Travaux Dirigés n°2
Régression Logistique

KMAXPP05

2025-2026

Exercice 1

On considère la fonction logistique :

$$\sigma : \mathbb{R} \to]0,1[$$
$$\sigma : t \mapsto \frac{1}{1+e^{-t}}$$

La fonction logistique est une fonction *sigmoïde*. Littéralement, c'est une fonction avec une forme de S. Mathématiquement, une fonction sigmoïde est croissante, continue et dérivable et admet deux asymptotes horizontales en $-\infty$ et $+\infty$

1. Montrer que la dérivée de la fonction $\sigma(t)$ est $\sigma(t)\,(1-\sigma(t))$
2. Montrer que σ est strictement croissante sur $\mathbb{R}$.
3. Montrer que $\lim_{t\to+\infty} \sigma(t) = 1$ et que $\lim_{t\to-\infty} \sigma(t) = 0$.
4. Quelle autre fonction (trigonométrique) est également une sigmoïde ?
5. Montrer que σ possède la propriété suivante $\sigma(t) = 1 - \sigma(-t) \quad \forall t \in \mathbb{R}$.
6. Déduire de la propriété précédente que $\sigma(0) = \frac{1}{2}$ et que $\sigma(t) \geq \frac{1}{2} \iff t \geq 0$.
7. Tracer approximativement la courbe de σ sur $\mathbb{R}$.

Exercice 2

On cherche à construire un classifieur C, basé sur une régression logistique, qui prend en entrée $x \in \mathbb{R}$ et qui le classifie en la classe 0 ou la classe 1. Ce classifieur va être paramétré par un vecteur $\beta \in \mathcal{M}_{2,1}(\mathbb{R})$ de composantes β_0 et β_1. La classe associée sera 1 si $C_\beta(x) \geq 1/2$ et 0 sinon. Le classifieur C est défini comme suit :

$$\begin{aligned} C_\beta(x) &= \sigma(\beta_0 + \beta_1 x) \\ &= \frac{1}{1+\exp(-\beta_0 - \beta_1 x)} \end{aligned}$$

Les réponses de l'exercice 1 peuvent être utilisées pour les questions de cet exercice.

1. Pour $\beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ quels x sont classés en 0 et quels x sont classés en 1 ?
2. Même question pour $\beta = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
3. Pour un $x \neq 0$ et β_0 fixés, lorsque β_1 devient assez grand ($\beta_1 \to +\infty$), comment est classé x ? Même question lorsque $\beta_1 \to -\infty$.

Exercice 3

Nous disposons des données $\mathcal{D}$ suivantes pour des joueurs de rugby. L'espace des caractéristiques d'entrée est défini par le poids x et noté $\mathcal{X} \subseteq \mathbb{R}$. L'espace des caractéristiques de sortie est défini par la position y noté $\mathcal{Y} = \{0, 1\}$:

Poids (kg)	**Position**
112	avant = 1
85	trois-quart = 0
135	avant = 1
92	trois-quart = 0

On suppose que les $n = 4$ observations sont indépendantes. Nous cherchons à établir un classifieur basé sur un modèle de régression logistique de la forme suivante, avec $\beta_0 \in \mathbb{R}$ et $\beta_1 \in \mathbb{R}$:

$$C_\beta(x) = \begin{cases} 0 & \forall x \quad \dfrac{1}{1 + \exp\left(-\beta_0 - \beta_1\, x\right)} < \dfrac{1}{2} \\ 1 & \forall x \quad \dfrac{1}{1 + \exp\left(-\beta_0 - \beta_1\, x\right)} \geq \dfrac{1}{2} \end{cases}$$

On note $\beta \in \mathcal{M}_{2,1}(\mathbb{R})$ le vecteur :

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

On associe à chaque observation i une variable aléatoire $Y^{(i)}$ suivant une loi de Bernoulli de paramètre $p^{(i)} = \sigma(\tilde{x}^{(i)}\beta)$:

$$\begin{cases} P(Y^{(i)} = 1) = p^{(i)} \\ P(Y^{(i)} = 0) = 1 - p^{(i)} \end{cases}$$

Les réponses de l'exercice 1 peuvent être utilisées pour les questions de cet exercice.

1. Quelle application $\Phi : \mathcal{X} \to \tilde{\mathcal{X}} \subseteq \mathbb{R}^2$ permet-elle de calculer la probabilité pour l'observation i sous la forme $\sigma(\tilde{x}^{(i)}\beta)$ où σ est la fonction logistique. Quelle est la matrice de design $\tilde{X}$ pour ce problème de classification ?
2. Vérifier que la probabilité de la loi de Bernoulli pour la variable aléatoire $Y^{(i)}$ peut s'écrire $P(Y^{(i)} = y^{(i)}) = \left(p^{(i)}\right)^{y^{(i)}} \left(1 - p^{(i)}\right)^{\left(1 - y^{(i)}\right)}$.
3. Pour une valeur fixe de β, quelle est la vraisemblance sur $\mathcal{D}$ selon les variables aléatoires de Bernoulli introduites ci-dessus ?
4. Calculer la log-vraisemblance négative moyenne sur $\mathcal{D}$. Sous quelle autre dénomination usuelle cette quantité est-elle aussi connue ?
5. Quelle condition doit satisfaire β pour être la meilleure estimation de ce problème de régression logistique ?
6. On rappelle que $\tilde{x}^{(i)}$ est le vecteur ligne de l'observation i dans l'espace augmenté $\tilde{\mathcal{X}}$. On pose $t^{(i)} = \tilde{x}^{(i)}\beta = \beta_0 + \beta_1 x^{(i)}$. Démontrer que $\dfrac{dt^{(i)}}{d\beta} = \tilde{x}^{(i)T}$.
7. Démontrer que $\dfrac{d}{dt^{(i)}} \ln\left[\sigma\left(t^{(i)}\right)\right] = 1 - \sigma\left(t^{(i)}\right)$.
8. En déduire que $\dfrac{d}{d\beta} \ln\left[\sigma\left(\tilde{x}^{(i)}\beta\right)\right] = \left[1 - \sigma\left(\tilde{x}^{(i)}\beta\right)\right] \tilde{x}^{(i)T}$.

9. Démontrer que $\dfrac{d}{dt^{(i)}} \ln\left[1-\sigma\left(t^{(i)}\right)\right] = -\sigma\left(t^{(i)}\right)$.
10. En déduire que $\dfrac{d}{d\beta} \ln\left[1-\sigma\left(\tilde{x}^{(i)}\beta\right)\right] = -\sigma\left(\tilde{x}^{(i)}\beta\right)\tilde{x}^{(i)T}$.
11. Démontrer que le gradient de la log-vraisemblance négative moyenne par rapport à β est égal à $-\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(y^{(i)}-\sigma(\tilde{x}^{(i)}\beta)\right)\tilde{x}^{(i)T}$.
12. Démontrer que ce même gradient peut s'écrire sous la forme suivante : $-\dfrac{1}{n}\tilde{X}^T\left(Y-\sigma(\tilde{X}\beta)\right)$.
13. On fixe $\beta = \begin{pmatrix} -90 \\ 1 \end{pmatrix}$. On pourrait calculer que le gradient de la log-vraisemblance négative moyenne est à peu près égal à $\begin{pmatrix} 0,22 \\ 20,4 \end{pmatrix}$. Que déduire de cette valeur de β ?
14. En gardant la même valeur de β, quel est le poids à partir duquel le classifieur va prédire une classe 1 ? En gardant $\beta_1 = 1$, quelle serait une meilleure alternative pour β_0 afin d'obtenir un meilleur classifieur ? Justifier votre choix pour cette alternative.
15. La figure ci-dessous permet de visualiser le logarithme base 10 de la norme du gradient. Le $\beta = \begin{pmatrix} -90 \\ 1 \end{pmatrix}$ de la question précédente est représenté. En prenant $\beta_1 = 1.0$ quelle valeur de β_0 permet de minimiser le gradient ? Quelle information cette visualisation du gradient apporte-t-elle ?