Les modèles linéaires

Exercices

Exercice : 1

Dans la régression linéaire, les prévisions empiriques sont données par $\hat{y} = Hy$, où :

$$H = X(X^T X)^{-1} X^T$$

Montrer que H est une matrice de projection (H^2=H). Donc, on a $\hat{y}$ est une projection y dans un espace. Qu'appelle-t-on cet espace ?

Exercice : 2

(a) Soit :

$$E_i(w) = max\big(0, 1 - y_i(w^T x_i)\big)$$

Montrer que si $y_i = w^T x_i$ on a $E_i(w)$ est continue et différentiable.

(b) Montrer que $\boldsymbol{E_i}(w)$ est une borne supérieur du $[sign(w^T x_i) \neq y_i]$. Donc, $\frac{1}{n}\sum_{i=1}^{n} E_i(w)$ est une borne supérieure de l'erreur de classification empirique $E_{emp}(w)$.

(c) Appliquer l'algorithme stochastique de la descente de gradient sur $\frac{1}{n}\sum_{i=1}^{n} E_i(w)$ (ignorer le cas où $w^T x_i = y_i$) et déduire un nouveau algorithme d'entrainement du Perceptron.

Exercice : 3

Extraire l'algorithme de programmation linéaire pour capturer un modèle linéaire de classification en utilisant les étapes suivantes. Un programme linéaire est un problème d'optimisation de la forme suivante :

$$min_z \ c^T z$$

Tel que : $Az \leq b$

A, b et c sont les paramètres du programme linéaire et z est une variable d'optimisation.

(a) Pour des données linéairement séparables, montrer que pour quelques w on a $y_i(w^T x_i) \geq 1$, pour $i = 1, \ldots, n$.

(b) Formuler le problème de programmation linéaire qui permet de trouver les valeurs de w pour des données linéairement séparables. Vous devez spécifier les paramètres de A, b, c et z.

(c) Si les données sont non-séparables, la condition dans (a) n'est plus valable pour tout i. Cela introduit la violation $\xi_i \geq 0$ pour capturer la quantité de violation pour x_i. Pour $i = 1, \dots, n$:
$y_i(w^T x_i) \geq 1 - \xi_i$
En effet, on doit minimiser la quantité de violation. Pour cela, il faut minimiser $\sum_{i=1}^{n} \xi_i$, donc, on veut trouver w qui permet de résoudre :

$$min_{w,\, \xi_i} \quad \sum_{i=1}^{n} \xi_i$$

$$\text{Tel que : } y_i(w^T x_i) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

Et cela pour tout $i = 1, \dots, n$. Formuler ce problème en un programme linéaire.

(d) Montrer que le programme linéaire trouvé dans (c) et le problème d'optimisation de l'exercice 2 sont équivalents.

Exercice : 4

Dans les paramètres de la régression, supposant que la fonction cible est linéaire, donc :

$$f(x) = x^T w_f$$

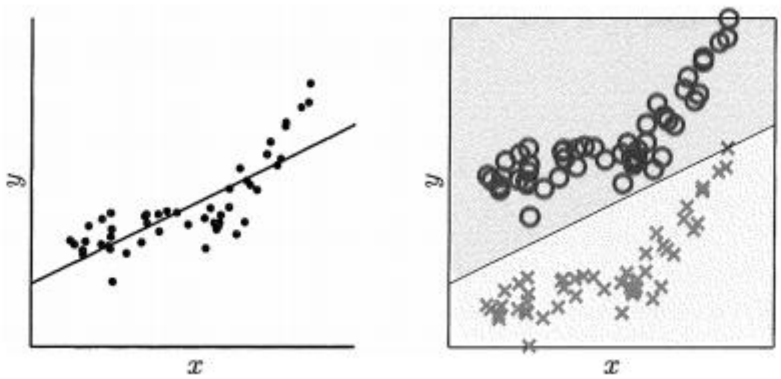
Et $y = Xw_f + \epsilon$, oû la moyenne de ϵ est 0 et sa variance égale à σ^2. Dans ce problème vous allez déduire le bias et la variance comme suit :

(a) Montrer que la fonction moyenne $\bar{g}(x) = f(x)$, quel que soit la taille des données, tant que $X^T X$ est inversible. Quel est le bias ?

(b) Quelle est la variance ?

Exercice : 5

Ce problème crée un algorithme de régression linéaire à partir d'un bon algorithme de classification linéaire. Comme illustré dans les figures suivantes, l'idée c'est de prendre les données originales et les décaler dans une direction pour avoir les points de données +1 ; puis les décaler dans l'autre direction pour avoir les points de données -1.



La figure à gauche présente les données originales pour un problème de régression à une seule dimension. La figure à droite, présente les données décalées comme un problème de classification à deux dimensions.

Généralement les données (x_n, y_n) peuvent être vues comme des points de données dans R^{d+1} en traitant les valeurs y comme étant la coordonnée d+1.

Construire les points de données positifs et négatifs :

$$D_+ = (x_1, y_1) + a, \dots, (x_n, y_n) + \text{a}$$

$$D_- = (x_1, y_1) - a, \dots, (x_n, y_n) - a$$

Où "a" est un paramètre de perturbation. Utiliser maintenant l'algorithme de programmation linéaire de l'exercice 3 pour séparer D_+ de D_-. L'hyperplan de séparation résultant peut être utilisé comme une adaptation de régression pour les données originales.

(a) Combien de poids sont appris dans le problème de classification ? Combien on a besoin de poids pour l'adaptation linéaire dans le problème de la régression ?

(b) L'hypothèse d'adaptation linéaire exige des poids w tel que : $f(x) = w^T x$. Supposons que les poids retournés par la résolution du problème de classification sont w_{class}. Donner une expression pour w en fonction de w_{class}.

(c) Générer l'ensemble de données $y_n = x_n^2 + \sigma\epsilon_n$ avec n=50, où x_n est uniforme [0,1] et ϵ_n est un bruit Gaussien possédant une moyenne égale à zéro ; soit $\sigma = 0.1$. Dessiner D_- et D_+ pour : $a = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$.

(d) Comparer les courbes adaptatives résultantes à partir de l'utilisation de l'approche de classification et de l'algorithme analytique de pseudo-inverse pour la régression linéaire.