

Nom de l'étudiant: \_\_\_\_\_

---

FACULTE DES ARTS ET DES SCIENCES  
DEPARTEMENT D'INFORMATIQUE ET DE RECHERCHE OPERATIONNELLE

TITRE DU COURS: **Algorithmes d'apprentissage**

SIGLE DU COURS : **IFT6266 A06**

NOM DU PROFESSEUR: Yoshua Bengio

DATE DE L'EXAMEN FINAL A06: 7 décembre 2006 HEURE: 13h - 14h55

SALLE: PAA-1411

DIRECTIVES PÉDAGOGIQUES: - Documentation permise.

- Répondre directement sur le questionnaire. Vous pouvez utiliser l'arrière des pages aussi si vous en avez besoin.
  - Soyez brefs et précis dans vos réponses.
  - **Si vous manquez de temps: l'important est de montrer que vous avez compris le problème, plutôt que les détails de la réponse.**
  - Échanger des informations lors d'un examen (ou autres formes de tricherie) est du **plagiat**, qui est passible de sanctions allant jusqu'à l'exclusion du programme.
  - Suggestion: commencez par les questions qui vous semblent les plus faciles
-



3. On s'intéresse à l'apprentissage non-supervisé ultime, c'est à dire à modéliser la distribution jointe entre des variables  $Z_1, \dots, Z_d$ . Montrez comment on peut le faire facilement si on a à notre disposition un algorithme d'apprentissage supervisé  $A$  qui fait de l'estimation de probabilité conditionnelle, c'est à dire qui prend un ensemble d'apprentissage formé de paires  $(x_i, y_i)$  et retourne une fonction  $f$  qui estime la loi  $P(Y|X)$  (i.e. cette fonction prend une paire  $(x, y)$  en argument et retourne un estimé de  $P(y|x)$ ). La réponse est très simple, mais justifiez-là.

4. Dans l'analyse en composante principales, on part des exemples  $\{x_1, \dots, x_n\}$  avec  $x_i \in \mathbf{R}^d$ , auxquels on soustrait leur moyenne  $\bar{x}$  puis que l'on projette sur le  $k$ -ieme vecteur propre  $v_k$  de la matrice de covariance des  $x_i$ , pour obtenir la  $k$ -ieme coordonnée réduite  $\alpha_k$ . Pour le point  $x$  on a

$$\alpha_k = (x - \bar{x}) \cdot v_k.$$

- (a) Pour justifier l'utilisation de la base des vecteurs propres de la matrice de covariance  $C$ , on a dit en classe que

$$COV(\alpha_i, \alpha_j) = \delta_{ij} \lambda_i$$

où  $\lambda_k$  est la  $k$ -ieme valeur propre de  $C$ . Prouvez cette assertion. (Ci-haut on parle de covariance empirique des  $\alpha$  et des  $x$ , mais un énoncé similaire est vrai pour les véritables covariances).

- (b) On peut reconstruire  $x$  en utilisant seulement  $k$  coordonnées réduites avec la formule  $\hat{x} = \bar{x} + \sum_{j=1}^k \alpha_j v_j$ . Pour justifier le fait que l'on veut garder seulement les  $k$  vecteurs propres principaux, commencez par montrer que l'erreur de reconstruction de  $x$  par  $\hat{x}$ ,  $\|\hat{x} - x\|^2$ , est égale à  $\sum_{j=k+1}^d \alpha_j^2$ . En combinant ce résultat avec celui de la sous-question précédente, montrez que l'erreur de reconstruction moyenne égale  $\sum_{j=k+1}^d \lambda_j$ .

5. **(BONUS)** Montrez que l'algorithme des  $k$ -moyennes effectue à chaque étape un pas de Newton dans l'erreur de reconstruction. Rappel: avec des données  $x_1, \dots, x_n$ , et des prototypes  $\mu_1, \dots, \mu_k$  (avec  $x_i \in \mathbf{R}^d, \mu_i \in \mathbf{R}^d$ ), l'algorithme met itérativement à jour chaque  $\mu_i$  par la moyenne des  $x_j$  qui sont les plus proches de  $\mu_i$ . L'erreur de reconstruction est la somme des carrés des distances entre chaque  $\mu_i$  et les  $x_i$  qui lui sont associés. L'algorithme de Newton minimise un critère en l'approximant à chaque itération par sa série de Taylor jusqu'au deuxième degré, autour de la valeur courante des paramètres.