

Apprentissage

Aubin SIONVILLE

MPI Clemenceau - 2021-2023

Définitions

Signature de données

Une donnée est un n -uplet de paires (nom, ensemble) :
 $(\text{nom}_1 : E_1, \text{nom}_2 : E_2, \dots, \text{nom}_n : E_n)$

Donnée

Étant donné une signature $\mathbb{S} = (n_1 : S_1, \dots, n_n : S_n)$,
Une donnée est un vecteur $\bar{v} = (v_1, \dots, v_n) \in S_1 \times \dots \times S_n$

Jeu de données

Étant donné une signature \mathbb{S}
Un jeu de données est un ensemble fini
de vecteurs de signature \mathbb{S}

Jeu de données classifié

Étant donné une signature \mathbb{S} et un ensemble fini de classes \mathcal{C} ,
Un jeu de données classifié est la donnée
– d'un jeu de données S
– d'une fonction de classification $f : S \rightarrow \mathcal{C}$

Apprentissage supervisé

k-plus proches voisins

Entrée : Un jeu de données classifié (S, c) , un entier k et un vecteur \bar{v} .

On trie S par distance croissante de \bar{v} en d_1, \dots, d_n
On crée et initialise à 0 un dictionnaire $D : \mathcal{C} \rightarrow \mathbb{N}$
Pour i allant de 1 à k faire :
 On incrémente $D[c(d_i)]$
On retourne la classe c qui maximise $D[c]$

Matrice de confusion

Pour algorithme de prédiction \mathcal{A} sur un jeu de données classifié (S, c)
On appelle matrice de confusion de \mathcal{A} la matrice M telle que
 $\forall (i, j) \in \mathcal{C}^2, M_{i,j} =$ nombre de vecteurs \bar{v} de S tels que $\begin{cases} \text{L'algorithme prédit que la classe de } \bar{v} \text{ est } i \\ \text{La véritable classe de } \bar{v} \text{ est } j \end{cases}$

Notation : Composantes d'un jeu de données

Étant donné un jeu de données $S, \forall v \in S,$

$$S^{\leq iv} = \{v' \in S \mid v'_i \leq v_i\}$$

$$S^{> iv} = \{v' \in S \mid v'_i > v_i\}$$

Création d'un arbre k-dimensionnel (\mathbb{F})

Entrée : Un jeu de donnée \mathcal{V} et $i \in \llbracket 0, n-1 \rrbracket$.

Si \mathcal{V} est vide,

Retourner Vide

Sinon,

On cherche $v \in \mathcal{V}$ tel que v_i est la médiane de $\{u_i \mid u \in \mathcal{V}\}$

Retourner un noeud $\left\{ \begin{array}{l} \text{Fils gauche : } \mathbb{F}((\mathcal{V} \setminus \{u\})^{\leq iv}, (i+1) \bmod n) \\ \text{Valeur : } (v, i) \\ \text{Fils droit : } \mathbb{F}((\mathcal{V} \setminus \{u\})^{> iv}, (i+1) \bmod n) \end{array} \right.$

TO BE CONTINUED