Correction des travaux dirigés de
Machine learning
Cycle pluridisciplinaire d'études supérieures
Université Paris sciences et lettres

Joon Kwon

vendredi 24 avril 2020

1) Pour $i \in [n]$, on a

$$x_i x_i^\top = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{id} \end{pmatrix} (x_{i1}, \dots, x_{id}) = (x_{ij}x_{ik})_{1 \leqslant j,k \leqslant d}.$$

Donc

$$A = \sum_{i=1}^n x_i x_i^\top = \left(\sum_{i=1}^n x_{ij} x_{ik} \right)_{1 \leqslant j,k \leqslant d}.$$

2) Pour $w \in \mathbb{R}^d$, F peut s'écrire :

$$F(w) = \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^d w_j x_{ij} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^d w_j^2.$$

F est donc différentiable par théorèmes d'opérations. Pour $w \in \mathbb{R}^d$ et $1 \leqslant j \leqslant$

d, on a

$$\begin{aligned}
\frac{\partial F}{\partial w_j}(w) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(-2x_{ij})\left(y_i - \sum_{j'=1}^{d} w_{j'}x_{ij'}\right) + \frac{\lambda}{2}2w_j \\
&= \frac{1}{n}\sum_{j'=1}^{d} w_{j'}\sum_{i=1}^{n} x_{ij}x_{ij'} - \frac{1}{n}\sum_{i=1}^{n} y_i x_{ij} + \lambda w_j \\
&= \frac{1}{n}(\mathrm{A}w)_j - \frac{1}{n}b_j + \lambda w_j.
\end{aligned}$$

Donc

$$\nabla F(w) = \left(\frac{\partial F}{\partial w_j}(w)\right)_{1\leqslant j\leqslant d} = \frac{1}{n}(\mathrm{A}w - b) + \lambda w.$$

3) $\hat{w}$ étant un minimiseur de la fonction F qui est différentiable sur $\mathbb{R}^d$ tout entier, on a $\nabla F(\hat{w}) = 0$, autrement dit :

$$\mathrm{A}\hat{w} - b + \lambda n\hat{w} = 0.$$

4) Soit μ une valeur propre de A et $u \neq 0$ un vecteur propre associé, autrement dit :

$$\mathrm{A}u = \mu u.$$

En multipliant à gauche par $u^\top$:

$$u^\top \mathrm{A}u = \mu u^\top u.$$

D'une part on a que

$$u^\top u = \sum_{j=1}^{d} u_j^2 > 0$$

et d'autre part :

$$u^\top \mathrm{A}u = u^\top\left(\sum_{i=1}^{n} x_i x_i^\top\right)u = \sum_{i=1}^{n} u^\top x_i x_i^\top u = \sum_{i=1}^{n}(x_i^\top u)^\top(x_i^\top u) = \sum_{i=1}^{n}\sum_{j=1}^{d}(x_i^\top u)_j^2 \geqslant 0.$$

Nécessairement, on a $\mu \geqslant 0$.

5) Montrons qu'aucune valeur propre de $A + \lambda n I_d$ n'est nulle, ce qui équivaut à dire que la matrice est inversible. Soit μ une valeur propre de $A + \lambda n I_d$ et $u \neq 0$ un vecteur propre associé. On a alors

$$(A + \lambda n I_d)u = \mu u$$

ce qui se réécrit :

$$Au = (\mu - \lambda n)u.$$

$\mu - \lambda n$ est donc une valeur propre de A et est donc positive d'après la question précédente. Donc

$$\mu \geqslant \lambda n > 0,$$

car $\lambda > 0$ par hypothèse. En particulier $\mu \neq 0$.

6) L'équation obtenue à la question 3 peut s'écrire

$$(A + \lambda n I_d)\hat{w} = b.$$

Donc

$$\hat{w} = (A + \lambda n I_d)^{-1} b.$$