

TRAVAUX DIRIGÉS DE
MACHINE LEARNING
CYCLE PLURIDISCIPLINAIRE D'ÉTUDES SUPÉRIEURES
UNIVERSITÉ PARIS SCIENCES ET LETTRES

Joon Kwon

mercredi 6 avril 2022



On se place dans un cadre de régression avec $\mathcal{X} = \mathbb{R}^d$ et $\mathcal{Y} = \mathbb{R}$, et la perte quadratique :

$$\forall y, y' \in \mathbb{R}, \quad \ell(y, y') = \frac{1}{2}(y - y')^2.$$

Pour $w \in \mathbb{R}^d$, on note h_w le prédicteur défini par :

$$\forall x \in \mathbb{R}^d, \quad h_w(x) = \langle w, x \rangle.$$

On considère la classe de prédicteurs $\mathcal{F} = \{h_w\}_{w \in \mathbb{R}^d}$. Soit $n \geq 1$ un entier et $S = (x_i, y_i)_{i \in [n]} \in \mathcal{S}(\mathcal{X}, \mathcal{Y})$ un échantillon d'apprentissage. Pour $i \in [n]$, on note x_{ij} ($1 \leq j \leq d$) les composantes de x_i . On pose :

$$A = \sum_{i=1}^n x_i x_i^\top \quad \text{et} \quad b = \sum_{i=1}^n y_i x_i,$$

où x_i^\top désigne la transposée de x_i , ce dernier étant vu comme un vecteur colonne.

1) Donner une expression simple des coefficients de la matrice A .

On s'intéresse à la minimisation du risque empirique avec régularisation ℓ_2 et un paramètre de régularisation $\lambda > 0$: elle donne le prédicteur $h_{\hat{w}}$ où \hat{w} est défini par :

$$\hat{w} = \arg \min_{w \in \mathbb{R}^d} F(w),$$

et où $F : \mathbb{R}^d \rightarrow \mathbb{R}$ est la fonction définie par

$$\forall w \in \mathbb{R}^d, \quad F(w) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (y_i - h_w(x_i))^2 + \frac{\lambda}{2} \|w\|_2^2.$$

- 2) Donner une expression du gradient de F faisant intervenir A et b .
- 3) En déduire une équation matricielle vérifiée par \hat{w} .
- 4) Montrer que les valeurs propres de A sont positives.
- 5) Montrer que la matrice $A + \lambda n I_d$ (où I_d désigne la matrice identité) est inversible.
- 6) En déduire une expression matricielle pour \hat{w} .

