Examen du cours Apprentissage Statistique
Master Informatique – Spécialité Intelligence Artificielle et Décision
19 – 11 – 2008

Durée 2 h – Documents autorisés

Mélanges de Bernoulli

L'exercice reprend la notion de mélange de densité dans un cas simple qui est le mélange de lois de Bernoulli. Il dérive les formules de réestimation de l'algorithme EM.

Exemple introductif

Considérons des images binaires où chaque pixel peut prendre la valeur 0 ou 1. On représente l'image par un vecteur de dimension D. Une image donnée peut être considérée comme la réalisation d'une variable aléatoire $\mathbf{x} = (x_1, \ldots, x_D)$ binaire multidimensionnelle. On suppose qu'un pixel x_i prend la valeur 1 avec une probabilité p_i. La loi de $\mathbf{x}$ sera alors caractérisée par le vecteur de paramètres $\mathbf{p} = (p_1, \ldots, p_D)$. On remarque que chaque pixel obéit à une loi différente.
Supposons que l'on dispose d'un ensemble d'images binaires représentant des A et des B manuscrits. On observe un ensemble d'images non étiquetées et on veut modéliser ces observations. On peut modéliser la variable image $\mathbf{x}$ comme un mélange de deux lois correspondant respectivement à la modélisation des A et des B. Il se pose alors le problème de l'identification des paramètres du mélange.

Exercice

Une variable x suit une loi de Bernoulli de paramètre p si $x \in \{0, 1\}$ et $p(x = 1) = p$.

On considère D variables de Bernoulli x_i, i=1..D, de paramètre p_i.
On note $\mathbf{x} = (x_1, \ldots, x_D)$ et $\mathbf{p} = (p_1, \ldots, p_D)$
1. Montrer

$$p(\mathbf{x}|\mathbf{p}) = \prod_{i=1}^{D} p_i^{x_i} (1 - p_i)^{(1-x_i)}$$

2. On considère un mélange de ces distributions

$$p(\mathbf{x}|\ \mathbf{p}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|\ \mathbf{p_k})$$

Avec $\mathbf{p} = (\mathbf{p_1}, \ldots, \mathbf{p_K})$ et $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ et $p(\mathbf{x}|\mathbf{p}_k) = \prod_{i=1}^{D} p_{ki}^{x_i} (1 - p_{ki})^{(1-x_i)}$. $\mathbf{p_i} = (p_{i1}, \ldots, p_{iD})$ est un vecteur de dimension D.
2.1 A quoi correspondent les $\mathbf{x}$, $\mathbf{p_k}$ et π_k dans le cas de l'exemple introductif ?
2.2 Ce modèle de mélange explique statistiquement la génération des données. Expliquer informellement comment les données sont générées selon ce modèle.
2.3 On observe des données $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$, $\mathbf{x_n} = (x_{n1}, \ldots, x_{nD})$ donner l'expression de la log vraisemblance, $\log p(\mathbf{X}|\ \mathbf{p}, \boldsymbol{\pi})$

2.4 Expliquer pourquoi il est difficile d'optimiser directement log p(**X**| **p**, **π**) pour trouver **p** et **π.**

3. On va maintenant dériver un algorithme EM pour optimiser cette log vraisemblance.
On considère des variables cachées **z** associées à chaque instance de **x**. **z** = (z_1, ..., z_K) est un vecteur binaire dont toutes les composantes sont 0 sauf une qui vaut 1. **z** est l'indicatrice de la composante.

3.1 Montrer

$$p(\mathbf{x}|\ \mathbf{z},\mathbf{p}) = \prod_{k=1}^{K} p(\mathbf{x}|\mathbf{p_k})^{z_k} \ \text{ et } \ p(\mathbf{z}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

3.2 Montrer

$$p(\mathbf{x}|\mathbf{p},\boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k\, p(\mathbf{x}|\mathbf{p_k})$$

Pour cela, on calculera $\sum_{Z} p(\mathbf{x},\mathbf{z}|\mathbf{p},\boldsymbol{\pi})$ où la somme porte sur toutes les valeurs possibles de **Z**.

3.3 On note **X** = {$\mathbf{x_1},\dots,\mathbf{x_N}$} N observations de la variable **x** et **Z** = {$\mathbf{z_1},\dots,\mathbf{z_N}$} les variables cachées correspondantes avec $\mathbf{z_n}$ = ($z_{n1}, \dots, z_{nD}$). Montrer que la log vraisemblance des données complètes, log p(**X, Z**|**p, π)** s'écrit :

$$\sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\log \pi_k + \sum_{i=1}^{D} (x_{ni} \log p_{ki} + (1-x_{ni}) \log(1-p_{ki})))$$

3.4 On note $\mathbf{E}_{\mathbf{Z}|\mathbf{X},\mathbf{p},\boldsymbol{\pi}}$ [.] l'espérance par rapport à la distribution a posteriori des variables latentes **Z**, sachant **X, p et π.**
Montrer que l'espérance par rapport à la distribution a posteriori de **Z** de la vraisemblance complète s'écrit :

$$\mathbf{E}_{\mathbf{Z}|\mathbf{X},\mathbf{p},\boldsymbol{\pi}}\left[\log p(\mathbf{X},\mathbf{Z}|\mathbf{p},\boldsymbol{\pi})\right] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} (\log \pi_k + \sum_{i=1}^{D} (x_{ni} \log p_{ki} + (1-x_{ni}) \log(1-p_{ki})))$$

Où $\gamma_{nk} = \mathbf{E}_{\mathbf{Z}|\mathbf{X},\mathbf{p},\boldsymbol{\pi}} [z_{nk}] = p(z_k=1|\ \mathbf{x_n})$
Par la suite, on note Q cette espérance.

3.5 Montrer (utiliser la règle de Bayes).

$$\gamma_{nk} = \frac{\pi_k\, p(\mathbf{x_n}|\mathbf{p_k})}{\sum_{j=1}^{K} \pi_j\, p(\mathbf{x_n}|\mathbf{p_j})}$$ [1]

4 On veut maximiser l'espérance Q sous la contrainte $\sum_{k=1}^{K} \pi_k = 1$

4.1 Ecrire le lagrangien associé à ce problème.
4.2 Dériver la formule permettant de calculer $\mathbf{p}_k$
4.3 Dériver la formule permettant de calculer $\boldsymbol{\pi}_k$
5. Formuler l'algorithme EM permettant la maximisation de Q

[1] Dans cette expression, les π_k, π_j, p_k, p_j sont les paramètres courants du modèle notés $W^{(t)}$ dans le cours. Lors de l'étape M, ils sont considérés comme des constantes et pas des variables.