

Early stopping for stochastic gradient descent in deep learning

Projet

En apprentissage profond (deep learning), les couches successives des réseaux de neurones sont la plupart du temps optimisées par gradient stochastique (ou l'une de ses variantes telles que le momentum). Outre une grande variabilité et les optima locaux éventuels, une difficulté majeure liée à l'utilisation de cet algorithme itératif est de choisir après combien d'étapes interrompre le processus itératif. Ce choix est d'autant plus cruciale qu'il détermine la performance statistique ainsi que le coût de calculs (généralement déjà élevé pour les réseaux de neurones).

Les règles d'arrêt prématuré (Early Stopping Rules-ESR) sont des règles fondées sur les données qui fournissent à l'utilisateur le nombre d'étapes d'optimisation à effectuer. Si de tels règles d'arrêt existent déjà pour la descente de gradient classique, il n'en existe pas pour le gradient stochastique, ce qui conduit à utiliser des approches de type Hold-out reposant sur des données laissées de côté (au prix d'un coût de calculs additionnels).

Les objectifs du stage sont :

1. comprendre et utiliser l'algorithme de gradient stochastique (et certaines de ses variantes),
2. utiliser le gradient stochastique dans un réseau de neurones simple pour prédire une valeur réelle,
3. comprendre et utiliser une règle d'arrêt simple appliquée à la descente de gradient, puis adaptée pour le gradient stochastique,
4. étudier (empiriquement) et comprendre leur fonctionnement sur des simulations,
5. appliquer ces stratégies sur des vraies données disponibles librement sur le net.

Prérequis

Outre de bonnes compétences en apprentissage statistique/machine learning, le candidat utilisera R, Python ou Matlab pour réaliser des simulations afin de vérifier empiriquement ses résultats.

Encadrant : Alain Celisse, maître de conférences HDR.

Length: 4-6 mois.

Opportunity: Thèse académique possible

Laboratory: MODAL équipe-projet Inria, Lille.

Contact: Alain Celisse (alain.celisse@math.univ-lille1.fr).

Bibliographie

- Raskutti, Wainwrigth, Yu (2014), JMLR, Early Stopping and Non-parametric Regression: An Optimal Data-dependent Stopping Rule.
- Wei, Wainwright, Pilanci (2017). Early stopping for kernel boosting algorithms: A general analysis with localized complexities.