Algorithmique des Réseaux Sociaux

3 Décembre

Exercice 1 – Distribution a priori pour les lois de Bernoulli et multinomiale

Nous reprenons l'exemple vu en cours : soit X_i une v.a. de Bernoulli avec $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. On note $x = \sum_{i=1}^n x_i$.

1. Calculer l'estimateur du maxium de vraisemblance $\hat{p}_{mle}$.

La distribution $Beta(\alpha, \beta)$ est définie par sa densité sur $[0, 1]$:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}.$$

2. Calculer la moyenne et la variance de la distribution $Beta(\alpha, \beta)$.
3. Calculer la distribution a posteriori pour p si la distributon a priori est $Beta(\alpha, \beta)$.
4. En déduire un estimateur bayésien $\hat{p}$.
5. Comparer les cas $Beta(c, c)$ pour $c = 1$, $c = 1/2$, $c \to \infty$ et $c \to 0$.

On définit la distribution de Dirichlet sur le simplex $\sum_{i=1}^k x_i = 1$, par $(\alpha_i \geq 0)$:

$$f(x_1, \ldots, x_k; \alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} x_1^{\alpha_1} \ldots x_k^{\alpha_k}.$$

6. Calculer le vecteur moyenne d'une distribution de Dirichlet.
7. Calculer la distribution a posteriori pour les paramètres d'une multinomiale avec une distribution a priori suivant une distribution de Dirichlet.

Exercice 2 – Classer avec des fonctions de préférence

Une fonction de préférence sur un ensemble fini X est une fonction $P : X \times X \to [0, 1]$. Une valeur $P(u, v)$ proche de 1 indique que u est préféré à v tandis qu'une valeur proche de 0 indique le contraire. La valeur $P(u, v) = 1/2$ est interprété comme une indifférence entre u et v.

Une fonction d'ordre $f : X \to S \subset \mathbb{R}$ est interprétée comme suit : si $f(u) > f(v)$ alors u est classé avant v. Pour un élément u qui n'est pas classé, nous introduisons $f(u) = \perp$ qui n'est pas comparable à un réel. Une fonction d'ordre induit une fonction de préférence de la manière suivante :

$$P_f(u, v) = \begin{cases} 1 & \text{si } f(u) > f(v), \\ 0 & \text{si } f(u) < f(v), \\ 1/2 & \text{autrement.} \end{cases}$$

Il est utile de représenter une fonction de préférence par un graphe orienté pondéré. Les sommets du graphe correspondent aux éléments de X. Chaque paire (u, v) est connectée par un arc orienté de poids $P(u, v)$.

1. Donner le graphe associé à P_f, P_g et $1/4P_f + 3/4P_g$ avec $f(a) = 1, f(b) = 2, f(c) = 0, f(d) = \perp$ et $g(a) = 0, g(b) = 2, g(c) = 1, g(d) = 2$.

Étant donné une fonction de préférence, le but est de calculer un ordre σ qui maximise :

$$A(\sigma, P) = \sum_{u,v,\, \sigma(u)>\sigma(v)} P(u, v).$$

On considère l'algorithme glouton suivant :

Données : X et une fonction de préférence P
Résultat : un ordre $\hat{\sigma}$
Soit $V = X$;
for *chaque* $v \in V$ **do**
  $\pi(v) = \sum_{u \in V} P(v, u) - \sum_{u \in V} P(u, v)$
end
while V *n'est pas vide* **do**
  soit $t = \arg\max_{u \in V} \pi(u)$;
  soit $\hat{\sigma}(t) = |V|$;
  $V = V - \{t\}$;
  for *chaque* $v \in V$ **do**
    $\pi(v) = \pi(v) + P(t, v) - P(v, t)$
  end
end

2. Donner le résultat de l'algorithme pour l'exemple $1/4P_f + 3/4P_g$.
3. Montrer que $A(\hat{\sigma}, P) \geq \frac{1}{2} \max_{\sigma} A(\sigma, P)$. On pourra utiliser le fait que $\sum_v \pi(v) = 0$.
4. Montrer qu'il existe des graphes tels que le facteur d'approximation soit proche de $1/2$.