décider laquelle de ces deux valeurs on devrait privilégier.
6. Rappeler le principe d'un test et dire en quoi un test est proche mais different d'un classifieur.
7. On suppose que $\frac{1}{100}\sum_{i=1}^{100} x_i = 0.56$. Faire un test à 95% pour savoir si il est possible que p vaille 0.5.

1.2 Détecter des anomalies sur des photo

Imaginons que vous êtes un moniteur en colonie de vacances. Les enfants ont des appareils photos qu'ils utilisent pendant le déjeûner. Hélas, il y a des restes de nourriture sur l'objectif, ce qui résulte en de grosses tâches noires sur leurs photos. Notez que ces applats ne sont pas tous aux mêmes endroits ni de la même taille.
Les enfants pleurent car leurs photos sont ratées. Pour les aider, vous voulez faire un programme qui détecte ces tâches.

1. Donnez une procedure statistique qui pourrait effectuer automatiquement cette tâche. Argumentez sur les forces et les faiblesses de votre approche.
2. Est-il en général possible de supprimer toutes les tâches à coup sûr ?
3. Certains enfants veulent vraiment supprimer toutes les tâches, quitte à abimer leurs photos, alors que d'autres sont moins gênés par cela. Donner une façon de régler cela à l'aide de la courbe ROC, afin de permettre aux enfants de bien choisir.

2 Partie 2

2.1 Classification par arbre

On considère un problème de classification à 2 classes pour les données suivantes (4 attributs binaires) :

Classe 1	Classe 2
a=0001	e=1010
b=1111	f=0100
c=0110	g=0000
d=1010	h=1011

Q 1. Utilisez le critère d'impurté entropique pour construire un arbre de décision pour ces données.
Q 2. Exprimez chaque catégorie à l'aide d'expressions logiques aussi simple que possible (c'est-à-dire avec le plus petit nombre possible de ET et de OU).
Q 3. A quelle catégorie appartient l'exemple `x=0111` ?

2.2 Encore une question de distance

Q 4. Calculez dans un tableau la distance *edit* entre les mots `tsunami` et `tiramisu`.