Modèle linéaire et modèles dérivés, partie 2 : session 2

Exercice 1

On se place dans le modèle linéaire généralisé inverse gaussien avec lien canonique. La loi inverse gaussienne de paramètres $\lambda > 0$ et $\mu > 0$ notée $\mathrm{IG}(\mu, \lambda)$ a pour densité

$$f_{\mu,\lambda}(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp -\frac{\lambda(y-\mu)^2}{2\mu^2 y}$$

sur $\mathbb{R}_+$.

1. Écrire cette densité sous la forme d'une densité de la famille exponentielle paramétrée, comme dans le cours, par θ et ϕ.
2. Écrire θ et ϕ en fonction de λ et μ. Que valent $\mathbb{E}(Y)$ et $\mathbb{V}(Y)$ quand Y a la loi $\mathrm{IG}(\mu, \lambda)$?
3. Quel est le lien canonique dans ce modèle ? On notera β le paramètre de régression.
4. On dispose d'un échantillon $(X_i, Y_i)_{i=1,\dots,n}$ dans ce modèle. Quelle équation vérifie l'estimateur au maximum de vraisemblance $\hat{\beta}$?
5. Quelle est la loi asymptotique de $\sqrt{n}(\hat{\beta} - \beta)$
6. Donner la forme des résidus de déviance dans ce modèle.
7. Proposer un estimateur de ϕ.

Exercice 2

1. On considère un problème de classification avec pour $i = 1, \dots, n$
   - des features $X_i \in \mathbb{R}^d$
   - des labels $Y_i \in \{-1, 1\}$

   et un ensemble de "weak learners" $\mathcal{H} = \{h_1, \dots h_L\}$ où chaque $h_l : \mathbb{R}^d \to \{-1, 1\}$ est un learner très simple

   On va procéder par boosting, c'est-à-dire en combinant additivement des weak learners

   $$g^{(B)}(x) = \sum_{b=1}^{B} \eta^{(b)} h^{(b)}(x)$$

   avec $\eta^{(b)} \geq 0$ pour espérer en obtenir un meilleur. L'algorithme de boosting le plus connu en classification est AdaBoost (ADAptive BOOSTing) dont le pseudo-code est

Algorithm 1 Adaboost

```
1: Posons $p^1(i) = 1/n$ pour $i = 1, \dots, n$
2: for $b = 1, \dots B$ do
3:    $h^{(b)} \in \mathrm{argmin}_{h \in \mathcal{H}} \sum_{i=1}^n p^{(b)}(i) \mathbb{1}_{Y_i h(X_i) < 0}$
4:    $\varepsilon^{(b)} \leftarrow \sum_{i=1}^n p^{(b)}(i) \mathbb{1}_{Y_i h^{(b)}(X_i) < 0}$
5:    $\eta^{(b)} \leftarrow \frac{1}{2} \log \left( \frac{1-\varepsilon^{(b)}}{\varepsilon^{(b)}} \right)$
6:    $Z^{(b)} \leftarrow 2\sqrt{\varepsilon^{(b)}(1-\varepsilon^{(b)})}$
7:    $p^{(b+1)}(i) \leftarrow p^{(b)}(i) \frac{e^{-\eta^{(b)} Y_i h^{(b)}(X_i)}}{Z^{(b)}}$
8: end for
9: return  Un boosting classifieur $g^{(B)}(x) = \sum_{b=1}^B \eta^{(b)} h^{(b)}(x)$.
```

On veut montrer que c'est équivalent à optimiser la perte exponentielle

$$\exp(-yu).$$

(a) (Question préliminaire) Pour un classifieur donné $h \in \mathcal{H}$, montrer que, pour tout $i = 1, \dots, n$, on a $Y_i h(X_i) \in \{-1, 1\}$. Puis montrer que $\varepsilon^{(b)}$ défini à la ligne 4 de l'algorithme s'écrit aussi

$$\varepsilon^{(b)} = 1 - \sum_{i : Y_i h(X_i) = 1} p^{(b)}(i).$$

(b) A l'itération $b+1$, on cherche $\eta^{(b+1)}$ et $h^{(b+1)}$ définis par

$$h^{(b+1)}, \eta^{(b+1)} = \mathrm{argmin}_{h \in \mathcal{H}, \eta \geq 0} L_n\Big((Y_1, X_1), \dots, (Y_n, X_n); h, \eta \Big)$$

avec

$$L_n\Big((Y_1, X_1), \dots, (Y_n, X_n); h, \eta \Big) = \sum_{i=1}^n \exp\Big(- Y_i \Big(g^{(b)}(X_i) + \eta h(X_i) \Big) \Big).$$

Donner la forme des poids $w_i^{(b+1)}$ qui vérifient

$$\sum_{i=1}^n \exp\Big(- Y_i \Big(g^{(b)}(X_i) + \eta h(X_i) \Big) \Big) = \sum_{i=1}^n w_i^{(b+1)} \exp\Big(- Y_i \eta h(X_i) \Big).$$

Puis montrer que

$$h^{(b+1)}, \eta^{(b+1)} = \mathrm{argmin}_{h \in \mathcal{H}, \eta \geq 0} \sum_{i=1}^n \frac{w_i^{(b+1)}}{\sum_{i=1}^n w_i^{(b+1)}} \exp\Big(- Y_i \eta h(X_i) \Big).$$

On note, pour la suite,

$$\pi_i^{(b+1)} = \frac{w_i^{(b+1)}}{\sum_{i=1}^n w_i^{(b+1)}}.$$

(c) Montrer que

$$\sum_{i=1}^n \pi_i^{(b+1)} \exp\Big(- Y_i \eta h(X_i) \Big) = \sum_{i=1}^n \pi_i^{(b+1)} \exp(-\eta) + \sum_{i : Y_i h(X_i) = -1}^n \pi_i^{(b+1)} \Big(\exp(\eta) - \exp(-\eta) \Big).$$

En déduire que, si $\pi_i^{(b+1)} = p_i^{(b+1)}$ pour tout i, alors, pour tout $\eta > 0$:

$$\mathrm{argmin}_{h\in\mathcal{H}} \sum_{i=1}^{n} \pi_i^{(b+1)} \exp\Big(- Y_i \eta h(X_i)\Big)$$

est égal à $h^{(b+1)}$ défini à la ligne 3 de l'algorithme.

(d) On fixe donc $h = h^{(b+1)}$, solution du problème précédent. Optimiser en η

$$\sum_{i=1}^{n} \pi_i^{(b+1)} \exp\Big(- Y_i \eta h^{(b+1)}(X_i)\Big)$$

Montrer ensuite que, si $\pi_i^{(b+1)} = p_i^{(b+1)}$, la solution se récrit comme $\eta^{(b+1)}$, défini à la ligne 5 de l'algorithme.

(e) Montrer que

$$\pi_i^{(b+1)} = \frac{w_i^{(b+1)}}{\sum_{i=1}^n w_i^{(b+1)}} = \pi_i^{(b)} \frac{\exp\Big(- \eta^{(b)} Y_i h^{(b)}(X_i)\Big)}{C^{(b)}}$$

où $C^{(b)}$ est une constante de normalisation. Montrer que $\sum_{i=1}^n \pi_i^{(b+1)} = 1$ et que c'est équivalent à

$$C^{(b)} = \sum_{i=1}^{n} \pi_i^{(b)} \exp\Big(- \eta^{(b)} Y_i h^{(b)}(X_i)\Big)$$

puis que $C^{(b)} = Z^{(b)}$, défini à la ligne 6 de l'algorithme. Conclure.

2. Nous considérons désormais une variante au problème de classification précédent dans laquelle les labels Y_i prennent leurs valeurs dans $\{-1, 0, 1\}$. Cela peut par exemple correspondre aux cas où le vrai label n'est pas observé et dans ce cas le label est mis à 0. Comme pour l'algorithme en classification classique, on se considère la perte 0/1 $\mathbb{1}_{y\neq u}$ et son approximation par la perte exponentielle. Proposer un algorithme de boosting dans ce cas en justifiant comme dans la question précédente toutes les étapes de l'algorithme.