Université Abdelmalek Essaâdi
Faculté des Sciences
Département de Mathématiques

Année : 2024-2025
MLIA-S5

Série 2

Régression Logistique

Exercice 1 : Classification binaire simple

On dispose d'un jeu de données avec deux caractéristiques x_1 et x_2, et une variable cible $y \in \{0,1\}$. Les données sont les suivantes :

x_1	x_2	y
2	3	1
1	5	0
3	1	1
0	2	0

Le modèle de régression logistique est donné par :

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}},$$

où $x = [1, x_1, x_2]^T$ (ajout du biais), et $\theta = [\theta_0, \theta_1, \theta_2]^T$ est le vecteur de paramètres.

1. Écrivez la fonction coût logistique pour ce problème :
2. Montrer que les gradients pour mettre à jour θ avec la méthode de descente de gradient sont:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}, \quad j = 0,1,2.$$

3. Implémentez une itération de descente de gradient pour un pas d'apprentissage $\alpha = 0.1$.

Exercice 2 : Évaluation d'un classificateur

Un modèle logistique a été entraîné avec les paramètres suivants :

$$\theta_0 = -3, \quad \theta_1 = 0.8, \quad \theta_2 = 1.2.$$

1) Pour chaque exemple de test (x_1, x_2), calculez la probabilité prédite $h_\theta(x)$ et la prédiction binaire $\hat{y}$: ($\hat{y} = 1$ si $h_\theta(x) \geq 0.5$, sinon $\hat{y} = 0$) :

- $x = (2,3)$,
- $x = (1,5)$,

- $x = (0, 1)$.

2) Supposons les vrais labels pour ces exemples soient $y = [1, 0, 0]$. Calculez les métriques suivantes :

a. **Exactitude (Accuracy)** :

$$\text{Exactitude} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total d'exemples}}.$$

b. **Précision (Precision)**, **Rappel (Recall)** et **F1-score** pour la classe $y = 1$.

Exercice 3 : Régression logistique multiclasse (One-vs-All)

On dispose d'un jeu de données pour une classification à 3 classes : $y \in \{0, 1, 2\}$. La régression logistique doit être utilisée avec la méthode **One-vs-All**.

Entraînez trois classificateurs logistiques $h_\theta^{(k)}(x)$ pour chaque classe $k \in \{0, 1, 2\}$, où :

$$h_\theta^{(k)}(x) = P(y = k|x, \theta).$$

1. Écrivez la fonction coût pour un classificateur donné k.
2. Implémentez les mises à jour des gradients pour un seul classificateur.
3. Une fois les trois modèles entraînés, donnez la règle de prédiction pour attribuer une classe :
$$\hat{y} = \arg\max_k h_\theta^{(k)}(x).$$

Exemple d'application

Pour un point $x = [1, 2]^T$, et les paramètres des trois modèles :

$$\theta^{(0)} = [-1, 0.5, 0.3], \quad \theta^{(1)} = [0.2, -0.1, 0.8], \quad \theta^{(2)} = [0.7, -0.4, 0.2],$$

calculez la probabilité associée à chaque classe et donnez la classe prédite.

Exercice 4 : Application avancée de la régression logistique

Nous considérons un ensemble de données de taille $m = 4$, avec deux caractéristiques x_1 et x_2 (y compris un biais implicite ajouté $x_0 = 1$ pour toutes les observations) et leurs labels y correspondants.

Exemple	x_1	x_2	y
1	0.5	1.2	0
2	1.5	0.7	1
3	2.0	1.0	1
4	1.0	1.5	0

Appliquer la regression logistique à ce problème

Exercice 5 : Analyse multicatégorie avec régression logistique

Considérons un problème de classification avec trois classes C_1, C_2, C_3. Vous disposez d'un jeu de données avec deux caractéristiques x_1 et x_2, ainsi que les étiquettes correspondantes.

1. Développez un modèle de régression logistique multinomial (softmax) pour prédire la probabilité d'appartenance à chaque classe.
2. Ècrivez la fonction coût multinomiale adaptée.
3. Implémentez une descente de gradient pour ajuster les paramètres.

Exercice 6 : Interaction des caractéristiques dans un modèle de régression logistique

Dans cet exercice, vous allez étudier l'impact des interactions entre les caractéristiques dans un modèle de régression logistique. Vous êtes invités à comparer deux approches : avec et sans termes d'interaction.

Données initiales

Les données suivantes sont fournies :

x_1	x_2	x_3	y
0.5	1.2	0.8	0
1.5	0.7	1.0	1
2.0	1.0	1.5	1
1.0	1.5	1.2	0

1. **Modèle avec interactions** Implémentez une régression logistique où la fonction hypothèse est donnée par :
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T \tilde{x}}},$$
avec $\tilde{x} = [1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3]^T$. Suivez les étapes suivantes :
   (a) Construisez la matrice étendue $\tilde{X}$ à partir des données initiales en incluant les termes d'interaction (x_1x_2, x_1x_3, x_2x_3).
   (b) Entraînez le modèle à l'aide de la descente de gradient en initialisant $\theta = 0$.
   (c) Affichez les paramètres θ appris et les prédictions pour un point de test donné $x_{\text{test}} = [1.2, 0.8, 1.1]$.

Exercice 7 : Comparaison de modèles de régression logistique avec et sans interactions

On considère un problème de classification binaire avec les données suivantes :

x_1	x_2	x_3	Classe (y)
0.5	1.2	0.8	0
1.5	0.7	1.0	1
2.0	1.0	1.5	1
1.0	1.5	1.2	0

(a) **Modèle avec interactions** Implémentez une régression logistique qui inclut :

- Un biais ($x_0 = 1$),
- Les caractéristiques originales (x_1, x_2, x_3),
- Les termes d'interaction ($x_1 \cdot x_2$, $x_1 \cdot x_3$, $x_2 \cdot x_3$).

Utilisez une descente de gradient pour ajuster les paramètres du modèle (θ). Testez la prédiction pour le point $[1.2, 0.8, 1.1]$. Affichez la probabilité prédite et la classe attribuée ($y = 0$ ou $y = 1$).

(b) **Modèle sans interactions** Implémentez une régression logistique en utilisant uniquement :

- Un biais ($x_0 = 1$),
- Les caractéristiques originales (x_1, x_2, x_3).

Testez également la prédiction pour le point $[1.2, 0.8, 1.1]$. Affichez la probabilité prédite et la classe attribuée.

(c) **Comparaison des résultats**

i. Comparez les paramètres appris (θ) dans les deux cas.

ii. Comparez les probabilités prédites et les classes attribuées pour le point testé.

iii. Discutez l'impact des termes d'interaction sur la performance du modèle. Dans quel cas leur inclusion peut-elle être bénéfique ?

Exercice 8 : Régularisation

Pour éviter le sur-apprentissage sur un petit ensemble de données, ajoutez une régularisation L_2.

(a) Développez la fonction coût régularisée :

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2.$$

(b) Implémentez la descente de gradient modifiée pour tenir compte de la régularisation.

(c) Testez sur un ensemble de données où une caractéristique est beaucoup plus importante que les autres.