Exercice complet : Régression logistique avec deux variables

Cet exercice vous guide à travers les étapes pour résoudre une régression logistique avec deux variables en utilisant la fonction de coût **log-loss** et l'algorithme de **descente de gradient**. À chaque étape, nous rappelons les formules nécessaires.

Étape 1 : Formulation de la régression logistique

La **régression logistique** prédit la probabilité qu'une observation appartienne à la classe 1 (malade) en utilisant la **fonction sigmoïde** :

$$\hat{y} = \frac{1}{1 + e^{-(w_1 \cdot x_1 + w_2 \cdot x_2 + b)}}$$

Où :

- w_1 et w_2 sont les **poids** pour les deux caractéristiques x_1 et x_2,
- b est le **biais** (ou intercept),
- $\hat{y}$ est la probabilité prédite.

Étape 2 : Fonction de coût Log-Loss

La **fonction de coût log-loss** mesure la différence entre les prédictions $\hat{y}_i$ et les véritables labels y_i. Elle est définie par la formule suivante :

$$\textbf{Log-Loss} = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

Où :

- y_i est la véritable étiquette (0 ou 1),
- $\hat{y}_i$ est la probabilité prédite pour y_i,
- n est le nombre d'observations.

Étape 3 : Exemple de données

Considérons trois observations avec deux caractéristiques x_1, x_2 et leurs labels y.

x_1	x_2	y
1.2	0.7	1
2.4	1.5	0
1.8	1.1	1

Nous allons utiliser ces données pour calculer les probabilités prédites $\hat{y}_i$, la fonction de coût log-loss, et ajuster les paramètres w_1, w_2, et b.

Étape 4 : Calcul des probabilités avec la fonction sigmoïde

Nous utilisons la fonction sigmoïde pour calculer la probabilité pour chaque donnée. Supposons que :

- $w_1 = 0.5$, $w_2 = 0.3$,
- $b = 0$ (initialement).

La probabilité prédite $\hat{y}_i$ pour chaque observation est calculée comme suit :

$$\hat{y}_i = \frac{1}{1 + e^{-(w_1 \cdot x_1 + w_2 \cdot x_2 + b)}}$$

Calculer les valeurs de $\hat{y}_1$, $\hat{y}_2$, et $\hat{y}_3$ à partir des données d'exemple.

Étape 5 : Calcul de la fonction de coût Log-Loss

Nous utilisons les vraies valeurs y et les probabilités prédites $\hat{y}$ pour calculer la fonction de coût log-loss :

$$\textbf{Log-Loss} = -\frac{1}{3} \sum_{i=1}^{3} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

Calculez chaque terme pour $i = 1$, $i = 2$, et $i = 3$, puis déterminez la valeur finale de la fonction de coût log-loss.

Étape 6 : Dérivation de la fonction de coût Log-Loss

Pour ajuster les paramètres, nous devons calculer les **dérivées** de la fonction de coût par rapport à w_1, w_2, et b.

Dérivée par rapport à w_1 :

$$\frac{\partial J}{\partial w_1} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i1}$$

Dérivée par rapport à w_2 :

$$\frac{\partial J}{\partial w_2} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i2}$$

Dérivée par rapport à b (biais) :

$$\frac{\partial J}{\partial b} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

Utilisez les données d'exemple pour calculer les dérivées par rapport à w_1, w_2, et b.

Étape 7 : Descente de gradient

L'algorithme de **descente de gradient** ajuste les paramètres w_1, w_2, et b en suivant la direction opposée du gradient.

Mise à jour des poids w_1 et w_2 :

$$w_1 := w_1 - \alpha \frac{\partial J}{\partial w_1}$$

$$w_2 := w_2 - \alpha \frac{\partial J}{\partial w_2}$$

Mise à jour du biais b :

$$b := b - \alpha \frac{\partial J}{\partial b}$$

Où α est le **taux d'apprentissage**.

Effectuez une mise à jour des paramètres w_1, w_2, et b avec un taux d'apprentissage $\alpha = 0.1$ à partir des dérivées calculées à l'étape précédente.

Étape 8 : Répétition de la descente de gradient

Maintenant que nous avons effectué une première mise à jour des paramètres w_1, w_2 et b, nous allons poursuivre le processus de **descente de gradient** en répétant les étapes suivantes jusqu'à convergence :

1. **Calcul des prédictions $\hat{y}_i$** : Pour chaque observation, nous recalculons les probabilités en utilisant les nouvelles valeurs de w_1, w_2 et b.

$$\hat{y}_i = \frac{1}{1 + e^{-(w_1 \cdot x_1 + w_2 \cdot x_2 + b)}}$$

2. **Calcul de la fonction de coût Log-Loss** : Une fois les nouvelles prédictions obtenues, nous recalculons la fonction de coût pour évaluer l'erreur totale.

$$\textbf{Log-Loss} = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

3. **Calcul des dérivées** : Nous recalculons les dérivées par rapport aux paramètres w_1, w_2 et b afin de déterminer la direction à suivre pour minimiser la fonction de coût.

   - Dérivée par rapport à w_1 :

$$\frac{\partial J}{\partial w_1} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i1}$$

$i=1$

- Dérivée par rapport à w_2 :

$$\frac{\partial J}{\partial w_2} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_{i2}$$

- Dérivée par rapport à b :

$$\frac{\partial J}{\partial b} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

Mise à jour des paramètres : Nous utilisons les nouvelles dérivées pour mettre à jour w_1, w_2 et b.

- Mise à jour de w_1 :

$$w_1 := w_1 - \alpha \frac{\partial J}{\partial w_1}$$

- Mise à jour de w_2 :

$$w_2 := w_2 - \alpha \frac{\partial J}{\partial w_2}$$

- Mise à jour de b :

$$b := b - \alpha \frac{\partial J}{\partial b}$$

5. **Répétition jusqu'à convergence** : Vous répétez ce processus jusqu'à ce que les changements dans les paramètres w_1, w_2, et b deviennent très petits, indiquant que la fonction de coût est proche d'un minimum.

Critère d'arrêt

Vous pouvez arrêter la descente de gradient lorsque :

- La différence dans la fonction de coût entre deux itérations est inférieure à un seuil (par exemple, 10^{-6}).
- Un nombre maximal d'itérations est atteint.

Exemple d'itérations supplémentaires

1. **Itération 1** :
   - Prédictions $\hat{y}_i$,
   - Log-Loss,
   - Dérivées,
   - Mise à jour des paramètres w_1, w_2, et b.
2. **Itération 2** :
   - Prédictions $\hat{y}_i$,
   - Log-Loss,
   - Dérivées,
   - Mise à jour des paramètres w_1, w_2, et b.

...

Continuez à suivre ces étapes pour ajuster les paramètres jusqu'à ce que la fonction de coût atteigne un minimum.

Conclusion

L'objectif de la descente de gradient est d'**ajuster les paramètres** de manière itérative en suivant la direction opposée du gradient, afin de **minimiser la fonction de coût log-loss**. Le processus est répété jusqu'à ce que le modèle converge, c'est-à-dire que les paramètres atteignent des valeurs stables qui minimisent l'erreur entre les prédictions et les vraies étiquettes.

Vous avez maintenant toutes les étapes pour résoudre une régression logistique avec deux variables, de la formulation de la fonction sigmoïde et de la fonction de coût log-loss à l'application des dérivées et de la descente de gradient.