

FONDEMENTS DE L'APPRENTISSAGE MACHINE (IFT3395/6390)

Professeur: Pascal Vincent

Examen Intra

Mercredi 14 octobre 2009

Durée: 1h45

Aucune documentation n'est permise.

Prénom:**Nom:****Code permanent:****IFT 3395 ou 6390?****Veillez répondre aux questions dans les zones de blanc laissées à cet effet.****Notations**

Les notations suivantes sont définies pour tout l'examen, là où elles ont un sens:

On suppose qu'on dispose d'un ensemble de données de n exemples: $D_n = \{z^{(1)}, \dots, z^{(n)}\}$. Dans le cas supervisé chaque exemple $z^{(i)}$ est constitué d'une paire *observation, cible*: $z^{(i)} = (x^{(i)}, t^{(i)})$, alors que dans le cas non-supervisé, on n'a pas de notion de cible explicite donc juste un vecteur d'observation: $z^{(i)} = x^{(i)}$. On suppose que chaque observation est constituée de d traits caractéristiques (composantes): $x^{(i)} \in \mathbb{R}^d$: $x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})$. On suppose aussi qu'un algorithme d'apprentissage permet de trouver une fonction $f_\theta(x)$ dont les paramètres sont paramétrés par θ et qu'on peut appliquer à tout nouveau point de test x .

ATTENTION: on vous demande de respecter *scrupuleusement* la notation définie dans cet énoncé. Par exemple souvenez-vous qu'ici la cible est notée t (ou $t^{(i)}$ s'il s'agit du i ème exemple).

1 Problèmes d'apprentissage et minimisation du risque (30 pts)

1.1 (6 pts) Dans le cadre général donné ci-dessus, expliquez brièvement dans vos propres mots la différence entre un problème de classification, un problème de régression, et un problème d'estimation de densité de probabilité.

1.2 (12 pts) Spécifiez pour *chacun* des types de problèmes: **a)** la nature de la cible t (quelle plage de valeurs elle peut prendre); **b)** la nature de la sortie que l'algorithme va calculer pour un point de test x ; **c)** la fonction de perte (ou de coût) précise $L(z, f)$ typiquement utilisée pour ce type de problème:

1. Classification binaire

2. Classification multiclasse

3. Régression

4. Estimation de densité

1.3 (6 pts)

En respectant les notations définies pour l'examen:

a) donnez la formule générale du *risque empirique* (le coût moyen sur un ensemble de données)

$$\hat{R}(\quad) =$$

b) Expliquez en Français dans vos propres mots en quoi consiste le *principe de minimisation du risque empirique*; que permet-il de trouver?

c) Donnez l'expression mathématique correspondant à la minimisation du risque empirique

1.4 (6 pts) Quelle technique va-t-on souvent utiliser pour trouver la valeur des paramètres qui minimisent le risque empirique? Expliquez sous la forme d'un très bref pseudo-code de haut niveau le fonctionnement de cette technique.

2 Exercice de classification (20 pts)

On a affaire à un problème de classification à 4 classes. L'ensemble de données D_n contient $n = 1000$ points, dont 400 sont de la classe 1, 400 sont de la classe 2, 100 sont de la classe 3, et 100 sont de la classe 4. On suppose qu'on a créé 4 estimateurs de densité $\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4$, et entraîné chacun uniquement sur les points d'une classe (\hat{f}_1 a été entraîné sur les points de la classe 1, \hat{f}_2 sur ceux de la classe 2, etc...).

Pour un nouveau point de test x que l'on désire classifier, on obtient en appliquant ces 4 estimateurs de densité à ce point:

$$\begin{aligned}\hat{f}_1(x) &= 0.5 \\ \hat{f}_2(x) &= 1.0 \\ \hat{f}_3(x) &= 2.5 \\ \hat{f}_4(x) &= 1.5\end{aligned}$$

a) Expliquez brièvement comment vous vous y prendriez pour calculer le vecteur des probabilités d'appartenance aux classes pour ce point x : $(P(t=1|x), P(t=2|x), P(t=3|x), P(t=4|x))$. **Calculez ce vecteur.**

Votre réponse:

$(P(t=1|x), P(t=2|x), P(t=3|x), P(t=4|x)) = ($, , ,)

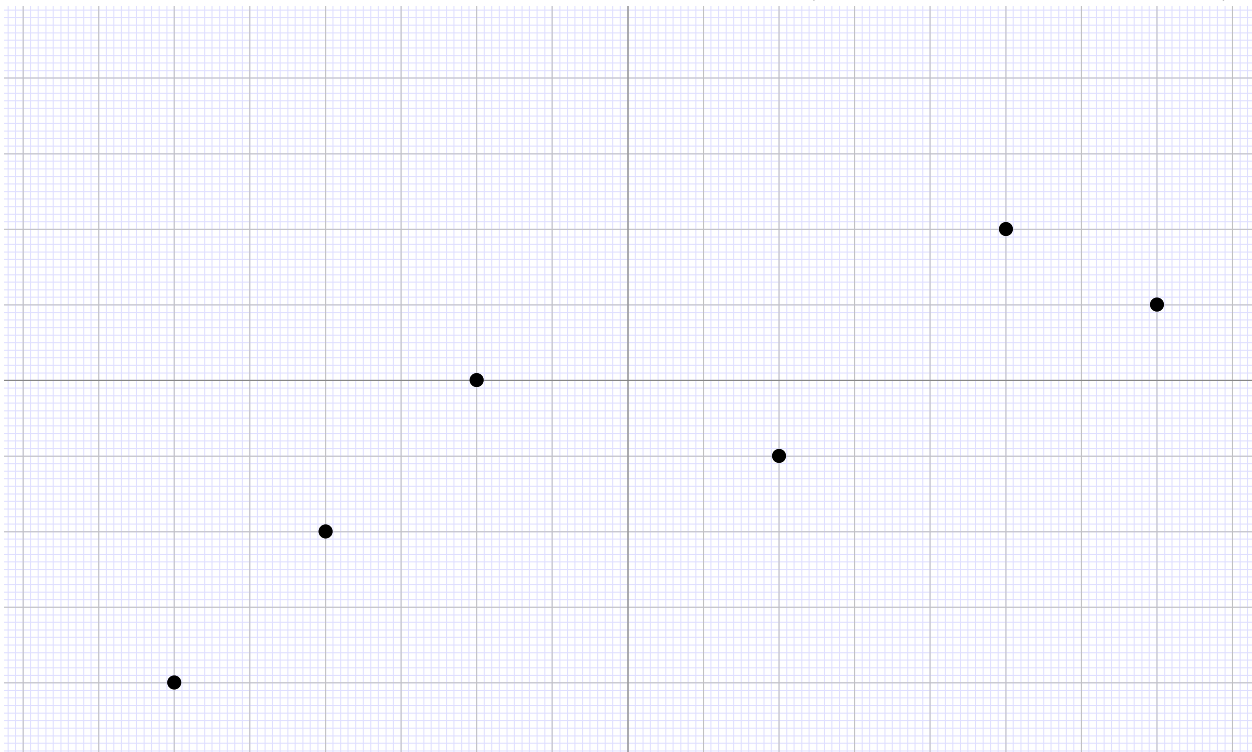
b) Quelle classe d'appartenance décidera-t-on pour ce point x ?

c) Comment s'appelle cette technique ou ce genre de classifieur?

3 K-plus proches voisins pour la régression (23 pts)

3.1 (8 pts) Expliquez brièvement mais clairement l'algorithme de **régression** des K plus proches voisins (K-ppv ou K-NN) pour un problème de **régression**. Expliquez précisément dans vos propres mots comment l'algorithme effectue une prédiction pour un point de test x .

3.2 (8 pts) Le graphique ci-dessous présente graphiquement l'ensemble de données D_n comportant $n = 6$ exemples pour un problème de régression en dimension $d = 1$ (x est en abscisse et t en ordonnée.)



- Tracez sur ce graphique **en trait plein** la prédiction de l'algorithme pour $K = 1$
- Tracez sur ce graphique **en trait pointillé** la prédiction de l'algorithme pour $K = n$ (c.a.d. ici $K = 6$)

3.3 (4 pts) On suppose que vous avez écrit la fonction `KNN_regression_predict(x, D_n, K)` qui calcule et retourne la prédiction pour un point de test x . Écrivez sous forme d'un pseudo-code l'algorithme qui permettrait de tracer sur le graphique ci-dessus la prédiction de l'algorithme dans l'intervalle $[-8, 8]$. Votre code doit pour cela faire appel à la fonction `KNN_regression_predict` précédemment écrite et à une fonction `affiche_point(x,y)` qui affiche un tout petit point en (x,y) .

3.4 (3 pts) Nommez une autre technique de régression non-paramétrique qu'on pourrait utiliser et qui ne donnerait pas une courbe en escalier (en paliers).

4 Classifieur linéaire (33 pts)

On considère l'utilisation d'un classifieur linéaire (ex: Perceptron) pour un problème de classification binaire (2 classes: +1 et -1).

4.1 (4 pts)

Donnez l'expression mathématique d'une fonction discriminante linéaire:

$$g(x) =$$

ainsi que la fonction de classification correspondante (qui va donner une classe +1 ou -1):

$$h(x) =$$

4.2 (4 pts) Précisez quels sont les *paramètres* de la fonction g (apparaissant dans l'expression que vous avez donnée ci-dessus) qu'on va vouloir apprendre. Précisez leur nom ainsi que leur *dimensionnalité*.

4.3 (5 pts)

Donnez une définition en français puis une définition mathématique pour les ensembles suivants:

a) la région de décision de la classe positive:

b) la frontière de décision:

c) Comment appelle-t-on l'objet géométrique qui correspond aux frontières de décision obtenues avec un classifieur linéaire?

4.4 (4 pts) Quand dit-on qu'un ensemble de données D_n est linéairement séparable? Expliquez-le d'abord dans vos propres mots, puis donnez une expression mathématique pour caractériser cette propriété.

4.5 (6 pts) Illustrez sur un graphique un ensemble de données 2D ($x \in \mathbb{R}^2$) de classification binaire, **linéairement séparable**, comportant 3 points de la classe $+$ et 3 points de la classe $-$ (les 3 points de chaque classe doivent être non alignés). Tracez en **trait plein** la frontière de décision qu'on obtiendrait avec l'algorithme du **1 plus proche voisin**. Hachurez ou grisez légèrement la région de décision des $+$. Tracez ensuite en **trait pointillé** un exemple de frontière de décision qu'on pourrait obtenir sur ce problème avec un algorithme de type **classifieur linéaire** (fonction discriminante linéaire).

4.6 (6 pts) Même question que la précédente, mais avec un ensemble de données **non linéairement séparable**.

4.7 (4 pts) Quelle technique simple peut-on utiliser pour construire un classifieur non-linéaire en utilisant un algorithme de classification linéaire? Expliquez-en brièvement le principe (en Français dans vos propres mots).