

Analyse de données - Résumé

November 28, 2023

THEVENET Louis

Table des matières

| | |
|---|---|
| 1. Introduction - Evaluating classifiers | 1 |
| 2. Statistical Classification | 1 |
| 2.1. Bayesian Rule | 1 |
| 2.2. MAP Classifier | 2 |
| 3. ACP - Analyse en Composantes Principales | 3 |

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, \dots, w_K et $x = (x_1, \dots, x_p)^T$ observations

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

où A est un ensemble d'actions a_1, \dots, a_q où $a_k =$ assigne x à la classe $w_k, \forall k \in \llbracket 1, \dots, n \rrbracket$

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 $w_k : f(x | w_k)$

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

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

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

2.2. MAP Classifier**Définition 2.2.1:**

$$d^*(x) = a_j \Leftrightarrow \forall k \in \llbracket 1, \dots, K \rrbracket : P(w_j | x) \geq P(w_k | x)$$

Définition 2.2.2:

Classes équiprobables : classificateur de maximum de vraisemblance

$$d^*(x) = a_j \Leftrightarrow \forall k \in \llbracket 1, \dots, K \rrbracket : P(x | w_j) \geq P(x | w_k)$$

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. ACP - Analyse en Composantes Principales