
TD 5

Exercice I

Soit $S = \{(x_i, y_i), 1 \leq i \leq n\}$, $x_i \in \mathbb{R}^d$ et $y_i \in \mathbb{R}^p$ un ensemble de données d'apprentissage. Soit X la matrice de taille $n \times d$ telle que $\mathrm{X}_{i,j} = (x_i)_j$ et Y la matrice de taille $n \times p$ telle que $\mathrm{Y}_{i,j} = (y_i)_j$.

1. Calculer W_{LS} la solution du problème de régression multivariée utilisant la méthode des moindres carrés.
2. Montrer que
$$\|Y - XW\|^2 = \|Y - XW_{LS}\|^2 + \|XW_{LS} - XW\|^2. \quad (1)$$
(Aide : $M = X(X^\top X)^{-1}X^\top$ est une matrice idempotente, c.a.d $M^2 = M$)

Exercice II

Le théorème d'Eckart-Young permet d'obtenir la meilleure approximation d'une matrice par une matrice de rang faible dans le cas où on désire minimiser la distance au sens de la norme de Frobenius. Plus formellement, la solution B^* du problème

$$B^* = \arg\min_B \|A - B\|^2 \quad \text{s.t.} \quad rang(B) = r,$$

est la matrice $B^* = U\Sigma_r V$, où $A = U\Sigma V$ est la décomposition en valeurs singulières de A et Σ_r contient que les r plus grandes valeurs singulières, les autres étant remplacées par 0.

1. Ecrire le problème de régression à rang faible.
2. Utilisant le théorème d'Eckart-Young et Equation 1, donner la solution du problème de régression à rang faible.
3. Proposer un moyen pour résoudre le problème de régression à rang faible régularisée dans le cas d'une régularisation L_2.

Exercice III

En apprentissage multi-tâche, on dispose de T jeux de données $S^t = \{(x_i^t, y_i^t), 1 \leq i \leq n^t\}$, $x_i^t \in \mathbb{R}^d$ et y_i^t un scalaire pour $t = 1, \ldots, T$, avec T le nombre de tâches.

1. Dans le cas d'une fonction coût de moindres carrées, formaliser le problème d'apprentissage multi-tâche comme étant un problème de régression multivariée. (Aide : utiliser un masque, c.a.d. une matrice qui contient des valeurs égales à 0 ou 1, pour identifier les positions des valeurs manquantes dans les sorties vectorielles)

2. Une façon de traiter le problème de classification multi-classe est de coder chaque classe (y_i) avec un vecteur au lieu d'une valeur réelle (numéro de la classe). Par exemple pour un problème multi-classe de 3 classes, on peut considèrer le codage suivant : $y_i = (1, 0, 0)$ (respectivement, $y_i = (0, 1, 0)$, $y_i = (0, 0, 1)$) si x_i est dans la classe 1 (respectivement, classe 2, classe 3). Donner une stratégie de classification multi-classe qui correspond à résoudre le problème multi-tâche suivant

$$\min_{f_1, f_2, f_3} \left\{ \frac{1}{n} \sum_{j=1}^{3} \sum_{i=1}^{n} (y_i^j - f^j(x_i))^2 + \lambda \|f^j\|^2 \right\}.$$