Feuille de TP 3 – Régressions avancées

1 Régression logistique : exemple des crabes

1. Charger la bibliothèque `MASS` et le jeu de données `crabs` sous R.
2. Avec la fonction `glm`, effectuer la régression logistique pour prédire le sexe à partir des variables RW et BD du jeu de données. On effectuera le calcul sur un échantillon d'apprentissage et on comparera les prédictions avec les données.
3. Par descente de gradient à pas constant, retrouver le résultat précédent.

2 Régression LASSO

Il s'agit ici de minimiser

$$J_n(\vartheta) = R_n(\vartheta) + \lambda \mathrm{pen}(\vartheta) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \langle x_i, \vartheta\rangle)^2 + \lambda \sum_{j=1}^{d} |\vartheta_j|.$$

On suppose que $d = n = 1$ et $x_1 = 1$.

1. Montrer que
$$\vartheta^* = \underset{\vartheta\in\mathbb{R}}{\mathrm{argmin}} \frac{1}{2}(y_1 - \vartheta)^2 + \lambda|\vartheta| = \mathrm{sign}(y_1)(|y_1| - \lambda)_+ = S_\lambda(y_1).$$
2. Tracer $J_n(\vartheta)$ en fonction de ϑ pour $|y_1| < \lambda$ et $|y_1| > \lambda$.
3. Tester l'algorithme de descente de gradient pour les valeurs suivantes :
$$\lambda = 3 \text{ et } y_1 = 11, \qquad \text{puis} \qquad \lambda = 3 \text{ et } y_1 = 2.$$
Que remarque-t-on ?
4. L'algorithme de descente de gradient adapté à ce cas s'écrit :
$$\theta^k = S_{\lambda\rho}\left(\theta^{k-1} - \rho\nabla R_n(\theta^{k-1})\right)$$
où ρ est le pas de l'algorithme de descente. Implémenter cet algorithme et vérifier qu'il converge.

On va maintenant utiliser cette idée pour la régression LASSO. Notons $L = \|X^*X\|_2/n$ la constante de Lipschitz de ∇R_n ($\|X^*X\|_2$ est aussi la plus grande valeur propre de la matrice X^*X). Alors l'algorithme de descente de gradient adapté à notre cas s'écrit :

$$\theta^k = S_{\lambda/L}\left(\theta^{k-1} - \frac{1}{L}\nabla R_n(\theta^{k-1})\right)$$

où pour $z \in \mathbb{R}^d$, $S_\rho(z)$ est le vecteur

$$S_\rho(z) = \mathrm{sign}(z) \odot (|z| - \rho)_+ = (\mathrm{sign}(z_i)(|z_i| - \rho)_+)_{1\leqslant i\leqslant d}.$$

Dans les faits le schéma fonctionne mieux si on remplace L par $L' > L$.

5. Implémenter cet algorithme et obtenir la valeur de ϑ pour $\lambda = 1$ et $\lambda = 10$. On utilisera encore le jeu de données `trees` du TP précédent.
6. Comparer avec ce qu'on obtient la fonction `glmnet`.