TD 6

Exercice 1 – Algorithme des K-moyennes

La taille du dictionnaire K est fixée, c'est un paramètre de l'algorithme.

- Initialiser aléatoiremet les K prototypes.
- Répéter jusqu'à (critère d'arrêt) :
  - ▶ partitionner les exemples en les affectant aux prtotypes dont ils sont le plus proche ;
  - ▶ redéfinir les protoypes (i.e. centres de gravité des partitions).

Q 1.1 Soit l'ensemble d'exemples en dimension 2 :
$D = \{(0,-4),(0,-3),(1,-3),(1,-2),(0,4),(-1,1),(-1,2),(0,3)\}$
Faire tourner l'algorithme des K-moyennes en prenant comme point de départ les prototypes $(0,-6)$ et $(-1,1)$.

Q 1.2 Quels critères d'arrêt préconisez-vous pour les méthodes de QV ?

Exercice 2 – Clustering et mélange de lois

On souhaite estimer une densité de probabilité par un modèle de type mélange de gaussiennes. La probabilité d'une observation x est donnée par : $p(x) = \sum_{l=1}^{L} \tau_l . p(x|\lambda_l)$ où les τ_l sont les probabilités a priori des lois et les $p(x|\lambda_l)$ sont des lois gaussiennes multi-dimentionnelles caractérisées par leur moyenne μ_l et leur matrice de co-variance Σ_l , i.e. $\lambda_l = (\mu_l, \Sigma_l)$.

Q 2.1 Dessiner la loi de probabilité pour $L = 2, \tau_1 = \tau_2 = 0.5$, et $\mu_1 = 1, \mu_2 = 3, \Sigma_1 = 1, \Sigma_2 = 10$.

Q 2.2 Quelles est la probabilité a posteriori qu'un exemple x ait été produit par la gaussienne multi-dimentionnelle l, $p(\lambda_l|x)$?

Q 2.3 Expliquer comment l'apprentissage d'un mélange de lois peut être utilisé pour faire du clustering.

Exercice 3 – Apprentisage d'un mélange de lois et maximum de vraissemblance

On souhaite apprendre le modèle de l'exercice précedent avec un critère de maximum de vraisemblance (MV) sur une base d'apprentissage $X = \{x_i\}, i = 1..N$.

Q 3.1 Exprimer la log-vraisemblance des données par le modèle en supposant que les x_i sont indépendants.

Q 3.2 Quelle est la difficulté avec cette log-vraissemblance ?

Q 3.3 L'idée de l'algorithme EM (Expectation-Maximization) est de se dire que si l'on avait des informations supplémentaires Z, on pourrait optimiser cette vraissemblance plus facilement. Quelles informations seraient utiles ici ? Donner la vraissemblance complétée par ces informations.

Q 3.4 On peut alors utiliser un algorithme dit algorithme EM (Expectation-Maximization) pour l'estimation de ce mélange de gaussiennes. Une variante de l'algorithme EM est la suivante :

- initialiser les paramètres $(\tau_i, \mu_i, \sigma_i)_{i=1..L}$;
- Répéter :
  - ▶ déterminer pour chaque x_i la gaussienne qui l'a produit avec la plus grande vraissemblance : pour $i = 1..N$, $I(x_i) = argmax_l \; p(\lambda_l|x_i)$;

- ► ré-estimer les paramètres des lois à partir des exemples qui lui ont été affectés : pour $l = 1..L$, ré-estimer λ_l à partir des $\{x_i \in E | I(x_i) = l\}$

Dans le cas où tous les τ_i sont égaux (equi-probabilité des gaussiennes) et où les matrices de covariance des lois sont fixées à l'identité, montrer que l'algorithme précédent est équivalent à un algorithme des K-Moyennes.

Q 3.5 L'algorithme précédent procède par affectations successives des éléments aux différents clusters. Quelle est la limite de ce genre d'approche ?

Q 3.6 La version classique de l'algorithme EM travaille en deux étapes :

- Expectation step (E step) : Calcul de l'espérance de la log-vraissemblance en fonction des probabilités conditionnelles des données manquantes Z étant donné les observations X selon les estimations courantes des paramètres $\theta^{(t)}$:

$$Q(\theta|\theta^{(t)}) = \mathrm{E}_{Z|X,\theta^{(t)}} [\log L(\theta; X; Z)]$$

- Maximization step (M step) : Recherche des paramètres θ qui maximisent cette quantité :

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)})$$

Q 3.7 Sachant que d'après l'inégalité de Gibbs, $\sum_{i=1}^{n} p_i \log p_i \geq \sum_{i=1}^{n} p_i \log q_i$ pour toutes paires de distributions de probabilités p et q, montrer que la suite des vraissemblances $p(X|\theta^{(t)})$, selon les paramètres $\theta^{(t)}$ calculés à chaque étape de l'algorithme EM, est croissante.

Q 3.8 Donner la formulation de l'espérance $Q(\theta|\theta^{(t)})$ selon les paramètres courants $\theta^{(t)}$.

Q 3.9 Donner alors les formulations des estimations des paramètres à l'itération t+1 selon les estimations à l'itération t