

Challenge 1

Par: Jeremy Qin

Crédit à Gabriel Fortin-Leblanc (fortement inspiré)

Février 2023

Le Challenge 1 du club d'intelligence artificielle de l'Université de Montréal portera sur l'implémentation de l'algorithme d'Adam, une version très populaire et la plus utilisée de l'algorithme itératif : la descente de gradient stochastique. L'idée derrière l'algorithme est de partir d'une valeur initiale du paramètre à estimer et de converger vers la valeur optimale en allant à chaque itération dans la direction du gradient de la fonction de coût/risque empirique qui mesure essentiellement combien on est proche de la vraie valeur. Plus de détails seront présentés en dessous. Bon défi !

Tâche 1

Soit (X, Y) , un jeu de données qui prend ses valeurs dans les réels. Considérons le modèle de régression suivant

$$Y = \log(X\beta) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I). \quad (1)$$

X représente la variable explicative et Y représente la variable réponse. On va vouloir trouver le paramètre optimal au sens des moindres carrés. Cela vient à dire que la fonction de perte est la norme l2 de la différence entre la vraie valeur et la valeur estimée.

- a) Commencez par charger le jeu de données dans le fichier csv donné et affichez-le dans une figure.

Tâche 2

Comme mentionné un peu plus tôt, l'algorithme utilise le gradient du risque empirique. Ainsi, pour faciliter l'implémentation, calculez la dérivée du risque empirique selon le paramètre β . Exprimez-la aussi sous forme matricielle. (Cela vous aidera lors de l'implémentation en numpy).

Tâche 3

Selon le modèle de l'équation 1, trouver l'estimateur optimale qui minimise le risque empirique en utilisant l'algorithme d'Adam. Je vous recommande fortement de lire l'article au complet si vous êtes intéressé en ML. C'est un algorithme utilisé partout. Pour l'initialisation de l'algorithme, utilisez les valeurs suivantes

- $\beta_0 = 3$;
- $\alpha = 0.001$;
- $\rho_1 = 0.9$, et ;
- $\rho_2 = 0.999$.

Tâche 4

- a) Affichez sur une figure la fonction du risque empirique sur le domaine $[-1, 4]$, les valeurs itérées des paramètres lors de l'exécution de l'algorithme et une droite verticale correspondante à la valeur optimale obtenue par l'algorithme.
- b) Afficher la fonction estimée avec la paramètre optimal sur le jeu de données.