Introduction à la classification Introduction & Perceptron

Guillaume Wisniewski guillaume.wisniewski@limsi.fr

Université Paris Sud — LIMSI

janvier 2016

Première partie I

La tâche de classification supervisée

Ce que l'on a vu...

Régression

- ▶ prédire le temps qu'il fera demain
- ▶ prédire la taille d'un cancer
- ▶ prédire le prix d'un appartement
- •
- ⇒ à partir d'exemples

Points communs

- ▶ apprendre l'association entre une observation (= description d'un exemple) et une étiquette
- connaissance d'un ensemble d'apprentissage : « résultats passés » = observations avec leur étiquette
- ▶ prédire = déterminer la valeur d'un exemple non observé

Formalisation

Tâche de classification

- $\mathcal{X} \to \text{espace des exemples/observations}$ (en pratique $\mathcal{X} = \mathbb{R}^n$)
- $\blacktriangleright \ \mathcal{Y} \to \mathsf{espace} \ \mathsf{des} \ \mathsf{\overset{\color{red}\mathsf{\'e}\mathsf{tiquettes}}{}}$
- ▶ il y a une dépendance fonctionnelle entre une observation et un étiquette : un expert (vous?) peut déterminer l'étiquette associée à une observation
- déterminer l'étiquette associée à une observation = classer/classifier cette observation
- $ightharpoonup \mathcal{Y} = ext{espace discret}$ (classification) ou $\mathcal{Y} = \mathbb{R}$ (régression)

Formellement, un classifieur est :

$$f: \mathcal{X} \mapsto \mathcal{Y}$$
$$x \to y$$

Exemple

Observations



Étiquette

- 1. spam
- 2. non-spam

attributs : présence d'un mot, nombre de lettres, heure d'envoie, langue, nombre de fautes, ...

Remarque : lorsqu'il y a deux classes, on parle de classification

binaire

Exemple (2)

Observation



Étiquette

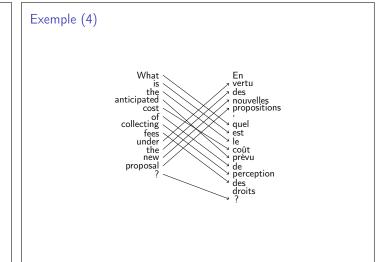
- basket
- ▶ football
- ► hockey sur gazon
- autre
- ⇒ classification multiclasse

 attributs: nombre de pixels blancs, nombre de pixels rouges, nombre de personnes, ...

Exemple (3)

He reckons the current account deficit will narrow to only \$1.8 billion in September.

[NPHe] [VPreckons] [NPthe current account deficit] [VPwill narrow] [PPto] [NPonly \$1.8 billion] [PPin] [NPSeptember.]



Exemple (5)

Comment va ton frère?

τι κανει ο αδελφος σου;

Exemple (5)

Comment va ton frère?

τι κανει ο αδελφος σου;

Remarque : il y a autant de classes que de phrases en français

Formalisation (2)

Apprentissage supervisé

- ▶ la fonction *f* est rarement connue
- \blacktriangleright la fonction f n'est pas forcément exprimable : décision floue \oplus combinaison de critères
- objectif : apprendre/estimer/inférer f à partir d'une base d'exemples et de leur étiquette associée
- ▶ apprentissage supervisé : on connait l'étiquette des exemples

Problématique

► Comment inférer une fonction « générale » à partir d'un nombre fini d'exemples ?

Deuxième partie II

Apprentissage supervisé

Définition



Un programme apprend à partir d'une base d'apprentissage E par rapport à une classe de tâche T et une mesure de performance P, si sa performance à l'exécution de tâches de T, mesurée par P, s'améliore au fur et à mesure que ${\it E}$ lui est présentée.

Éléments

1. un ensemble de données étiquetées :

$$\left(\mathbf{x}^{(i)}, y^{(i)}\right)_{i=1}^{N}$$

- 2. une mesure d'évaluation \Rightarrow capable de mesurer les « progrès »
- 3. une classe d'hypothèses
 - « structure » du programme que l'on veut apprendre
 - exemple : classe des fonctions linéaires

Troisième partie III

Évaluation

Contexte

Objectif de l'apprentissage

- être capable de « (re)trouver » l'étiquette associée à tous les exemples de ${\mathcal X}$
- ightharpoonup approximer la dépendance fonctionnelle f...
- ► ...mais f reste inconnue

Évaluation

- ▶ objectif : est-ce que l'on peut faire « confiance » à un classifieur appris automatiquement / est-ce que l'on a envie de
- ▶ idée de base : comparer les sorties obtenues avec les sorties attendues

Fonction de coût loss function

- évalue le coût associé à une (mauvaise) prédiction
- élément de base : fonction de coût : $\ell(\hat{y}, y^*) : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$

 - \hat{y} est l'étiquette prédite y^* est l'étiquette réelle (donnée par un oracle)
- ▶ définition : $\ell(y,y) = 0$, plus ℓ est grand, plus la décision est mauvaise
- pour la régression :

$$\ell_{\rm RSE} = (y^* - \hat{y})^2$$

▶ pour la classification (binaire) :

$$\ell_{0/1} = \mathbb{1}_{\hat{y}=y^*}$$

$$= \begin{cases} 1 & \text{si } \hat{y} \neq y^* \\ 0 & \text{sinon} \end{cases}$$

Principe de l'évaluation

Un classifieur est évalué par la moyenne de la fonction de coût calculé sur un ensemble des données :

$$\mathcal{E} = \frac{1}{n} \sum_{i} \ell(\hat{y}_i, y_i^*)$$
$$= \frac{1}{n} \sum_{i} \ell(f(x_i), y_i^*)$$

- ▶ plus ce taux d'erreur (error rate) est faible, meilleur est le
- ▶ à mesurer sur un ensemble de test (indépendant des données utilisées en apprentissage)