

Feuille de TD 5 : Bases du ML et de la classification

Exercice 1 (Régression logistique)

Dans un contexte de classification binaire, on considère *modèle logistique*, défini par

$$\mathbb{P}[Y = 1|X = x] = \frac{\exp(\langle \beta^*, x \rangle)}{1 + \exp(\langle \beta^*, x \rangle)} \quad \text{pour tout } x \in \mathbb{R}^d, \quad (1)$$

avec $\beta^* \in \mathbb{R}^d$. Dans ce cas, on a $\mathbb{P}[Y = 1|X = x] > 1/2$ si et seulement si $\langle \beta^*, x \rangle > 0$, de sorte que la frontière de décision du prédicteur de Bayes est un hyperplan, de direction orthogonale β^* .

On peut estimer le paramètre β^* en maximisant la vraisemblance conditionnelle de Y sachant X :

$$\hat{\beta} \in \operatorname{argmax}_{\beta \in \mathbb{R}^d} \prod_{i=1}^n \left[\left(\frac{\exp(\langle \beta, X_i \rangle)}{1 + \exp(\langle \beta, X_i \rangle)} \right)^{Y_i} \left(\frac{1}{1 + \exp(\langle \beta, X_i \rangle)} \right)^{1-Y_i} \right],$$

et construire le classifieur $\hat{h}_{\text{logistic}}(x) = \mathbf{1}_{\langle \hat{\beta}, x \rangle > 0}$ pour tout $x \in \mathbb{R}^d$.

Notre objectif est maintenant de calculer des bornes de confiance pour β^* .

1. Vérifier que le gradient et la Hessienne $H_n(\beta)$ de

$$\ell_n(\beta) = - \sum_{i=1}^n [Y_i \langle X_i, \beta \rangle - \log(1 + \exp(\langle X_i, \beta \rangle))]$$

sont donnés par¹

$$\nabla \ell_n(\beta) = - \sum_{i=1}^n \left(Y_i - \frac{e^{\langle X_i, \beta \rangle}}{1 + e^{\langle X_i, \beta \rangle}} \right) X_i \quad \text{et} \quad H_n(\beta) = \sum_{i=1}^n \frac{e^{\langle X_i, \beta \rangle}}{(1 + e^{\langle X_i, \beta \rangle})^2} X_i X_i^\top.$$

On suppose $H_n(\beta)$ non singulière. Que peut-on dire de la fonction ℓ_n ?

2. Montrer qu'il existe $\tilde{\beta}$ tel que $\|\tilde{\beta} - \beta^*\| \leq \|\hat{\beta} - \beta^*\|$ et

$$\hat{\beta} - \beta^* = -H_n(\tilde{\beta})^{-1} \nabla \ell_n(\beta^*).$$

Dans la suite, on suppose que les X_i sont uniformément bornés, que $\hat{\beta} \rightarrow \beta^*$ p.s., et qu'il existe $H(\beta)$ continue et non singulière telle que $n^{-1}H_n(\beta)$ converge vers $H(\beta)$ uniformément dans une boule centrée en β^* .

3. (Optionnel) On pose $p_i(\beta) = e^{\langle X_i, \beta \rangle} / (1 + e^{\langle X_i, \beta \rangle})$. Vérifier que

$$\begin{aligned} \mathbb{E} e^{-n^{-1/2} \langle t, \nabla \ell_n(\beta^*) \rangle} &= \prod_{i=1}^n \left(1 - p_i(\beta^*) + p_i(\beta^*) e^{\langle t, X_i \rangle / \sqrt{n}} \right) e^{-p_i(\beta^*) \langle t, X_i \rangle / \sqrt{n}} \\ &= \exp \left(\frac{1}{2} t^\top (n^{-1} H_n(\beta^*)) t + O(n^{-1/2}) \right). \end{aligned}$$

1. On préfère en général prendre $Y \in \{1, -1\}$ car la log-vraisemblance est plus sympathique : $\ell_n(\beta) = \sum_{i=1}^n \log(1 + \exp(-Y_i \langle X_i, \beta \rangle))$

4. Quelle est la loi asymptotique de $-n^{-1/2}\nabla\ell_n(\beta^*)$? de $\sqrt{n}(\hat{\beta} - \beta^*)$?
5. Proposer un intervalle de confiance $\mathcal{I}_{n,\alpha}$ tel que $\beta_j^* \in \mathcal{I}_{n,\alpha}$ avec une probabilité asymptotique $1 - \alpha$.
6. Proposer un ellipsoïde de confiance $\mathcal{E}_{n,\alpha}$ tel que la probabilité que $\beta^* \in \mathcal{E}_{n,\alpha}$ soit asymptotiquement $1 - \alpha$.
7. Proposer deux méthodes de sélection de coordonnées de x utiles pour prédire la classe d'un nouvel individu x .