

Data Science

Travaux dirigés - Séance 1

1. Cadre de la classification

On considère un problème de classification de données dans un espace X continu dans un ensemble fini de classes $C = \{C_1, \dots, C_K\}$. On note $p(x)$ la densité de probabilité en x , $p(C_k/x)$ la probabilité a posteriori de la classe C_k sachant x , $P(C_k)$ la probabilité a priori de la $k^{\text{ème}}$ classe.

Une règle de décision est une application de X dans C . On considère qu'il s'agit d'une fonction déterministe qui associe une classe à une forme de X .

1. Probabilité d'erreur d'une règle de décision

On caractérise une règle de décision par sa probabilité d'erreur, c'est à dire l'espérance de la fonction indicatrice $1_{[r(x) \neq C(x)]}$ sur l'espace joint $X \times C$. Exprimer cette espérance.

2. Probabilité d'erreur d'une règle de décision pour un x donné dans X

Exprimer la probabilité de se tromper pour une donnée x particulière en fonction des quantités décrites plus haut. En déduire la règle optimale de classification pour cet exemple. L'erreur de cette règle optimale est elle nulle ? Pourquoi ?

3. En déduire la règle optimale de classification pour le problème de classification considéré.

2. Régression linéaire – résolution analytique

Comme vu en cours, le critère de qualité d'une régression linéaire, apprise sur un ensemble de N points $\{(x^i, y^i), i = 1..N\}$ s'écrit :

$$R_{emp}(\mathbf{w}) = \sum_{i=1}^N l(f_{\mathbf{w}}(x^i), y^i)$$

où l est la fonction de perte (ou loss) et où w désigne les paramètres de la régression linéaire. On considère un problème de régression linéaire à valeur réelle, c'est à dire que la prédiction du modèle de régression est une valeur réelle.

1. Critère d'erreur quadratique

Réécrivez ce critère en utilisant la fonction de perte *erreur quadratique* et en faisant apparaître explicitement w .

2. Exprimer le gradient de $C(w)$ par rapport à w .

La fonction $C(w)$ est convexe en w , ce qui signifie qu'elle n'a qu'un minimum et c'est un minimum global. Le vecteur w optimisant le critère $C(w)$ est donc le seul pour lequel le gradient s'annule. Raison pour laquelle l'expression du gradient nous intéresse.

3. Mise sous forme vectorielle

En formant la matrice X des données (autant de lignes que d'exemple et autant de colonnes que la dimension des données), exprimez comment obtenir Y le vecteur des prédictions, par le modèle de paramètres w , pour toutes les données de X .

Exprimez maintenant le critère $C(w)$ à l'aide de X et de w .

4. Dérivations matricielles

Montrer que si A est une matrice $n \times p$ et z un vecteur $p \times 1$ alors $\frac{\partial Az}{\partial z} = A^T$

Montrer que si A est une matrice carrée $p \times p$ et z un vecteur $p \times 1$ alors $\frac{\partial z^T A z}{\partial z} = 2A^T z$

5. Solution analytique de la régression linéaire

En déduire la solution analytique du problème de régression linéaire, en supposant que la matrice $X^T X$ est inversible.

3. Construction d'estimateurs

Soit (X_1, \dots, X_n) un échantillon de loi uniforme $U([0, \theta])$ avec θ inconnu. Le but de l'exercice est d'estimer θ à l'aide de la méthode des moments et du maximum de vraisemblance. On rappelle que la densité de X s'écrit $f(x) = \frac{1}{\theta} 1_{[x < \theta]}$ et son espérance est $E[X] = \frac{\theta}{2}$.

1. Méthode des moments

Tout d'abord on cherche la fonction f telle que $\theta = E[f(X_1)]$. Ensuite on remplace l'espérance par la moyenne empirique et on conclut à l'aide de la loi des grands nombres.

2. Méthode du maximum de vraisemblance

On écrit la fonction de vraisemblance du modèle définie par $V_\theta(X_1, \dots, X_n) = f_\theta(X_1) \dots f_\theta(X_n)$ et on cherche la valeur de θ qui maximise cette fonction.