Analyse de données - Résumé

November 28, 2023

THEVENET Louis

Table des matières

1. Introduction - Evaluating classifiers

Définition 1.1: Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	60	10
Actual Positive	5	25

Définition 1.2: Precision, Recall and F1-score

$$\text{Precision} = \frac{\text{True positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2. Statistical Classification

2.1. Bayesian Rule

Définition 2.1.1:

Pour K classes $w_1, ..., w_K$ et $x = (x_1, ..., x_p)^T$ observations

$$d : \begin{cases} X \to A \\ x \mapsto d(x) \end{cases}$$

où A est un ensemble d'actions $a_1, ..., a_q$ où $a_k =$ assigne x à la classe $w_k, \forall k \in [\![1, ..., n]\!]$

On peut ajouter $a_0 =$ ne pas classer x pour avoir une option de rejet.

Théorème 2.1.1: Bayesian Rule

- Probabilité *à priori* de la classe w_k : $P(w_k)$
- Densité de probabilité de x sachant la classe w_k : $f(x \mid w_k)$

On en conclut la probabilité *à posteriori* que x appartiennent à w_k :

$$P(w_k \mid x) = \frac{f(x \mid w_k) P(w_k)}{f(x)}$$

avec $f(x) = \sum_{k=1}^{K} f(x \mid w_k) P(w_k)$

2.2. MAP Classifier

Définition 2.2.1:

$$d^*(x) = a_j \Leftrightarrow \forall k \in [\![1, ..., K]\!] : P(w_j \mid x) \geq P(w_k \mid x)$$

Définition 2.2.2:

Classes équiprobables : classificateur de maximum de vraisemblance

$$d^*(x) = a_j \Leftrightarrow \forall k \in [\![1, ..., K]\!] : P(x \mid w_j) \geq P(x \mid w_k)$$

Proposition 2.2.1: Le MAP classifier minimise la probabilité d'erreur :

$$P_e = \sum_{k=1}^{K} P[d(x) = a_k \cap x \notin w_k]$$

3. Support Vector Machine (SVM)

Ici on associe des 1et -1 et on définit un hyperplan (une droite par exemple)

Définition 3.1:

$$\mathcal{B} = \{(x_{1,1}), ..., (x_n, y_n)\}$$

où $x_1, ..., x_n \in (\mathbb{R}^p)^n$ et $y_1, ..., y_n$ sont booléens tels que

$$\forall i \in [\![1, ..., n]\!] y_i = \begin{cases} 1 \text{ si } x_i \in w_1 \\ -1 \text{ si } x_i \in w_2 \end{cases}$$

L'hyperplan : $g_{w,b}(x) = w^T x - b = 0$

avec

$$g_{w,b}(x_i) \begin{cases} > 0 \text{ si } x_i \in w_1 \\ < 0 \text{ si } x_i \in w_2 \end{cases}$$

On classifie de la manière suivante : $f(x) = \operatorname{sign}\left[g_{w,b}(x)\right]$

Définition 3.2: Formulation du problème (hyperplan séparateur optimal)

Marge de x_i avec label y_i (distance à l'hyperplan) :

$$\gamma_i(\tilde{w}) = \gamma_{i(w,b)} = \left(y_i \frac{w^T x_i - b}{\|w\|}\right)$$

Marge du set de donnée : $\gamma_{\mathcal{B}}(\tilde{w}) = \min_i \gamma_i(\tilde{w})$

Théorème 3.1: Primal formulation

$$\begin{cases} \min_{w \in \mathbb{R}^n, b \in \mathbb{R}} \frac{1}{2} \|w\|^2 \\ \forall i \in [\![1, ..., n]\!] : y_i (w^T x_i - b) \geq 1 \end{cases}$$

Car on veut maximiser $\gamma_{\mathcal{B}}(\tilde{w}) = \frac{1}{\|w\|}$

On maximise le min des distances à l'hyperplan