

Mastere M2 MVA 2009/2010 - Modèles graphiques

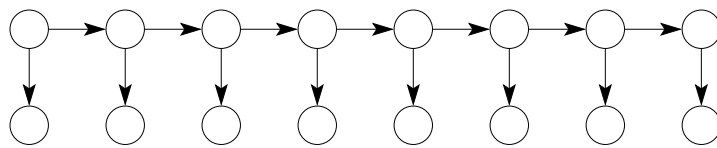
Exercices à rendre pour le 18 novembre 2009.

Ces exercices peuvent s'effectuer par groupe de deux élèves.

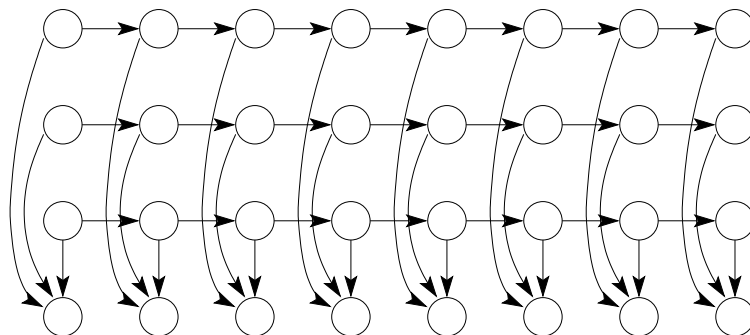
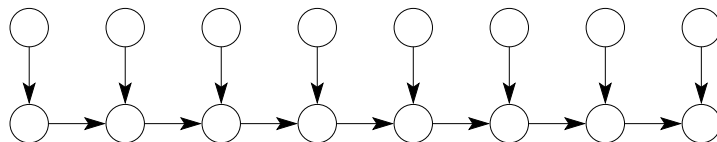
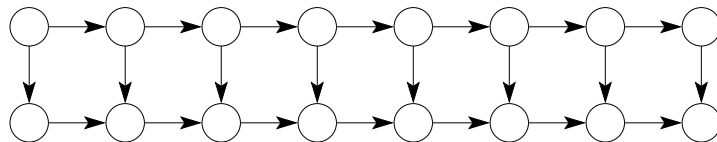
1 Elimination pour modèles dynamiques

Pour chacun des 4 graphes suivants,

- Déterminer la largeur arborescente (treewidth) après moralisation. Preuve non exigée.
- Exhiber un ordre d'élimination et le graphe reconstitué correspondants



(exemple de modèle de Markov caché)



(exemple de modèle de Markov caché *factoriel*)

2 Implémentation - mélange de Gaussiennes

Le fichier “EMGaussienne.dat” contient un ensemble de données (x_n, y_n) où $(x_n, y_n) \in \mathbb{R}^2$. Le but de cet exercice est d’implémenter l’algorithme EM pour certains mélanges de K Gaussiennes dans \mathbb{R}^d (dans cet exercice, $d = 2$ et $K = 4$), avec des données IID. (NB : dans cet exercice, il n’est pas nécessaire de démontrer les formules utilisées).

Le langage de programmation est libre (MATLAB, Octave, Scilab ou R sont néanmoins recommandés). Le code source doit être remis avec les résultats. Le corrigé sera codé en MATLAB.

- (a) Implémenter l’algorithme K-means (chapitre 10 du polycopié). Représenter graphiquement les données d’apprentissage, les centres obtenus et les différents groupes (“clusters”). Essayer plusieurs initialisations et comparer les résultats (centres et mesures de distortions).
- (b) Implémenter l’algorithme EM pour un mélange de Gaussiennes avec des matrices de covariance proportionnelles à l’identité (initialiser l’algorithme EM avec les moyennes trouvées par K-means).

Représenter graphiquement les données d’apprentissage, les centres et les covariances obtenus (une manière élégante de représenter graphiquement est de représenter l’ellipse contenant un certain pourcentage (e.g., 90%) de la masse de la Gaussienne). Estimer et représenter la variable latente pour chaque point (pour le jeu de paramètres appris par EM).

- (c) Implémenter l’algorithme EM pour un mélange de Gaussiennes avec des matrices de covariance générales. Représenter graphiquement les données d’apprentissage, les centres et les covariances obtenus. Estimer et représenter la variable latente pour chaque point (pour le jeu de paramètres appris par EM).
- (d) Commenter les différents résultats obtenus. En particulier, comparer les log-vraisemblances des deux modèles de mélanges, sur les données d’apprentissage, ainsi que sur les données de test (dans “EMGaussienne.test”).

3 BONUS : Equivalence factorisation/indépendances conditionnelles pour les modèles graphiques orientés

Soit $G = (V, E)$ un DAG et X un ensemble de variables aléatoires discrètes indexées par V et de loi jointe $p(x)$. Montrer par récurrence sur $\#(V)$ la proposition suivante :

Proposition 1 *Si pour tout $v \in V$, $X_v \perp X_{\text{non-descendants}(v)} | X_{\text{parents}(v)}$, alors $p(x)$ se factorise dans $L(G)$.*