Université Abdelmalek Essaâdi
Faculté des Sciences
Département de Mathématiques

Année : 2024-2025
MLIA-S5

Série 3

Super Vecteur Machine

Exercice 1 : Compréhension des différences entre SVM et régression logistique

1. Écrivez les fonctions de coût utilisées pour :

- La régression logistique :

$$J(w,b) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(h_\theta(x_i)) + (1-y_i)\log(1-h_\theta(x_i))\right],$$

où $h_\theta(x_i) = \frac{1}{1+e^{-(w^T x_i + b)}}$.

- Les machines à vecteurs de support (SVM) avec la perte hinge :

$$J(w,b) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\max(0, 1 - y_i(w^T x_i + b)).$$

2. Comparez les deux approches en termes d'objectifs :

- La régression logistique cherche à minimiser une fonction de coût probabiliste.
- Les SVM cherchent à maximiser la marge tout en minimisant la perte des points mal classés.

3. Montrez mathématiquement pourquoi la fonction de coût SVM pénalise les points mal classés différemment de la régression logistique.

Exercice 2 : Implémentation mathématique de SVM pour classification

Soit un ensemble de données $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, où $y_i \in \{-1, +1\}$.

1. Écrivez le problème d'optimisation dual pour les SVM :

$$\max_{\alpha} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j,$$

sous les contraintes :

$$\sum_{i=1}^{n}\alpha_i y_i = 0 \quad \text{et} \quad \alpha_i \geq 0.$$

2. Expliquez comment résoudre ce problème pour obtenir :

- $w = \sum_{i=1}^{n} \alpha_i y_i x_i$,
- b à partir d'un vecteur de support :

$$b = y_k - \sum_{i=1}^{n} \alpha_i y_i K(x_i, x_k),$$

où $K(x_i, x_k)$ est le noyau et x_k est un vecteur de support ($\alpha_k > 0$).

3. Discutez comment intégrer un noyau $K(x_i, x_j)$ dans cette formulation.

Exercice 3 : Régression logistique vs SVM sur un exemple simple

Considérez les points suivants dans $\mathbb{R}^2$:

$$\{(1, 2, +1), (2, 3, +1), (3, 3, -1), (4, 5, -1)\}.$$

1. Utilisez la régression logistique pour calculer les probabilités de classification pour un point $(3, 4)$. Utilisez :

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

avec :

$$w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad b = 0.$$

2. Appliquez le SVM avec une fonction de marge :

- Calculez le score $s = w^T x + b$.
- Classez le point selon le signe de s.

3. Comparez les résultats des deux méthodes.

Exercice 4 : Régularisation dans SVM et régression logistique

1. Pour la régression logistique régularisée, écrivez la fonction de coût avec un terme de régularisation L_2 :

$$J(w, b) = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(h_\theta(x_i)) + (1 - y_i) \log(1 - h_\theta(x_i))] + \frac{\lambda}{2} \|w\|^2.$$

2. Pour les SVM, expliquez comment le paramètre C contrôle le compromis entre la régularisation et la perte hinge :

$$J(w, b) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i + b)).$$

3. Donnez des exemples où la régularisation peut améliorer les performances pour des données bruitées.

Exercice 5 : Théorie et interprétation des SVM

1. Expliquez pourquoi les SVM cherchent à maximiser la marge fonctionnelle définie comme :

$$\gamma = \min_i \frac{y_i(w^T x_i + b)}{\|w\|}.$$

Comment cela contribue à une meilleure généralisation ?

2. Déduisez les conditions de KKT (Karush-Kuhn-Tucker) pour les SVM et expliquez leur rôle dans la résolution du problème.

3. Soit un SVM entraîné sur un ensemble de données. On ajoute un point x_k très éloigné de l'hyperplan mais dans la classe correcte ($y_k(w^T x_k + b) > 1$). Discutez comment cela affecte le modèle et la marge.

—

Exercice 6 : Régularisation et généralisation

1. Pour la régression logistique, considérez les deux fonctions de coût suivantes :

$$J(w) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(h_\theta(x_i)) + (1 - y_i)\log(1 - h_\theta(x_i))\right],$$

avec régularisation L_1 et L_2 :

$$J_{L_1}(w) = J(w) + \lambda\|w\|_1, \quad J_{L_2}(w) = J(w) + \frac{\lambda}{2}\|w\|_2^2.$$

- Comparez L_1 et L_2 en termes d'impact sur les poids w. - Donnez des exemples où l'utilisation de L_1 est préférable à L_2.

2. Pour un modèle SVM, expliquez le compromis entre la largeur de la marge et les erreurs de classification contrôlé par C. Comment ajuster C pour un ensemble de données bruitées ?

—

Exercice 7 : Cas pratique avec un SVM linéaire

Considérez un ensemble de données dans $\mathbb{R}^2$:

$$\{(1, 1, +1), (2, 2, +1), (2, 0, -1), (0, 1, -1)\}.$$

1. Calculez l'hyperplan séparateur (w et b) pour un SVM linéaire en résolvant le problème primal :

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{sous les contraintes } y_i(w^T x_i + b) \geq 1,\ \forall i.$$

2. Vérifiez si le point $(1.5, 1.5)$ appartient à la classe $+1$ ou -1 en utilisant l'hyperplan trouvé.

3. Appliquez un noyau polynomial de degré 2, $K(x, z) = (x^T z + 1)^2$, et reformulez le problème dual. Résolvez-le pour obtenir les coefficients α_i.

—

Exercice 8 : Régression logistique multiclasse

1. Expliquez comment étendre la régression logistique binaire à un problème multiclasse en utilisant : - Une approche un-contre-tous (OvR). - Une approche un-contre-un (OvO).

2. Considérez les données suivantes avec trois classes ($y \in \{1, 2, 3\}$) :

$$\{(1, 1, 1), (2, 2, 1), (1, 3, 2), (3, 3, 2), (4, 4, 3), (5, 5, 3)\}.$$

Construisez trois modèles logistiques pour la méthode OvR. Écrivez les équations de décision pour chaque modèle.

3. Discutez les avantages et inconvénients des approches OvR et OvO.

—

Exercice 9 : Visualisation et interprétation

1. Tracez dans un espace 2D les données suivantes :

$$\{(2, 3, +1), (3, 4, +1), (1, 1, -1), (0, 2, -1)\}.$$

Entraînez un SVM linéaire et un modèle de régression logistique. Superposez : - L'hyperplan séparateur du SVM. - Les courbes de niveau de probabilité $P(y = +1|x) = 0.5$, $P(y = +1|x) = 0.8$, etc., pour la régression logistique.

2. Expliquez les différences dans les frontières de décision obtenues et leur interprétation.

—

Exercice 10 : Régression logistique régulière vs non régulière

Considérez les données dans $\mathbb{R}^2$:

$$\{(1, 1, 0), (2, 2, 0), (3, 3, 1), (4, 4, 1)\}.$$

1. Entraînez un modèle de régression logistique sans régularisation. Trouvez les paramètres w et b qui minimisent $J(w, b)$.

2. Ajoutez une régularisation L_2 avec $\lambda = 1$ et réentraînez le modèle. Comparez les valeurs des poids w obtenus.

3. Tracez les frontières de décision dans les deux cas et discutez l'impact de la régularisation.

—