



Communauté Economique et Monétaire de l'Afrique Centrale

INSTITUT SOUS-REGIONAL DE STATISTIQUE ET D'ECONOMIE APPLIQUEE

Organisation Internationale

Projet de Machine Learning-Optimization Effet des algorithmes d'optimization sur l'erreur de généralisation

About The Authors

RÉDIGÉ PAR : HUNNAKEY Josué - NDOBO Pierre - TAGNE TCHINDA Rinel

Elèves Ingénieurs Statisticien Economiste (ISE)

SOUS LA SUPERVISION DE : Mme. MBIA NDI

Lien Github du Travail : [GitHub - isea-mloptimization-ise3-2023-2024](https://github.com/isea-mloptimization-ise3-2023-2024)

Résumé

Dans notre étude, nous avons examiné l'impact d'algorithmes d'optimisation sur l'erreur de généralisation dans le domaine de la santé, spécifiquement pour le cancer du sein. L'objectif était d'évaluer comment chaque méthode influençait les performances des modèles prédictifs. Nous avons testé plusieurs algorithmes, tels que Gradient, Proximal Gradient, NAG, Adam, RMSprop, Bayes et PSO, en mesurant des métriques comme l'erreur de généralisation, la perte finale, le temps de convergence et l'accuracy. Les résultats ont révélé des variations significatives : l'algorithme Gradient a affiché une erreur de généralisation de 0.000000, tandis que Newton et PSO ont obtenu des performances de 0.069767 et 0.023256. NAG et Adam se sont distingués par des temps de convergence rapides tout en maintenant une haute précision. Ces résultats soulignent l'importance du choix de l'algorithme d'optimisation pour développer des modèles robustes dans le secteur médical. Pour les données d'images, le dataset CIFAR-10 montre qu'Adam converge plus rapidement et généralise mieux que SGD, indiquant que le choix de l'optimiseur dépend des spécificités du problème..

Abstract

In our study, we examined the impact of various optimization algorithms on generalization error using a breast cancer dataset. Our goal was to assess how different methods influenced model performance. We tested algorithms like Gradient, NAG, Adam, and PSO, measuring metrics such as generalization error, final loss, convergence time, and accuracy. Results showed significant differences ; for example, the Gradient algorithm achieved a generalization error of 0.000000, while Newton and PSO had errors of 0.069767 and 0.023256, respectively. NAG and Adam demonstrated faster convergence with high accuracy. These findings highlight the importance of selecting the right optimization algorithm for robust medical models.

1 Introduction

Dans le domaine de l'apprentissage automatique et de l'intelligence artificielle, l'objectif principal d'un modèle est de bien performer non seulement sur les données d'entraînement, mais aussi de généraliser efficacement à des données invisibles. L'erreur de généralisation, qui évalue la performance sur de nouvelles données, est une métrique clé pour juger du succès d'un algorithme. Bien que l'architecture du modèle et la qualité des données soient cruciales, le choix de l'algorithme d'optimisation est tout aussi important pour déterminer la capacité de généralisation.

Les algorithmes d'optimisation, comme la Descente de Gradient, Adam ou l'Optimisation par Essaim Particulaire, jouent un rôle central dans l'apprentissage en guidant la convergence des paramètres du modèle. Cependant, leur impact sur la généralisation est complexe : certains convergent rapidement mais risquent le surapprentissage, tandis que d'autres explorent mieux l'espace des paramètres, améliorant la généralisation au détriment de la vitesse. Cette étude examine comment différents algorithmes influencent l'erreur de généralisation, permettant de choisir des stratégies d'optimisation adaptées pour renforcer la robustesse et la fiabilité des modèles dans des applications réelles.

2 Revue de la littérature

L'étude de l'impact des algorithmes d'optimisation sur l'erreur de généralisation suscite un intérêt croissant dans la communauté de l'apprentissage automatique. Cette revue de la littérature synthétise les travaux existants, mettant en lumière les avancées, défis et perspectives futures.

1.Fondements Théoriques : Les premiers travaux sur l'optimisation en apprentissage automatique se sont concentrés sur des méthodes classiques comme la Descente de Gradient (Robbins-Monro, 1951). Bien que simples, ces méthodes ont posé des bases théoriques essentielles pour comprendre la convergence des paramètres et son influence sur la performance des modèles. Cependant, elles présentent des problèmes de convergence lente et de sensibilité aux minima locaux, nuisant ainsi à la gé-

néralisation. Des études ultérieures, comme celles de Bottou (2010), ont démontré que la Descente de Gradient Stochastique (SGD) permet une convergence plus rapide et une meilleure exploration de l'espace des paramètres, améliorant ainsi la généralisation. L'importance du taux d'apprentissage et de la taille des lots (batch size) dans la réduction de l'erreur de généralisation a également été soulignée.

2.Algorithmes Adaptatifs : L'émergence des algorithmes adaptatifs, tels qu'Adam (Kingma-Ba, 2015), a marqué un tournant dans l'optimisation des modèles d'apprentissage profond. Adam combine les avantages de la méthode du momentum et de la mise à l'échelle adaptative des gradients, permettant une convergence plus stable et rapide. Toutefois, des études récentes, comme celles de Wilson et al. (2017), montrent qu'Adam peut parfois nuire à la généralisation en convergeant vers des minima plus "pointus" dans le paysage de la fonction de coût. D'autres algorithmes adaptatifs, tels que RMSProp et Adagrad, ont également été étudiés pour leur capacité à ajuster dynamiquement le taux d'apprentissage, bien que leur impact sur la généralisation dépende fortement de la structure du problème et des hyperparamètres.

3.Exploration vs Exploitation : Les algorithmes d'optimisation basés sur l'exploration, comme les algorithmes génétiques et l'Optimisation par Essaim Particulaire (PSO), ont été explorés pour leur capacité à éviter les minima locaux. Ces méthodes, bien que coûteuses en termes de calcul, peuvent améliorer la généralisation en explorant un espace de paramètres plus large. Par exemple, les travaux de Zhang et al. (2020) montrent que PSO peut être particulièrement efficace pour les problèmes non convexes, là où les méthodes basées sur le gradient peuvent rester coincées dans des minima locaux. Cependant, ces méthodes souffrent souvent d'un manque de scalabilité pour les modèles de grande taille, limitant leur applicabilité dans les réseaux de neurones profonds modernes.

4.Régularisation et Généralisation : La relation entre les algorithmes d'optimisation et la régularisation a également été largement étudiée. Par exemple, il a été démontré que SGD avec momentum a un effet de régularisation implicite (Zhang et al., 2019), améliorant la généralisation en empêchant le surapprentissage. Des techniques comme

weight decay et dropout interagissent fortement avec les algorithmes d'optimisation, influençant l'erreur de généralisation. Des travaux récents, comme ceux de Li et al. (2021), ont exploré comment les algorithmes d'optimisation modernes, tels que LAMB (Layer-wise Adaptive Moments), peuvent être combinés avec des techniques de régularisation pour améliorer la généralisation dans les modèles de grande taille, comme les Transformers.

3 Présentation des bases de données

1. Load Breast Cancer Dataset Le jeu de données Breast Cancer est un ensemble de données couramment utilisé en apprentissage automatique pour la classification binaire. Il contient des caractéristiques dérivées d'images numérisées de masses mammaires, avec pour objectif de prédire si une tumeur est maligne (cancéreuse) ou bénigne (non cancéreuse). Ce domaine relève de la santé et de l'oncologie, et il est largement utilisé pour développer des modèles de diagnostic assisté par ordinateur. Son intérêt pour les organismes de santé, les hôpitaux et les entreprises de technologie médicale réside dans sa capacité à améliorer la précision et la rapidité des diagnostics, réduisant ainsi les coûts et sauvant des vies. Pour les chercheurs, il sert de référence pour tester de nouveaux algorithmes de classification et d'optimisation.

2. Adult Dataset (UCI)

Le jeu de données Adult (ou "Census Income") est un ensemble de données utilisé pour prédire si le revenu d'un individu dépasse 50 000 dollars par an, en fonction de caractéristiques démographiques telles que l'âge, l'éducation, la profession et le statut marital. Ce domaine relève de l'économie, de la sociologie et de la gestion des ressources humaines. Il est souvent utilisé pour étudier les inégalités de revenus, les tendances du marché du travail et les politiques publiques. Pour les entreprises, il peut aider à segmenter les clients ou à prédire le potentiel de revenus. Pour les gouvernements, il est utile pour évaluer l'impact des politiques sociales et économiques. Ce dataset est également un benchmark populaire pour tester des algorithmes de classification et de prédiction.

Ces deux jeux de données illustrent l'importance de l'apprentissage automatique dans des domaines

critiques comme la santé et l'économie, offrant des opportunités pour améliorer les décisions et les politiques dans divers secteurs.

4 Principaux resultats

1. Base Breast Cancer (Petit Nombre d'Observations)

	Méthode	Erreur de Généralisation	Perte Finale	Temps de Convergence	Accuracy
0	Gradient	0.000000	0.080085	0.716364	1.000000
1	Proximal Gradient	0.046512	0.287908	1.439825	0.994872
2	NAG	0.011628	0.038240	1.067085	1.000000
3	Adam	0.000000	0.079250	1.246936	1.000000
4	RMSprop	0.011628	0.057804	2.142840	0.988372
5	Adagrad	0.023256	0.081544	1.080520	0.976744
6	Newton	0.069767	1.329707	0.052183	0.989103
7	recuit	0.023256	0.153039	0.063754	0.976744
8	bayes	0.023256	0.131175	64.819427	0.976744
9	pso	0.023256	0.180877	2.178173	0.976744

Source : nos travaux sous python à partir des données des données sur le cancer

Gradient Descent et NAG montrent une erreur de généralisation de 0.000000 et 0.011628 respectivement, avec une précision de 1.000000, ce qui indique une excellente performance sur cette base de données.

Adam et RMSprop ont également une erreur de généralisation faible (0.000000 et 0.011628) mais avec des temps de convergence plus élevés.

Newton et Recuit Simulé montrent des erreurs de généralisation plus élevées (0.069767 et 0.023256) mais avec des temps de convergence très rapides.

Bayes et PSO ont des temps de convergence très longs, ce qui les rend moins pratiques pour cette base de données.

On peut donc retenir que : Les algorithmes d'optimisation tels que Gradient et Adam ont montré des performances exceptionnelles en termes d'erreur de généralisation (pas d'erreur (0.000000)). Cela indique que ces méthodes sont particulièrement adaptées pour des échantillons réduits, permettant une meilleure généralisation.

Les résultats révèlent également que des algorithmes comme RMSprop et NAG, bien que performants, présentent une erreur de généralisation plus élevée. Cela peut suggérer qu'ils sont moins adaptés à des ensembles de données plus petits où la variance est plus significative.

2. Base Adult Dataset (Grand Nombre d'Observations)

	Méthode	Erreur de Généralisation	Perte Finale	Temps de Convergence	Accuracy
0	Gradient	0.150051	0.339091	45.920226	0.896796
1	Proximal Gradient	0.205322	0.493245	63.655628	0.847219
2	NAG	0.144115	0.318635	61.470855	0.902662
3	Adam	0.147390	0.335586	60.344333	0.902095
4	RMSprop	0.144524	0.318752	64.307851	0.855476
5	Adagrad	0.145343	0.318941	63.562387	0.854657
6	Newton	0.145548	0.320779	54.507297	0.901266
7	recuit	0.145343	1.121573	1.653550	0.854657
8	bayes	0.145343	2.840041	372.772579	0.654043
9	pso	0.145343	0.615441	244.201341	0.854657

Source : nos travaux sous python à partir des données des données sur le cancer

Gradient Descent et NAG montrent des erreurs de généralisation relativement faibles (0.150051 et 0.144115) avec des précisions élevées (0.896796 et 0.902662).

Adam et RMSprop ont des performances similaires, avec des erreurs de généralisation légèrement plus élevées mais des temps de convergence comparables.

Newton et Recuit Simulé montrent des erreurs de généralisation similaires à celles des autres méthodes, mais avec des temps de convergence plus rapides.

Bayes et PSO ont des temps de convergence extrêmement longs, ce qui les rend peu pratiques pour cette base de données.

On peut donc retenir que :

- Pour cette base, l'erreur de généralisation est significativement plus élevée, avec le Gradient affichant 0.150501. Cela peut indiquer que les modèles ont plus de difficulté à généraliser sur des ensembles de données plus vastes, ce qui est souvent observé lorsque la complexité des données augmente.

- Les algorithmes comme NAG, Adam, et Adagrad montrent des performances acceptables, mais l'erreur de généralisation reste notablement supérieure à celle observée dans la base Breast Cancer, ce qui n'est que chose normal puisque les bases ne sont pas les même en termes de taille.

REMARQUE GENERALE :

La NAG, la Gradient Descent et Adam sont les méthodes les plus performantes en termes d'erreur de généralisation et de précision, bien que NAG soit légèrement plus rapide. Newton et Recuit Simulé sont rapides mais peuvent avoir des erreurs de généralisation plus élevées pour des petits échantillons. La proximal gradient, Bayes et PSO sont peu pratiques en raison de leurs temps de convergence très longs. Erreur de Généralisation : Les méthodes comme NAG, Gradient Descent et Adam montrent une erreur de généralisation très faible, ce qui indique une bonne capacité à généraliser. Temps de Convergence : Les méthodes comme Newton et Recuit Simulé sont très rapides mais peuvent sacrifier un peu de précision. Accuracy : L'accuracy peut être un bon indicateur, mais elle ne doit pas être considérée isolément. Il est essentiel d'examiner l'équilibre entre l'accuracy et l'erreur de généralisation pour évaluer la robustesse d'un modèle.

5 Cas particulier des données images

Le dataset CIFAR-10 comprend 60 000 images de 32x32 pixels réparties en 10 classes, utilisé pour évaluer les modèles de classification d'images. Les algorithmes d'optimisation courants incluent SGD, simple mais efficace, Adam, qui ajuste le taux d'apprentissage pour chaque paramètre, et RMSprop, qui utilise une approche similaire.

Les images sont normalisées pour stabiliser l'apprentissage, et chaque algorithme est testé sur le même modèle de réseau de neurones pour une comparaison équitable. Les résultats montrent qu'Adam converge plus rapidement vers une perte plus faible que SGD, avec une meilleure erreur de généralisation. RMSprop, bien qu'efficace, peut présenter des oscillations.

En conclusion, le choix de l'optimiseur impacte significativement la généralisation, Adam étant souvent préféré pour sa rapidité et sa stabilité, tandis que le choix doit dépendre de la nature du problème et des ressources disponibles.

RECOMMANDATIONS ET AMELIORATIONS FUTURES :

Optimisation des Hyperparamètres Une exploration plus approfondie des hyperparamètres comme le taux d'apprentissage pour les méthodes comme Adam et RMSprop pourrait améliorer leurs performances.

Combinaison de Méthodes : Utiliser une combinaison de méthodes rapides comme Newton pour une convergence initiale, suivie de méthodes plus précises comme NAG pour affiner les résultats.

Réduction du Temps de Convergence : Des recherches pourraient être menées pour réduire le temps de convergence des méthodes comme Bayes et PSO, peut-être en utilisant des techniques de parallélisation ou d'approximation.

Validation sur Plus de Bases de Données : Étendre l'étude à d'autres bases de données pour confirmer la généralisabilité des résultats.

Expérimentation avec d'autres algorithmes : Tester des algorithmes d'optimisation supplémentaires, comme ceux basés sur des approches évolutionnaires.

6 references :

cours de ML-optimization : Mbiandi M.T
<https://github.com/MBIANDI/MLoptimization>