

Résumé

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

1 Un peu de statistiques

On suppose que l'on dispose d'un mélange de variables aléatoires provenant de deux sources différentes : une $N(\theta_1, 1)$ et une $N(\theta_2, 1)$ (par exemple de deux couleurs différentes bruités). On suppose que la proportion de variables générés par la $N(\theta_1, 1)$ est p .

Q 1. Rappel de la densité d'une loi normale de moyenne m et de variance σ .

Q 2. Ecrire la densité de probabilité d'une variable aléatoire générée par ce mélange.

Q 3. On dispose d'un échantillon de 100 réalisations de notre variable aléatoire, $(x_i)_{1 \dots 100}$. Rappel de ce qu'est la vraisemblance. Ecrire la vraisemblance de ce modèle.

Q 4. On se propose de tester sur l'échantillon si $p = 1$. Proposer un test pour cela sachant que $\theta_1 = 0$ et $\theta_2 = 1$.

2 Des RNA pour la classification

On dispose d'images de chiens et de chats et on souhaite les classifier. Pour cela, on dispose d'un outil qui permet d'extraire des caractéristiques de ces images. On suppose maintenant que on dispose d'une collection de variables $(t_i, p_i, m_i)_i$ indexées par i . Ces variables, pour l'image i , représentent respectivement la taille, la longueur du poil, et le fait d'avoir ou non des moustaches.

Q 5. On veut classifier à l'aide d'un réseau de neurone unicouche. De quoi a-t-on besoin en plus de notre collection de caractéristique sur chaque image pour calibrer le réseau ? A-t-on besoin de connaître la loi de nos variables aléatoires ?

Q 6. Proposer un algorithme simple pour faire cela, avec un schéma. Quels sont les paramètres ? Vont-ils converger vers une solution unique ?

Q 7. Proposer un ordre de grandeur pour les paramètres du réseau, si on prend une fonction de lien indicatrice avec seuil à zéro et que on observe dans la nature que les chiens sont plus grands que les chats (en moyenne), que les poils des chiens et des chats prennent différentes longueurs indépendamment de la race et que les chats ont des moustaches et pas les chiens (mais que parfois les capteurs se trompent).

Q 8. Le classifieur sera-t-il parfait ? Pourquoi ? Et que se passera-t-il si on lui demande de classifier une photo d'alligator ?

3 Question de distance

Q 9. Calculez dans un tableau la distance *edit* entre les mots `continûment` et `contiguë`.

4 Les arbres et les décisions

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

Classe 1	Classe 2
a=0110	e=1011
b=1010	f=0000
c=0011	g=0100
d=1111	h=1110

Q 10. Utilisez le critère d'impureté entropique pour construire un arbre de décision pour ces données

Q 11. 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 12. A quelle catégorie appartient l'exemple $\mathbf{x}=0111$?