24. Exemples d'algorithmes d'apprentissage supervisés et non supervisés.

Jill-Jênn Vie

4 janvier 2022

Programme MPI / NSI

Apprentissage supervisé

- Algorithme des k plus proches voisins (NSI 1re)
- Matrice de confusion
- Arbres k-dimensionnels
- Arbre de décision (algorithme ID3)

Apprentissage non supervisé

- Algorithme des k-moyennes
- Algorithme de classification hiérarchique ascendante

Jeux et heuristiques

- Jeux d'accessibilité à 2 joueurs sur un graphe
- Algorithme min-max avec une heuristique
- Élagage alpha-bêta
- Algorithme A* (ex. projet NSI)

Apprentissage machine

- Apprendre à partir d'exemples
- Généraliser, inférer de nouvelles connaissances

Exemples ?

Formalisation apprentissage supervisé

Échantillons : $(\boldsymbol{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y}$

Les caractéristiques (descripteurs) $\boldsymbol{x}_i$ et les étiquettes y_i.

Habituellement $\mathcal{X} = \mathbf{R}^d$
Si $\mathcal{Y}$ est discret : classification, sinon régression.

Modèle $f_\theta : \mathcal{X} \to \mathcal{Y}$ ayant des paramètres θ

Une fonction d'erreur $\mathcal{L}$ à minimiser qui dépend des paramètres θ.

But : trouver $\theta = \operatorname{argmin}_\theta \mathcal{L}(\theta, \boldsymbol{x}_i, y_i)$

Régression linéaire

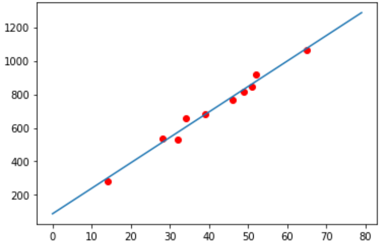
Modèle $f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\theta} \simeq y_i$

Erreur : moindres carrés
$\mathcal{L} = \sum_i ||\boldsymbol{x}_i^T \boldsymbol{\theta} - y_i||^2$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = \sum_i 2\boldsymbol{x}_i(\boldsymbol{x}_i^T \boldsymbol{\theta} - y_i) = 0$$

$$\left(\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^T\right) \boldsymbol{\theta} = \sum_i \boldsymbol{x}_i y_i$$

$$\boldsymbol{\theta} = (X^T X)^{-1} X^T \boldsymbol{y}$$

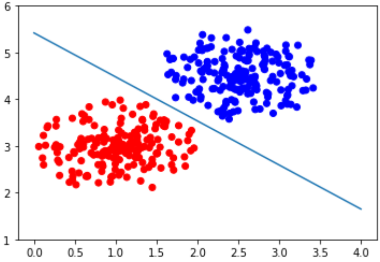


Régression logistique (classification)

$f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) = \sigma(\boldsymbol{x}_i^T \boldsymbol{\theta}) = \Pr(Y_i = 1)$
où $\sigma : x \mapsto 1/(1 + e^{-x})$

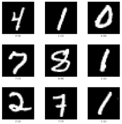
Pratique car $\sigma' = \sigma(1 - \sigma)$

Pas de formule close pour $\boldsymbol{\theta}$
donc apprentissage de $\boldsymbol{\theta}$ par
méthode de Newton



Exemple : reconnaissance de caractères

Algorithme du perceptron (Rosenblatt, 1957)



k plus proches voisins (Russell et Norvig, 2020, p. 687)

Algorithme k plus proches voisins parmi n en d dimensions

---

pour tout i de 1 à n **faire**
    calculer la distance de $\boldsymbol{x}$ à $\boldsymbol{x}_i$
fin pour
identifier les k plus proches voisins de $\boldsymbol{x}$
calculer la moyenne de leurs valeurs, ou la classe majoritaire

---

- Complexité de trouver les voisins de $\boldsymbol{x}$ dans X ? $O(nd + kn)$
- Avec *median of medians* : $O(nd)$
- Peut être accéléré avec un *k-d tree*[1] en $O((n + k) \log n)$

Pour une version recommandation dans des bases de données, cf. (Abiteboul et al., 2011, p. 365)

---

1. « arbre k dimensionnel » en vieux françois

Formalisation apprentissage non supervisé

Échantillons : $(\boldsymbol{x}_i) \in \mathbf{R}^d$

Les caractéristiques (descripteurs) $\boldsymbol{x}_i$

But : apprendre une représentation des données, clustering

Algorithme des K-moyennes (Bishop, 2006, p. 424)

$$\text{Minimiser } \mathcal{L} = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} ||\boldsymbol{x}_i - \boldsymbol{\mu}_k||^2 \quad \sum_k r_{ik} = 1$$

Le cluster de $\boldsymbol{x}_i$ est le k tel que $r_{ik} = 1$. $\mathcal{C}_k \triangleq \{i | r_{ik} = 1\}$

Minimisation

$$r_{ik} = \begin{cases} 1 & \text{si } k = \text{argmin}_j \, ||\boldsymbol{x}_i - \boldsymbol{\mu}_j|| \\ 0 & \text{sinon.} \end{cases}$$

Moyenne

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}_k} = \sum_i 2 r_{ik} (\boldsymbol{\mu}_k - \boldsymbol{x}_i) = \sum_{i \in \mathcal{C}_k} 2(\boldsymbol{\mu}_k - \boldsymbol{x}_i) = 0$$

$$|\mathcal{C}_k| \boldsymbol{\mu}_k = \sum_{i \in \mathcal{C}_k} \boldsymbol{x}_i$$

$$\boldsymbol{\mu}_k = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \boldsymbol{x}_i .$$

- Abiteboul, Serge et al. (2011). *Web Data Management*. Cambridge University Press. URL : http://webdam.inria.fr/Jorge/files/wdm.pdf.
- Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*.
- Russell, Stuart et Peter Norvig (2020). *Artificial Intelligence: A Modern Approach*.