Corrigé

DS 01 ING2 EILCO/FISEA - Ingénierie Mathématique

Octobre 2024

Nom :
Prénom :

Total: 15 points
Durée: 1h
Instructions générales: Vous êtes libres de rédiger vos réponses sur des pages supplémentaires en veillant toutefois à bien indiquer le numéro de chaque question. Une fois le DS terminé, Assurez vous de bien écrire votre nom (de façon lisible) sur chacune des pages. Répondez à un maximum de questions, en commençant par les questions qui vous semblent les plus abordables.

Question 1 (5pts) *Indiquer si les affirmations suivantes sont vraies ou fausses*

Vrai / Faux — *Dans le cadre d'une régularisation basée sur les normes ℓ_p, plus p est grand, plus la boule associée sera concentrée le long des axes*

Vrai / Faux — *Pour tout $\lambda > 0$, le minimum de la formulation de type Ridge peut être déterminé à l'aide des équations normales.*

Vrai / Faux — *L'estimateur de maximum de vraissemblance correspond à un estimateur de maximum a posteriori pour un a priori uniforme sur les coefficients de régression.*

Vrai / Faux — *Pour tout $j > 0$, l'itération de descente de gradient pour une fonction de coût de type Ridge est donnée par*

$$\beta_j \leftarrow \beta_j + \frac{2\eta}{m}\sum_{i=1}^{m}\left(t^{(i)} - (\beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_D x_D^{(i)})\right)x_j^{(i)} + 2\lambda\eta\beta_j$$

Vrai / Faux — *La règle de Bayes est donnée par $P(A|B) = P(B|A)P(B)/P(A)$*

Vrai / Faux — *Dans la décomposition biais-variance de l'erreur quadratique moyenne, le biais s'écrit $\mathbb{E}_{\mathcal{D}_i}\{h_{\mathcal{D}_i}(\boldsymbol{x}) - \mathbb{E}_{\mathcal{D}_i}\{h_{\mathcal{D}_i}(\boldsymbol{x})\}\}$ où $h_{\mathcal{D}_i}(\boldsymbol{x})$ représente la prédiction du modèle $h(\boldsymbol{x})$ lorsque celui-ci est entrainé sur les données du sous-ensemble $\mathcal{D}_i$*

Vrai / Faux — *La solution d'une régularisation de type Ridge correspond à un estimateur de maximum de vraissemblance avec un a-priori Gaussien*

Question 2 (3pts)

[3pts] *On considère les paires $(x^{(i)}, t^{(i)})$ suivantes: $(x^{(i)}, t^{(i)}) \in \{(1,1),(2,2)\}$. On souhaite entrainer un modèle linéaire $h_\beta(x) = \beta_0 + \beta_1 x$ de façon à ce que $h_\beta(x^{(i)}) = t^{(i)}$. En utilisant les équations normales, déterminer la valeur des coefficients β_0, β_1.*

Pour rappel, l'inverse d'une matrice 2×2 de la forme $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ est donnée par $\boldsymbol{A}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Question 3 (4pts) *En utilisant les mêmes données qu'à la question précédente, on souhaite à présent entrainer le modèle à l'aide d'une formulation de type Ridge. Soit $\lambda = .1$ et la fonction de coût*

$$\ell(\beta_0, \beta_1) = \frac{1}{2}\sum_{i=1}^{2}\left(t^{(i)} - (\beta_0 + \beta_1 x)\right)^2 + \lambda\beta_1^2.$$

Question 2 (solution)

On a $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$ $\quad t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

les equations normales sont données par

$$X^T X \beta = X^T t$$

soit $\beta = (X^T X)^{-1} X^T t$

en utilisant les expressions de X et t donnés plus haut,

$$X^T X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$X^T t = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

en utilisant l'expression de l'inverse d'une matrice 2×2,

$$(X^T X)^{-1} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

soit $\quad \beta = (X^T X)^{-1} X^T t = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Question 3 (solution)

On a $t_c = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{1}{2} \cdot 3 = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$

$$x_c = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{3}{2} = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$$

Question 3 (Solution - suite)

La formulation Ridge est donnée par

$$\frac{1}{2}\sum_{i=1}^{2}\left(t_c^{(i)} - x_c^{(i)}\beta_1 - \beta_0\right)^2 + \lambda\beta_1^2$$

En prenant la dérivée par rapport à β_0 on trouve par exemple

$$\frac{\partial}{\partial\beta_0}\left[\frac{1}{2}\sum_{i=1}^{2}\left(t_c^{(i)} - x_c^{(i)}\beta_1 - \beta_0\right)^2 + \lambda\beta_1^2\right]$$

$$= -\sum_{i=1}^{2}\left(t_c^{(i)} - x_c^{(i)}\beta_1 - \beta_0\right)$$

$$= -\left[\left(-\frac{1}{2} + \beta_1/2 - \beta_0\right) + \left(1/2 - \frac{\beta_1}{2} - \beta_0\right)\right] = 2\beta_0$$

ce qui, en fixant $\frac{\partial \ell}{\partial\beta_0} = 0$ donne bien $\beta_0 = 0$

Pour β_1, un raisonnement semblable donne

$$\frac{\partial}{\partial\beta_1}\left[\frac{1}{2}\sum_{i=1}^{2}\left(t_c^{(i)} - x_c^{(i)}\beta_1\right)^2 + \lambda\beta_1^2\right] = 0$$

$$\Leftrightarrow \quad 2\lambda\beta_1 + \left[\left(-1/2 + \beta_1/2\right)(1/2) + \left(1/2 - \beta_1/2\right)(-1/2)\right] = 0$$

$$\Leftrightarrow \quad 2\lambda\beta_1 - 1/2 + \beta_1/2 = 0$$

$$\Leftrightarrow \quad \beta_1\left(2\lambda + 1/2\right) = 1/2$$

$$\Leftrightarrow \quad \beta_1 = \frac{1}{2}\frac{1}{\left(2\lambda + 1/2\right)}$$

Question 4 (Solution) : Ridge, Ridge, fit, predict, Xpoly-test