

# TP Régression logistique

## SD211

Groupes en français : Roland Badeau, Olivier Fercoq, Eugène Ndiaye  
Group in English in C124 : Rahul Mourya  
12 mai 2017

Vous ferez un rapport accompagné des fonctions associées aux questions à envoyer au plus tard le 22 mai sur le site pédagogique. Nous attacherons de l'importance à l'interprétation des résultats obtenus. Vous pouvez rendre le rapport sous forme d'un notebook python ou d'un fichier pdf.

## 1 Régularisation de Tikhonov

On veut résoudre le problème suivant

$$\min_{w_0 \in \mathbb{R}, w \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(x_i^\top w + w_0))) + \frac{\rho}{2} \|w\|_2^2$$

### Question 1.1

Calculer le gradient de  $f_1 : (w_0, w) \mapsto \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(x_i^\top w + w_0))) + \frac{\rho}{2} \|w\|_2^2$  et sa matrice hessienne. La fonction est-elle convexe ?

### Question 1.2

Coder une fonction qui retourne la valeur de la fonction, son gradient et sa hessienne. Vous utiliserez la base de données cervical cancer fournie et  $\rho = 1/n$ . Il pourra être pratique de rajouter une colonne de uns à la matrice  $X$ .

Tester votre calcul de gradient avec la fonction `check_grad`.

### Question 1.3

Coder la méthode de Newton et la lancer avec comme condition initiale  $(w_0^0, w^0) = 0$  et comme test d'arrêt  $\|\nabla f_1(w)\|_2 < 10^{-10}$ . Afficher la norme du gradient en fonction des itérations en échelle logarithmique.

### Question 1.4

Lancer avec comme condition initiale  $(w_0^0, w^0) = 0.3e$  où  $e_i = 1$  pour tout  $i$ . Qu'observez-vous ?

### Question 1.5

La solution classique au problème observé à la question précédente est de rajouter une étape de recherche linéaire. Coder la recherche linéaire d'Armijo en justifiant vos choix des paramètres.

## 2 Régularisation pour la parcimonie

On s'intéresse toujours au problème de régression logistique mais on change la régularisation.

$$\min_{w_0 \in \mathbb{R}, w \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(x_i^\top w + w_0))) + \rho \|w\|_1$$

### Question 2.1

Pourquoi ne peut-on pas utiliser la méthode de Newton pour résoudre ce problème ?

### Question 2.2

Écrire la fonction objectif sous la forme  $F_2 = f_2 + g_2$  où  $f_2$  est dérivable et l'opérateur proximal de  $g_2$  est simple. Donner la formule de l'opérateur proximal de  $g_2$ . Calculer le gradient de  $f_2$ . La fonction objectif est-elle convexe ?

### Question 2.3

Coder le gradient proximal avec recherche linéaire. Ici, on prendra  $\rho = 0.1$ . Quel test d'arrêt proposeriez-vous ?

## 3 Comparaison

### Question 3.1

Comparer les propriétés des deux problèmes d'optimisation.

### Question 3.2

Comparer les solutions obtenues avec les deux types de régularisation.