

Techniques d'apprentissage  
IFT 603 - 712

**Examen Intra : XYZ %**  
**XYZ mars 20XY, 10 h 30 à 12 h 20**

Professeur : Pierre-Marc Jodoin

---

Nom :

Prénom :

Signature :

Matricule :

---

## Consignes

1. Cet examen comporte **huit (8) questions** au total.
2. Vous avez droit à des feuilles manuscrites **non photocopiées** pour toute documentation.
3. Les calculatrices sont permises.
4. Répondez sur le questionnaire.

## Question 1 (X points)

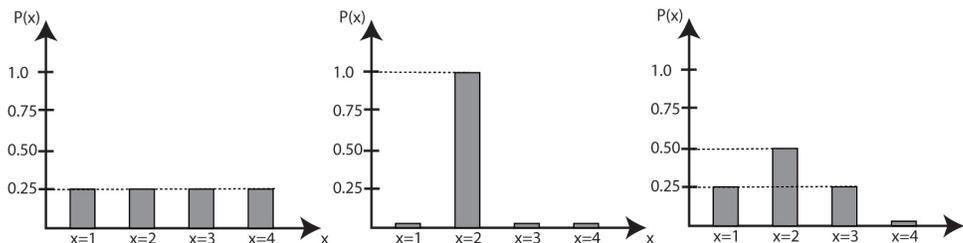
Décrivez ce qu'est un algorithme d'apprentissage et précisez en quoi consiste ses entrées et sorties. Expliquez également la différence qu'il y a entre un algorithme d'apprentissage supervisé et un algorithme non supervisé.

## Question 2 (X points)

Expliquez en quoi consiste la capacité d'un modèle et en quoi elle peut mener au sous-apprentissage ou au sur-apprentissage.

### Question 3 (X points)

Donnez la valeur numérique de l'entropie des trois histogrammes que voici :



### Question 4 (X points)

Soient  $x, y \in R$  deux variables aléatoires indépendantes,  $var(\cdot)$  la fonction de variance,  $E(\cdot)$  l'espérance mathématique et  $covar(\cdot)$  la fonction de covariance. Prouvez que

1.  $covar(x, y) = 0$
2.  $E(x + y) = E(x) + E(y)$
3.  $var(x + y) = var(x) + var(y)$ .

### Question 5 (X points)

Soit un problème de régression linéaire dont les données suivent une distribution conditionnelle gaussienne. (a) Expliquez ce qu'implique le fait que les données suivent une distribution conditionnelle gaussienne. (b) Prouvez que partant de la fonction de coût L2 suivante :  $\sum_{n=1}^N (t_n - \vec{W}^T \vec{x})^2$ , on peut obtenir le même résultat à l'aide d'un Lagrangien qu'à l'aide d'une approche par Maximum a Posteriori (MAP).

### Question 6 (X points)

Dites à quoi sert une fonction de base  $\phi(\cdot)$  dans le contexte de la régression.

### Question 7 (X points)

Nous avons vu en classe que certains algorithmes de classification s'appuient sur le concept de produit scalaire entre les paramètres  $\vec{w}$  du modèle et un vecteur de données  $\vec{x}$ . Ainsi, la fonction :

$$y_{\vec{w}}(\vec{x}) = \vec{w}^T \vec{x}$$

permet d'effectuer une classification binaire en retournant une valeur positive lorsque  $\vec{x}$  est situé devant l'hyperplan défini par  $\vec{w}$  et une valeur négative lorsque  $\vec{x}$  est derrière. Dans ce contexte, donnez une fonction de coût pouvant être utilisée pour calculer le vecteur de paramètres  $\vec{w}$ . Expliquez comment on peut généraliser  $y_{\vec{w}}(\vec{x})$  ainsi que la fonction de coût afin de segmenter des données de façon non linéaire.

### Question 8 (X points)

Nous avons vu en classe que l'approche par noyau implique le calcul d'une matrice de Gram  $K$  de taille  $N \times N$  où  $N$  est le nombre de données dans l'ensemble d'apprentissage. L'élément  $K_{ij} = k(\vec{x}_i, \vec{x}_j)$  de la matrice est la distance noyau entre les données  $\vec{x}_i$  et  $\vec{x}_j$  de l'ensemble d'entraînement. Donnez une fonction noyau ayant pour conséquence d'assigner la classe de l'élément le plus proche dans l'ensemble d'entraînement.