Donnée et Apprentissage

Nicolas Vayatis
ENS Cachan

Table des matières

Notre professeur fait partie du labo CMLA (*Centre de Mathématiques et de Leurs Applications*) faisant parti du MLMDA (*Machine Learning and Massive Data Analysis*) étudiant principalement :

- Optimisation séquentielle et apprentissage actif
- Machine Learning sur des signaux temporels
- Processus de diffusion dans les réseaux avec des applications dans la santé publique (virus) et propagation d'information (virus informatique et réseaux sociaux)

Le laboratoire effectue principalement de la recherche (publications) et de la vulgarisation (pour les industries, les politiques, etc) et également des outils numérique (codes, portails, logiciel).

1 Introduction

Il existe trois grands problèmes avec les données :

- Classification (Apprentissage supervisé) : labels discret ($\rightarrow$ catégories)
- Régression (Apprentissage supervisé) : labels continus ($\rightarrow$ prédire un prix)
- Clustering ou segmentation (Apprentissage non supervisé $\rightarrow$ pas de labels)

Il existe également deux approches :

- L'approche statistique classique *paramétrique*
- L'approche par apprentissage *non paramétrique*

Principe : De par des données historiques $Z_i = (\underbrace{X_i}_{\text{mesures}}, \underbrace{Y_i}_{\substack{\text{label si} \\ \text{cadre supervisé}}})$, après un apprentissage A fournit une transformée $\hat{f}$ permettant une *règle de décision*, qui est le résultat de $\hat{f}(X_{n+1})$. La fonction $\hat{f}$ est testée, ce qui permet de juger ses performances avant son utilisation.

En statistique classique :

- On fixe la famille de lois qui génère les Z_i
- On en estime les paramètre de la loi, par exemple avec des méthodes de maximum de vraisemblance
- On fait du plug-in pour construire la règle de décision

En apprentissage :

- On fixe une structure de fonctions pour les règles de décision Par exemple

$$f_{\omega,\omega_0}(x) = \begin{cases} \text{"chat"} & \text{si} \quad \omega^T x + \omega_0 > 0 \\ \text{"chien} & \text{sinon} \end{cases}$$

- On fixe un critère de performance, par exemple :

$$\hat{L}_n(f) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{(f(X_i) = \text{"chien"} \wedge Y_i = \text{"chat"}) \vee (f(X_i) = \text{"chat"} \wedge Y_i = \text{"chien"}\}$$

- L'algorithme d'apprentissage doit minimiser $\hat{L}_n(f_{\omega,\omega_0})$ sur $\mathcal{F}$

Loi du mélange Soit X un vecteur aléatoire sur $\mathbb{R}^d$ suivant une loi de mélange. On considère K composantes de loi de densité f_k pour $1 \leq k \leq K$.

Un paramètre du mélange est $p = (p_1, ...p_K) \in \Delta_K \subset \mathbb{R}^K$ ou Δ_K est un simplexe de $\mathbb{R}^K$:

$$\Delta_K = \left\{p = (p_1, ..., p_K)^T \in \mathbb{R}_+^K : \sum_{k=1}^{K} p_k = 1\right\}$$

La densité du mélange est la loi de X :

$$f_X(x) = \sum_{k=1}^{K} p_k f_k(x) \qquad \forall x \in \mathbb{R}^d$$

Variables latentes (labels) On note $Y = (Y_1, ..., Y_k)^T$ avec $Y_K \in \{0, 1\}$, $\sum_{k=1}^{K} Y_k = 1$. Les Y_k sont des drapeaux exclusifs.

Quels Problème décisionnels ?

1. Clustering ou classification non supervisée :

1.1 Expectation Maximisation

Definition 1 (Maximum de vraisemblance)**.** *Soit $Z_1, ..., Z_n$ des variables aléatoires indépendantes et de même loi $z \mapsto f(z, \theta*)$, $\theta*$ paramètre inconnu. On appelle maximum de vraisemblance l'estimateur de $\theta*$ défini par*

$$\hat{\theta} \in \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f(Z_i, \theta)$$

On utilise souvent la log-vraisemblance du modèle du mélange gaussien pour l'observation X :