EXAMEN D'ANALYSE NUMERIQUE

Alex Aussem, Jonas Koko, Philippe Mahey

Décembre 2003

Exercice I - Considérons un réseau linéaire à 2 couches qui à chaque vecteur $\mathbf{x}_k$ présenté en entrée du modèle fait correspondre un vecteur de sortie $\mathbf{s}_k$, selon les équations :

$$s_j^k = \sum_i w_{ij} x_i^k, \quad \forall j, \forall k.$$

On cherche à ajuster les paramètres ajustables du modèle, w_{ij}, de façon réaliser *au mieux* un ensemble d'association $\{(\mathbf{x}^1, \mathbf{y}^1), \ldots, (\mathbf{x}^p, \mathbf{y}^p)\}$, c'est-à-dire celle qui minimise la **somme des résidus d'erreur** $\sum_k \sum_j (y_j^k - s_j^k)^2$.

Rappels : Soit $\mathbf{X}$ une matrice (n, m) quelconque, il existe une unique matrice $\mathbf{X}^\dagger$ de dimension (m, n) telle que $\mathbf{X}^\dagger\mathbf{X}$ soit le projecteur orthogonal sur $\mathrm{Ker}(\mathbf{X})^\perp$ et $\mathbf{X}\mathbf{X}^\dagger$ soit le projecteur orthogonal sur $\mathrm{Im}(\mathbf{X})$. On supposera en outre que $(\mathbf{X}^\dagger)^T = (\mathbf{X}^T)^\dagger$ et que $(\mathbf{X}^\dagger)^\dagger = \mathbf{X}$.

1. Déterminer l'expression de $\mathbf{x}^\dagger$, où $\mathbf{x}$ est un vecteur non nul.
2. Déterminer l'expression de $\mathbf{X}^\dagger$ lorsque les colonnes de $\mathbf{X}$ sont linéairement indépendantes, en se fondant sur l'expression du projecteur orthogonal sur $\mathrm{Im}(\mathbf{X})$ vu en cours. Que devient $\mathbf{X}^\dagger$ si $\mathbf{X}$ est inversible ?
3. Montrer que $\mathrm{Ker}(\mathbf{X})^\perp \subset \mathrm{Im}(\mathbf{X}^\dagger)$ et $\mathrm{Ker}(\mathbf{X}^\dagger) \subset \mathrm{Im}(\mathbf{X})^\perp$. En déduire $\mathrm{Ker}(\mathbf{X}^\dagger) = \mathrm{Im}(\mathbf{X})^\perp$ et $\mathrm{Im}(\mathbf{X}^\dagger) = \mathrm{Ker}(\mathbf{X})^\perp$.
4. Montrer qu'on peut formuler le problème ci-dessus comme un problème de moindres carrés matriciel suivant : Trouver $\mathbf{W}$ qui minimise $\|\mathbf{Y} - \mathbf{W}\mathbf{X}\|^2$ où on explicitera la norme matricielle et les matrices $\mathbf{W}, \mathbf{X}$ et $\mathbf{Y}$.
5. Montrer que $\mathbf{Y} = \mathbf{W}\mathbf{X}$ admet (au moins) une solution si $\mathbf{Y}\mathbf{X}^\dagger\mathbf{X} = \mathbf{Y}$.
6. Inversement, établir que si $\mathbf{Y} = \mathbf{W}\mathbf{X}$ admet une solution, alors $\mathrm{Ker}(\mathbf{X}) \subset \mathrm{Ker}(\mathbf{Y})$. En déduire que si $\mathbf{Y} = \mathbf{W}\mathbf{X}$ admet une solution, alors $\mathbf{Y}\mathbf{X}^\dagger\mathbf{X} = \mathbf{Y}$.
7. Montrer que l'ensemble des matrices solution de $\mathbf{W}\mathbf{X} = \mathbf{0}$, où $\mathbf{0}$ est la matrice nulle, est de la forme $\mathbf{W} = \mathbf{Z}(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger)$ où $\mathbf{Z}$ est une matrice quelconque. En déduire la forme générale des solutions de $\mathbf{Y} = \mathbf{W}\mathbf{X}$ lorsque $\mathbf{Y}\mathbf{X}^\dagger\mathbf{X} = \mathbf{Y}$.
8. Montrer que la solution de norme minimale est $\mathbf{W} = \mathbf{Y}\mathbf{X}^\dagger$ au sens de la norme matricielle suivante : $\|\mathbf{W}\|^2 = \mathrm{Trace}(\mathbf{W}\mathbf{W}^T)$.
9. Calculer $< \mathbf{W}\mathbf{X}, \mathbf{Y} - \mathbf{Y}\mathbf{X}^\dagger\mathbf{X} >$. En déduire que $\mathbf{W} = \mathbf{Y}\mathbf{X}^\dagger$ réalise le minimum de $\|\mathbf{Y} - \mathbf{W}\mathbf{X}\|^2$. Que devient ce résultat lorsque $\mathbf{X}$ est inversible ? Interpréter.
10. Etablir que $\|\mathbf{Y} - \mathbf{W}\mathbf{X}\|^2 = \sum_j \|\mathbf{y}_j - \mathbf{X}^T\mathbf{w}_j\|^2$ avec $\mathbf{y}_j = (y_j^1, \ldots, y_j^p)^T$ et $\mathbf{w}_j = (w_{1j}, \ldots, w_{pj})^T$. Retrouver la solution $\mathbf{W} = \mathbf{Y}\mathbf{X}^\dagger$ en minimisant chaque terme de la somme. On supposera que les lignes de $\mathbf{X}$ sont linéairement indépendantes.
11. Si on adjoint à $\mathbf{X}$ un vecteur $\mathbf{x}$ de taille n tel que dans $\mathbf{x} \notin \mathrm{Im}(\mathbf{X})$, montrer que
$$(\mathbf{X}, \mathbf{x})^\dagger = \begin{pmatrix} \mathbf{X}^\dagger(\mathbf{I} - \mathbf{x}\mathbf{u}^\dagger) \\ \mathbf{u}^\dagger \end{pmatrix}$$
avec $\mathbf{u} = (\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger)\mathbf{x}$. Quelle signification donner à $\mathbf{u}$?
12. En déduire que la nouvelle matrice de connexion vérifie la formule de récurrence, $\mathbf{W}_{k+1} = \mathbf{W}_k + (\mathbf{y} - \mathbf{W}_k\mathbf{x})\mathbf{u}^\dagger$.