

The SDCA and PEGASOS Algorithms

Ismail HAMMOUNOU

May 5, 2017

On a essayé d'appliquer le SDCA sur trois bases de données, on a crée une et les deux autres bases, on les a trouvé sur le site de *Kaggle* et de *Yahoo/finance*. celle de Kaggle contient des données sur les voix des personnes et la variable d'intérêt est le sexe ("Homme", "femme"). Pour l'autre, de Yahoo, on a essayé de créer des *Features* sur la base des données qu'on a et la variable d'intérêt est le fait d'agir (Achat/Vente d'action) ou pas.

la fonction de coût est la fonction de hinge et il s'agit donc de l'algorithme SVM dans sa version basique sans introduction d'un biais. On a réussi à avoir de la convergence pour les deux algorithmes (SDCA, Pegasos). En effet, On a remarqué, comme le graphe le montre, que le SDCA est plus lent au début mais après un certain laps de temps, il converge plus rapidement que l'algorithme PEGASOS dans sa version **Batch**. Alors que ce dernier, au début, est plus rapide mais il est en générale plus lent que le SDCA.

L'algorithme SDCA qu'on a implémenté nous donnent une idée sur la convergence de les fonctions objectives dual et primal. En effet, On remarque que la solution primal converge plus rapidement que la solution dual et la différence n'est pas très grande comme on peut le voir sur le graphe.

L'application de l'algorithme SDCA sur les deux bases de données n'a pas été à la hauteur. On n'a pas réussi à avoir de la convergence pour la base de données GSPC. Par contre, on a réussi à avoir de la convergence pour la base de données SEXE sauf que la convergence était très lente et la fonction objective a beaucoup fluctué au début et à converger à la fin après un nombre d'itérations très élevé.

L'algorithme SDCA nécessite plus de ressource que celui de PEGASOS et il est plus lent que ce dernier. En effet, le SDCA nécessite 10.2 secondes par boucle dans ses meilleurs états, alors que PEGASOS nécessite 14 ms par boucle.

Il y a encore d'autres implémentations possibles de SDCA comme celle qui utilise le SGD pour initialiser les paramètres qui nous intéressent. Aussi, le SDCA qui au lieu de prendre un exemple d'une façon uniformément aléatoire, il prend une permutation aléatoire des exemples.