Méthode cumulative pour la descente de gradient.

auteur:	Denis Rouiller
github:	https://github.com/drrr-deudeu

Abstract:

Dans ce document, nous détaillons un développement mathématiques et une méthode permettant de corriger/d'actualiser les paramètres d'un modèle linéaire post entraînement/ajustement en déterminant un terme de correction $\delta\theta$ issue d'un nouvel ensemble de données $\mathbb{E}_{n+1,n+m}$. La formule empirique du gain de temps obtenu avec cette méthode de correction est :

$$2n + 4n^2$$

Introduction:

[Bref récap du machine learning, un truc bateau pour introduire]
[l'émergence de l'IA actuelle avec l'augmentation des données et de la puissance de calcul]
[parler de la taille des datasets nécessaires et tjrs de plus en plus groa]
[du temps que l'entrainement peut prendre]
[la prise en compte de nouvelles données permettant de corriger le modèle pour l'obtention de meilleures prédictions] [d'autres idées ?]
Nous détaillons dans ce document un développement mathématiques permettant de corriger les paramètres θ du modèle à partir d'un nouvel ensemble de données $\mathbb{E}_{(n+1,n+m)}$ sans devoir réentraîner le modèle avec l'ensemble de données initiales $\mathbb{E}_{(1,n+m)}$, puis nous exposons les résultats de cette nouvelle méthode en comparaison d'un réentraînement classique à partir de l'ensemble de données $\mathbb{E}_{(1,n+m)} = \mathbb{E}_{(1,n)} + \mathbb{E}_{(n+1,n+m)}$.

Méthode:

Soit $n \in \mathbb{N}$, tel que $1 < n$
Nous notons $\mathbb{E}_{(1,n)}$ l'ensemble des données issus de la réunion de $\mathbb{X}_{(1,n)} = \{x_1, \ldots, x_i, \ldots, x_n\}$ et $\mathbb{Y} = \{y_1, \ldots, y_i, \ldots, y_n\}$.
Nous avons alors:

$$\forall i \in \mathbb{N} \backslash (1 \leq i \leq n), (x_i, y_i) \in \mathbb{X} \times \mathbb{Y} = \mathbb{E}$$

Définitions:

$\forall x \in \mathbb{X} = \{x_1, ..., x_i, ..., x_n\}$, on définit $\overline{x}(n)$ la moyenne des n éléments x_i telle que:

$$\overline{x}(n) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (a)$$

Notons au passage qu'on a donc:

$$n\overline{x}(n) = \sum_{i=1}^{n} x_i \quad (b)$$

On définit maintenant $\sigma_{(n)}(x)$ l'écart-type des n élément x_i et la variance $V_{(n)}(x)$ tels que:

$$\sigma_{(n)}(x) = \sqrt{V_{(n)}(x)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_{(n)})^2}$$

On peut écrire:

$$V_{(n)}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2\overline{x}_{(n)}x_i + \overline{x}_{(n)}^2)$$

$$V_{(n)}(x) = \frac{1}{n}(\sum_{i=1}^{n} x_i^2 - 2\overline{x}_{(n)}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \overline{x}_{(n)}^2)$$

$$V_{(n)}(x) = \frac{1}{n}(\sum_{i=1}^{n} x_i^2 - 2n\overline{x}_{(n)}\overline{x}_{(n)} + n\overline{x}_{(n)}^2)$$

Donc:

$$V_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}_{(n)}^2 \quad (c)$$

Développement

Soit maintenant $m \in \mathbb{N}$, tel que $1 < m$.
Soit l'ensemble $\mathbb{X}_{(n+m)} = \{x_1, ..., x_i, ..., x_n, x_{n+1}, ..., x_{n+m}\}$. Alors la moyenne des $n+m$ éléments x_i de $\mathbb{X}_{(n+m)}$, notée $\overline{x}(n+m)$:

$$\overline{x}_{(n+m)} = \frac{1}{n+m}\sum_{i=1}^{n+m} x_i$$

$$\overline{x}_{(n+m)} = \frac{1}{n+m}(\sum_{i=1}^{n} x_i + \sum_{i=n+1}^{m} x_i)$$

en utilisant l'équation (b) on en déduit:

$$\overline{x}_{(n+m)} = \frac{1}{n+m}(n\overline{x}_{(n)} + m\overline{x}_{(m)})$$

Calculons maintenant la variance de $\mathbb{X}_{(n+m)}$

$$V_{(n+m)}(x) = \frac{1}{n+m}\sum_{i=1}^{n+m}(x_i - \overline{x})^2$$

$$V_{(n+m)}(x) = \frac{1}{n+m}\sum_{i=1}^{n+m}x_i^2 - \overline{x}^2_{(n+m)}$$

$$V_{(n+m)}(x) = \frac{1}{n+m}(\sum_{i=1}^{n}x_i^2 + \sum_{i=n+1}^{m}x_i^2) - \overline{x}^2_{(n+m)}$$

On définit la normalisation de l'ensemble $\mathbb{X}_{(n)}$,la transformation faisant passer x_i à X_i comme suit, dans l'ensemble $\mathbb{X}_{(n)}$:

$$X_{(n)i} = \frac{x_i - \overline{x}}{\sigma_{(n)}(x)} \quad (d)$$

Notons au passage les résultats suivants:

$$\overline{X}_{(n)} = \frac{1}{n\sigma_{(n)}(x)}\sum_{i=1}^{n}(x_i - \overline{x}) = 0 \quad (e)$$

$$\sum_{i=1}^{n}X_i^2 = \sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\sigma(x)^2} = \sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{n}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sum_{i=1}^{n}(x_i-\overline{x})^2 = n \quad (f)$$

Application dans le cas d'une droite de régression linéaire:

Nous nous intéressons ici à l'application de la méthode de la descente de gradient à une droite de régression linéaire telle que:

$$\hat{Y} = F(X) = \theta_0 + \theta_1 X \quad (g)$$

La méthode consiste à s'approcher itérativement des valeurs optimales des θ i.e. celles pour lesquelles le coût est minimum. Les équations proposées pour résoudre ce problème ne sont valables qu'avec des données normalisées: Chaque itération donne des nouvelles valeurs des θ:

$$\theta_0 \leftarrow \theta_0 - d\theta_0$$

$$\theta_1 \leftarrow \theta_1 - d\theta_1$$

Avec:

$$d\theta_0 = \frac{\alpha}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)\ et\ \ d\theta_1 = \frac{\alpha}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)X_i$$

avec α définit comme étant le "learning~ rate", $\hat{Y}_i$ la valeur prédite en X_i et Y_i la valeur normalisée de y_i.
On peut donc écrire:

$$\frac{d\theta_0}{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i - \overline{Y}(n) = \frac{1}{n}\sum_{i=1}^{n}(\theta_0 + \theta_1 X_i) - \overline{Y}_{(n)}$$

En appliquant (e):

$$\frac{d\theta_0}{\alpha} = \theta_0$$

On a donc à chaque itération:

$$\theta_0 \leftarrow \theta_0 - \alpha\theta_0$$

De même:

$$\frac{d\theta_1}{\alpha} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)X_i$$

$$\frac{d\theta_1}{\alpha} = \frac{1}{n}\sum_{i=1}^{n}(\theta_0 + \theta_1 X_i - Y_i)X_i$$

$$\frac{d\theta_1}{\alpha} = \frac{\theta_0}{n}\sum_{i=1}^{n}X_i + \frac{\theta_1}{n}\sum_{i=1}^{n}X_i^2 - \frac{1}{n}\sum_{i=1}^{n}X_iY_i$$

Mais d'après (e), on a $\overline{X}_{(n)} = 0$, d'où :

$$\frac{d\theta_1}{\alpha} = \frac{\theta_1}{n}\sum_{i=1}^{n}X_i^2 - \frac{1}{n}\sum_{i=1}^{n}X_iY_i$$

et en appliquant (f), on trouve:

$$\frac{d\theta_1}{\alpha} = \theta_1 - \frac{1}{n}\sum_{i=1}^{n}X_iY_i$$

Donc, $\theta_1 \leftarrow \theta_1 - \alpha(\theta_1 - \frac{1}{n}\sum_{i=1}^{n}X_iY_i)$

Notons, que nous avons:

$$\sum_{i=1}^{n}X_iY_i = \sum_{i=1}^{n}\frac{(x_i - \overline{x}_{(n)})(y_i - \overline{y}_{(n)})}{\sigma_{(n)}(x)\sigma_{(n)}(y)} = \frac{1}{\sigma_{(n)}(x)\sigma_{(n)}(y)}\sum_{i=1}^{n}(x_iy_i - x_i\overline{y}_{(n)} - \overline{x}_{(n)}y_i + \overline{x}_{(n)}\overline{y}_{(n)})$$

$$=> \sum_{i=1}^{n}X_iY_i = \frac{1}{\sigma_{(n)}(x)\sigma_{(n)}(y)}(\sum_{i=1}^{n}x_iy_i - n\overline{x}_{(n)}\overline{y}_{(n)}) \quad (h)$$

Pour l'ensemble (n+m) on aura :

$$\theta_0 \leftarrow \theta_0 - \alpha\theta_0$$

$$\theta_1 \leftarrow \theta_1 - \alpha(\theta_1 - \frac{1}{n+m}\sum_{i=1}^{n+m}X_iY_i)$$

Qui peut encore s'écrire:

$$\theta_1 \leftarrow \theta_1 - \alpha[\theta_1 - \frac{1}{(n+m)\sigma_{(n+m)}(x)\sigma_{(n+m)}(y)}(\sum_{i=1}^{n+m} x_i y_i - (n+m)\overline{x}_{(n+m)}\overline{y}_{(n+m)})]$$

Avec $\sigma_{(n+m)}(x) = \sqrt{V_{(n+m)}(x)}$ et $\sigma_{(n+m)}(y) = \sqrt{V_{(n+m)}(y)}$

Enfin la dénormalisation des θ:
D'après (d) et (f), on a:

$$\frac{\hat{y}i - \overline{y}(n)}{\sigma_{(n)}(y)} = \theta_0 - \theta_1 \frac{x_i - \overline{x}(n)}{\sigma(n)(x)}$$

$$\hat{y}i = \sigma(n)(y)(\theta_0 - \frac{theta_1}{\sigma_{(n)}(x)}\overline{x}(n)) + \overline{y}(n) + \theta_1 \frac{\sigma_{(n)}(y)}{\sigma_{(n)}(x)} x_i$$

D'où, par identification:

$$THETA(0) = \sigma_{(n)}(y)(\theta_0 - \frac{\theta_1}{\sigma_{(n)}(x)}\overline{x}(n)) + \overline{y}(n)$$

$$THETA(1) = \theta_1 \frac{\sigma_{(n)}(y)}{\sigma_{(n)}(x)}$$

Observation:

Le calcul des θ pour un ensemble $n+m$, ne nécessitent que la connaissance des grandeurs suivantes:

- Le cardinal n, de notre ancien ensemble
- Les θ normalisées sur l'ensemble $X_{(n)}$
- Les grandeurs: $\overline{x}(n)$, $\overline{y}(n)$, $\sum_{i=1}^{n} x_i^2$, $\sum_{i=1}^{n} y_i^2$, $\sum_{i=1}^{n} x_i y_i$
- l'ensemble des X_i et des Y_i appartenant aux ensembles

Cas particulier:

Avec l'ajoût d'une seule valeur x_{n+1}, on aura:

$$\overline{x}(n+1) = \frac{1}{n+1}(n\overline{x}(n) + x_{n+1})$$

$$V_{(n+1)}(x) = \frac{1}{n+m}(\sum_{i=1}^{n} x_i^2 + x_{n+1}^2) - \overline{x}_{(n+m)}^2$$

$$\sigma_{(n+1)}(x) = \sqrt{V_{(n+1)}}$$

$$X_{n+1} = \frac{x_{n+1} - \overline{x}_{(n+1)}}{\sigma_{(n+1)}(x)}$$

Méthodologie:

Analytique

Résultats:

Conclusion: