

PRESENTATION DE RMSPPROP

GROUPE 11

Introduction

RMSprop Root Mean Squared Propagation— est un algorithme d'optimisation non publié conçu pour les réseaux de neurones, proposé pour la première fois par **Geoff Hinton** dans la conférence 6 du cours en ligne « **Neural Networks for Machine Learning** ».

RMSprop se situe dans le domaine des méthodes de taux d'apprentissage adaptatif, qui ont gagné en popularité ces dernières années. Il est célèbre pour ne pas avoir été publié, mais il est pourtant très connu.

Introduction

RMSProp est conçu pour accélérer le processus d'optimisation, par exemple diminuer le nombre d'évaluations de fonctions nécessaires pour atteindre les optima, ou pour améliorer la capacité de l'algorithme d'optimisation, par exemple pour obtenir un meilleur résultat final.

Il est lié à une autre extension de la descente de gradient appelée Adaptive Gradient, ou AdaGrad.

Problème de la descente de gradient et AdaGrad

Une limitation de la descente de gradient est qu'elle utilise la même taille de pas (taux d'apprentissage) pour chaque observation d'entrée.

AdaGrad en abrégé, est une extension de l'algorithme d'optimisation de descente de gradient qui permet d'adapter automatiquement la taille du pas dans chaque dimension utilisée par l'algorithme d'optimisation en fonction des gradients observés pour la variable (dérivées partielles) au cours de la recherche.

Une limitation d'AdaGrad est qu'il peut entraîner une très petite taille de pas pour chaque paramètre à la fin de la recherche, ce qui peut trop ralentir la progression de la recherche et peut signifier ne pas localiser les optima.

Presentation de RMSProp

La taille du pas doit être faible lorsque nous approchons des minima globaux et elle doit être élevée lorsque nous sommes loin des minima globaux.

Donc, fondamentalement, ce dont nous avons besoin, c'est que si nous partons d'un point aléatoire, nous devons descendre rapidement pour arriver à un autre point (puisque nous sommes loin des minima globaux, donc la taille du pas doit être grande dans ce cas) et lorsque nous sommes proches des minima globaux, la taille du pas doit être faible .

Maintenant, comment la taille du pas doit être calculée pour que notre objectif de faire varier la taille du pas puisse être atteint.

Presentation de RMSProp

RMSProp peut être considéré comme une extension d'AdaGrad en ce sens qu'il utilise une moyenne décroissante ou une moyenne mobile des dérivées partielles au lieu de la somme dans le calcul du taux d'apprentissage pour chaque paramètre.

RMSProp utilise une moyenne décroissante de gradients partiels dans l'adaptation de la taille du pas pour chaque paramètre.

Presentation de RMSProp

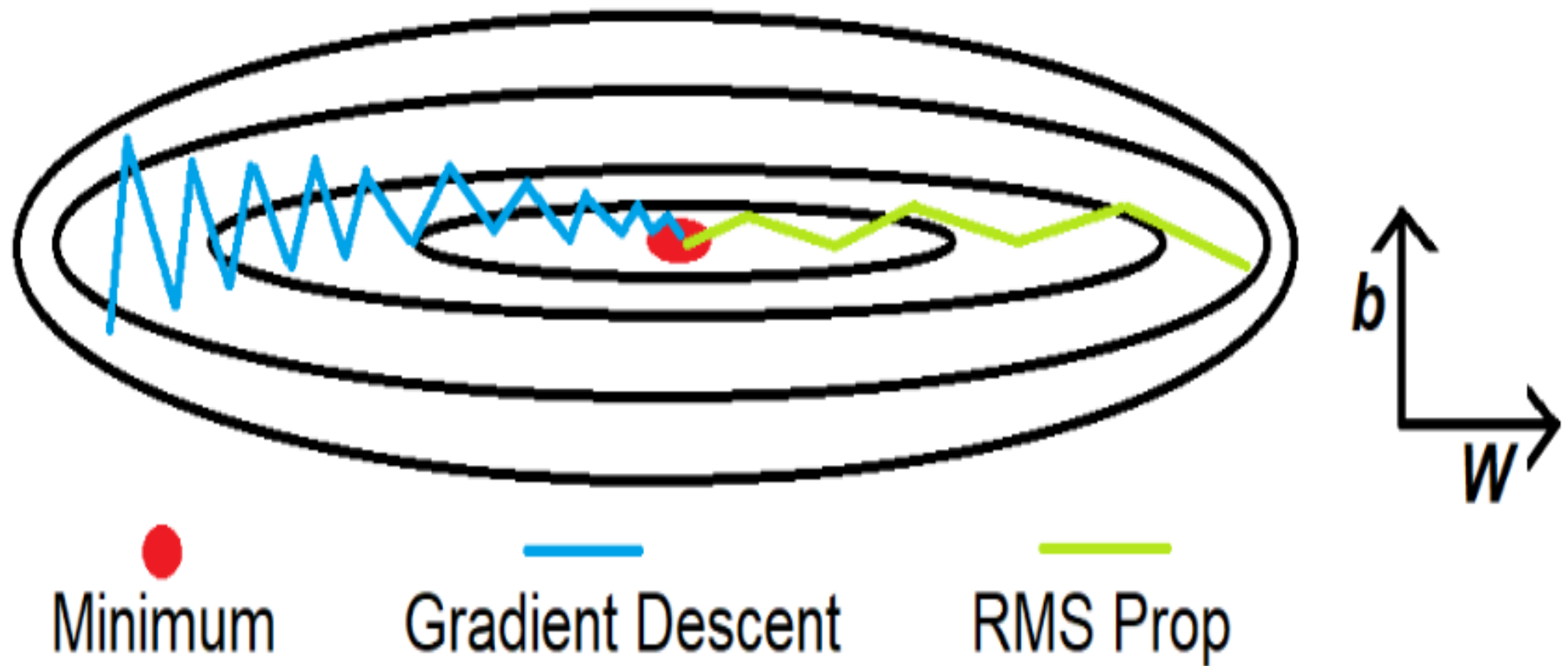
L'utilisation d'une moyenne mobile décroissante permet à l'algorithme d'oublier les gradients précoces et de se concentrer sur les gradients partiels les plus récemment observés au cours de la progression de la recherche, surmontant la limitation d'AdaGrad.

RMSProp utilise une moyenne décroissante de façon exponentielle pour éliminer l'historique du passé extrême afin qu'il puisse converger rapidement après avoir trouvé une surface convexe.

Intuition

La racine carrée moyenne de la propagation RMS Prop est similaire à Momentum, c'est une technique pour amortir le mouvement dans l'axe y et accélérer la descente du gradient. Pour une meilleure compréhension, notons l'axe Y comme le biais b et l'axe X comme le poids w . Il s'appelle Root Mean Square parce que nous mettons au carré les dérivées des paramètres w et b .

Intuition fig 1



Intuition

L'intuition est que lorsque l'on divise un grand nombre par un autre nombre, le résultat devient petit. Dans notre cas, le premier grand nombre est db et le deuxième grand nombre que nous utilisons est la moyenne pondérée de db^2 . Nous introduisons deux nouvelles variables Sdb et SdW , pour suivre la moyenne pondérée de db^2 et dW^2 . La division de db et Sdb donne une valeur plus petite qui amortit le mouvement sur l'axe y . Le ϵ est introduit pour éviter l'erreur de division par 0.

Intuition fig 2 : Optimiseur RMSprop

$$v_{dw} = \beta \cdot v_{dw} + (1 - \beta) \cdot dw^2$$

$$v_{db} = \beta \cdot v_{db} + (1 - \beta) \cdot db^2$$

$$W = W - \alpha \cdot \frac{dw}{\sqrt{v_{dw}} + \epsilon}$$

$$b = b - \alpha \cdot \frac{db}{\sqrt{v_{db}} + \epsilon}$$

Intuition

L'idée est de ralentir l'apprentissage sur la direction de l'axe y et d'accélérer l'apprentissage sur la direction de l'axe x . A chaque itération les dérivées dw et db sont calculées.

les moyennes mobiles exponentiellement appelée SdW et Sdb sont également calculer. Maintenant, le paramètre moyen pondéré exponentiellement SdW est relativement petit et nous le divisons donc par un petit nombre pour obtenir le poids.

Le paramètre α est le taux d'apprentissage et le ϵ est introduit pour éviter l'erreur de division par 0.

Intuition

Au contraire, Sdb est relativement volumineux et cela permet de ralentir les mises à jour de dimension verticale b . En fait, en regardant la figure fig 1, la dérivée dans la dimension horizontale est toujours petite et la dérivée dans la dimension verticale est grande.

L'effet net est d'accélérer l'apprentissage vertical et en même temps de ralentir l'apprentissage vertical. Le résultat est la ligne verte de la figure 2.

formule de mise à jour des poids

fig 3

$$E[g^2]_t = \beta E[g^2]_{t-1} + (1 - \beta) \left(\frac{\delta C}{\delta w} \right)^2$$
$$w_t = w_{t-1} - \frac{\eta}{\sqrt{E[g^2]_t}} \frac{\delta C}{\delta w}$$

$E[g]$ – moyenne mobile des gradients au carré ou la moyenne quadratique des gradients.

dC/dw – gradient de la fonction de coût par rapport au poids.

η – taux d'apprentissage.

Beta - paramètre de moyenne mobile (bonne valeur par défaut 0,9)

conclusion

Dans la descente de gradient, le problème était que la taille du pas était la même à chaque fois, mais nous modifions le taux d'apprentissage ou la taille du pas en utilisant adagrad, mais c'est un problème dans ce cas, et si le dénominateur est très grand, maintenant, pour contrôler cela, nous donnons pondération inférieure au poids total précédent et pondération supérieure au poids moyen récent ou au poids moyen mobile. Et nous pouvons contrôler cela dans RMSprop.

