Université de Sherbrooke **Département d'informatique**

Techniques d'apprentissage
IFT 603 - 712

Examen Intra

Professeur : Pierre-Marc Jodoin

Nom :	Prénom :
Signature :	Matricule :

Consignes

1. Cet examen comporte **six (6) questions** au total.
2. Vous avez droit à des feuilles manuscrites pour toute documentation.
3. Les calculatrices sont permises.
4. Répondez sur le questionnaire.

Question 1 (15 points)

Vous devez effectuer une régression non linéaire à l'aide de la fonction de coût de type Ridge. Pour ce faire, vous souhaitez utiliser une fonction de base polynomiale $\phi(.)$. (a) Identifiez les paramètres ainsi que les hyperparamètres du modèle. (b) Donnez une méthode pour trouver les meilleurs paramètres au sens de la fonction de Ridge. (c) Donnez une méthode pour trouver les meilleurs hyperparamètres. Soyez clairs dans vos réponses.

Question 2 (15 points)

Soit un réseau de neurones permettant de classifier linéairement des données issues de cinq (5) classes différentes. Si on initialise au hasard les poids W du réseau, dites quelle entropie-croisée moyenne pour l'ensemble des données d'entraînement on est en droit de s'attendre avant que l'algorithme d'entraînement ne soit lancé.

Question 3 (15 points)

Soient les variables aléatoires suivantes : x l'âge d'un(e) étudiant(e) universitaire, y un programme d'étude (informatique ou mathématique), z la taille d'une personne en mètre(s), d une variable binaire exprimant si une personne a un diplôme universitaire ou non et s le genre d'une personne (féminin ou masculin).

(a) Dites dans vos propres mots en quoi consiste les formulations probabilistes que voici :

1. $p(z|s)$
2. $p(y)$
3. $p(d|x)$
4. $p(x|s)$
5. $p(x, y, s)$

(b) Répondez par vrai ou faux aux affirmations que voici en considérant le contexte d'application mentionné précédemment (justifiez votre réponse par une courte phrase) :

1. $p(x, y) = p(x)p(y)$
2. $p(z|s) = p(z)$
3. $p(x, y) = \sum_s p(x, y, s)$
4. $p(y|x) = \frac{p(x|y)p(y)}{p(x)}$
5. $p(y, z) = p(y|z)p(z|y)$

Question 4 (20 points)

Soit un problème de classification 2 classes. Prouvez mathématiquement que le maximum a posteriori suivant :

$$\vec{w} = \arg\max P(\vec{w}|X, T)$$

est équivalent dans certains cas à minimiser la fonction suivante

$$E(\vec{w}) = \frac{1}{2}\sum_{n=1}^{N}(t_n - \vec{w}^{\mathrm{T}}\vec{x}_n)^2 + \frac{\lambda}{2}\vec{w}^{\mathrm{T}}\vec{w}.$$

Veuillez définir chaque variable et les hypothèses nécessaires à votre preuve.

Question 5 (15 points)

Soit un algorithme d'apprentissage supervisé ayant pour objectif de reconnaîre la signature sur une tablette électronique de cinq (5) personnes à savoir Pierre, Julie, Frank, Hélène et Jérémy. Pour ce faire, chaque signature dispose des caractéristiques suivantes : 1) la vitesse de signature (mm/s) 2) la largeur de la signature (mm) 3) la hauteur de la signature (mm) et 4) l'inclinaison de la signature (degrés).

(a) Décrivez en quoi consiste la base de données d'apprentissage et la base de données de test associées à un tel projet.

(b) Si les données associées à chaque classe suivent une distribution Gaussienne, dites en quoi consiste une classification générative de type maximum de vraisemblance. Précisez suivant quel critère on classifie une donnée dans une classe ou dans une autre.

(c) Si les données associées à chaque classe suivent une distribution Gaussienne, dites en quoi consiste une classification générative de type maximum a posteriori. Précisez suivant quel critère on classifie une donnée dans une classe ou dans une autre.

Question 6 (20 points)

Soit la distribution gaussienne 1D suivante :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

sachant que l'espérance mathématique d'une variable gaussienne est $E[x] = \mu$ et que $E[(x-\mu)^2] = \sigma^2$ prouvez que l'entropie d'une variable gaussienne est

$$H[x] = \frac{1}{2}(1 + \ln(2\pi\sigma^2))$$

sachant que $H[x] = E[-\ln(p(x))]$.