Apprentissage statistique

Liva Ralaivola

Master I Informatique LIF, UMR 6166 CNRS
Université de Provence
liva.ralaivola@lif.univ-mrs.fr

14 janvier 2008

De quoi s'agit-il ?

Définition (Une définition de haut niveau)

Etant donné un ensemble fini *de données identiquement et indépendamment distribuées selon un distribution D fixe et inconnue, déterminer une fonction h d'une famille de fonctions $\mathcal{H}$ telle qu'elle permette une modélisation adéquate de certaines caractéristiques particulières des éléments tirés selon D*

Quelques exemples

- Etant donné un ensemble fini de chansons étiquetées comme `plaisantes` ou `non plaisantes`, créer un *classifieur* de musiques capable d'estimer si une nouvelle musique vous plaira ou non
- Etant donné un ensemble fini de mesures sur le temps de décharge de plusieurs examplaires d'une même batterie en fonction des conditions de leur utilisation, déterminer une fonction qui est capable de prédire précisément le temps d'utilisation restant d'une batterie de la même série
- ...

De quoi s'agit-il ?

Définition (Une définition de haut niveau)

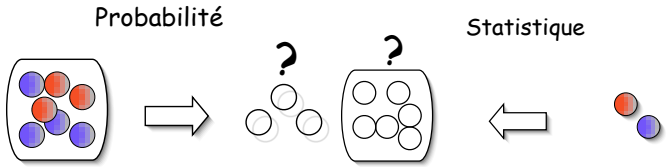
Etant donné un ensemble fini *de données identiquement et indépendamment distribuées selon un distribution D fixe et inconnue, déterminer une fonction h d'une famille de fonctions $\mathcal{H}$ telle qu'elle permette une modélisation adéquate de certaines caractéristiques particulières des éléments tirés selon D*

Définition (GENERALISATION ! ! !)

Notion primordiale. La fonction h apprise est-elle capable de modéliser de façon précise des caractéristiques d'instances non vues au cours de l'apprentissage.

Les origines : statistiques inférentielles

Définition (Statistique $\neq$ probabilité)

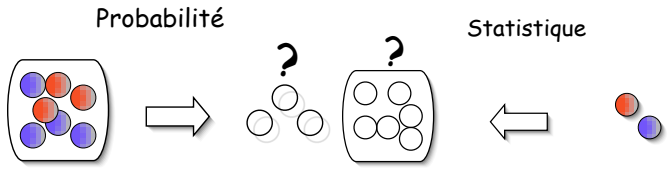


Amusons-nous

- Combien de chars : comment estimer le nombre de chars d'une armée à partir de l'observation des numéros de série de n chars (2nde guerre mondiale, guerre de Corée)
- Mise en place d'un protocole d'enquête anonyme pour l'estimation de la proportion de personnes usant de substances prohibées au sein d'une population

Les origines : statistiques inférentielles

Définition (Statistique $\neq$ probabilité)



Particularité de l'apprentissage statistique

Egale importance des aspects

- de modélisation (semi-paramétrique, non paramétrique)
- algorithmiques
- théoriques

Cadre de Vapnik

Apprentissage à partir de données

Formalisation

- $\mathcal{S}$ échantillon aléatoire sur $\mathcal{X} \times \mathcal{Y}$ selon $p(\mathbf{x}, y) = p(\mathbf{z})$ *inconnue*
- $\mathcal{F}$ famille de fonctions

$$\text{Trouver } f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \int Q(\mathbf{z}, f(\mathbf{z}))\, dp(\mathbf{z})$$

Cadre de Vapnik

Apprentissage à partir de données

Formalisation

- $\mathcal{S}$ échantillon aléatoire sur $\mathcal{X} \times \mathcal{Y}$ selon $p(\mathbf{x}, y) = p(\mathbf{z})$ *inconnue*
- $\mathcal{F}$ famille de fonctions

$$\text{Trouver } f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \underbrace{\int Q(\mathbf{z}, f(\mathbf{z}))\, dp(\mathbf{z})}_{R(f)}$$

Cadre de Vapnik

Conception d'algorithmes $\mathcal{A}$

Algorithme $\mathcal{A}$

$$\mathcal{A} : \bigcup_{\ell=1}^{\infty} (\mathcal{X} \times \mathcal{Y})^{\ell} \to \mathcal{F}$$

minimisant l'erreur de généralisation

$$\begin{aligned} R(\mathcal{A}, \mathcal{S}) &= R(\mathcal{A}(\mathcal{S})) - \min_{f \in \mathcal{F}} R(f) \\ &= R(\mathcal{A}(\mathcal{S})) - R(f^*) \end{aligned}$$

Principes d'induction

- minimisation de fonctions de risque empirique régularisé
- maximisaton de la marge
- ...

Sujets abordés

Classification supervisée

Ensemble d'apprentissage $\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\}$, $\mathbf{x}_i \in \mathcal{X}$, $y_i \in \{-1, +1\}$

- arbres de décision
- régression logistique
- perceptron, perceptron multi-couches, perceptron à noyau
- k-plus-proches voisins

Régression supervisée

Ensemble d'apprentissage $\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\}$, $\mathbf{x}_i \in \mathcal{X}$, $y_i \in \mathbb{R}$

- régression linéaire

Apprentissage non supervisé

Ensemble d'apprentissage $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\}$, $\mathbf{x}_i \in \mathcal{X}$

- K-moyennes
- classification hiérarchique ascendante
- analyse en composantes principales, multi-dimensional scaling, locally-linear embedding

Sujets chauds !

Nombreux

- apprentissage semi-supervisé
- apprentissage actif
- filtrage collaboratif
- espace d'entrées et de sorties structurés
- ranking
- modèles graphiques
- question answering
- résumé de textes
- ...