Exercices de révision - Cours d'Apprentissage Statistique

Aude Genevay - aude.genevay@ens.fr

12 janvier 2017

Exercice 1 - Naive Bayes

On considère un problème de classification avec données d'entrée binaires, $X \in \{0,1\}^d$, et une variable de sortie binaire $Y \in \{0,1\}$.

On considère le modèle génératif suivant :

- Y est distribuée comme une variable de Bernoulli de paramètre $\pi \in [0,1]$,
- sachant $Y = y$ avec $y \in \{0,1\}$, les composantes $X_1, \ldots, X_d$ de X sont indépendantes et chacune distribuée comme une loi de Bernoulli avec paramètres $\mu_{i,y} \in [0,1]$, $i = 1, \ldots, d$.

(a) Ecrire la log-vraisemblance $\log p(x,y)$ pour la paire (X,Y), comme fonction "bi-affine" (avec termes constants, linéaires et bi-linéaires) de $x \in \{0,1\}^d$ et $y \in \{0,1\}$.

(b) Montrer que $\log p(Y=1|x) - \log p(Y=0|x)$ peut s'exprimer comme fonction affine de x (avec paramètres μ et π).

(c) En déduire que $p(Y=1|X=x) = \sigma(a^\top X + b)$ en exprimant a et b en fonction de μ et π, où $\sigma(u) = (1+\exp(-u))^{-1}$.

(d) On considère n observations $(x^1, y^1), \cdots, (x^n, y^n)$ indépendantes et identiquement distribuées de ce modèle. Ecrire la vraisemblance des données $\log p(x^1, y^1, \ldots, x^n, y^n)$ et trouver des estimateurs de μ et π en formules analytiques simples.

(e) Ecrire la vraisemblance conditionnelle $\log p(y^1, \ldots, y^n | x^1, \ldots, x^n)$. A quel modèle vu en cours cela correspond-il ? Décrire un algorithme d'estimation.

(f) Quel est le lien avec l'analyse discriminante linéaire ?

(g) Ce modèle est très utilisé pour la prédiction textuelle, i.e., quand X représente un document, d le nombre total de mots disponibles, et Y une classe quelconque (i.e., article parlant de sport vs. article parlant de finance). Comment utiliser le modèle décrit ?

(h) Dans le cadre d'une application de prédiction textuelle, l'hypothèse d'indépendance de $X_1, \ldots, X_d$ sachant Y est-elle justifiée ? Que se passe-t-il si on la retire ?

Exercice 2 - Seuillage doux et projection sur la boule ℓ_∞

Soit $z \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}_+^*$. On considère le problème d'optimisation suivant :

$$\text{minimiser } \frac{1}{2}\|x - z\|_2^2 + \lambda \|x\|_1 \text{ par rapport à } x \in \mathbb{R}^n,$$

où $\|a\|_p = \left(\sum_{i=1}^n |a_i|^p\right)^{1/p}$ pour $p \in [1, \infty)$, et $\|a\|_\infty = \max\{|a_1|, \cdots, |a_n|\}$.

(a) Montrer que le minimum est *atteint* et qu'il est unique.

(b) Exprimer simplement la solution x en fonction de z. On utilisera la séparabilité par rapport à chacune des composantes de x pour se ramener à n problèmes uni-dimensionnels. Conseil : considérer le cas où $x_i = 0$ est une solution séparément en premier.

(c) Soit $a \in \mathbb{R}^n$. Calculer $\inf_{x \in \mathbb{R}^n} a^\top x + \lambda \|x\|_1$.

(d) En introduisant la variable $u = x - z \in \mathbb{R}^n$ et des multiplicateurs de Lagrange pour ces contraintes, montrer qu'un problème dual est la projection orthogonale sur une boule pour la norme $\|\cdot\|_\infty$.