Internet Analytics (COM-308)
Problem Set 6 - Solutions

Problem 1

In this exercise we verify the two formulas which we used (in the lecture notes and lab) for updating μ_k and Σ_k in the M–step of EM for GMM.

We know that the log–likelihood function for a GMM is

$$L = \log p(X_1, \ldots, X_n | \pi, \mu, \Sigma) = \sum_n \log \sum_k \pi_k N(X_n | \mu_k, \Sigma_k).$$

(a) Show that $\mu_k = \frac{1}{N_k} \sum_n \gamma_{nk} x_n$ with $\gamma_{nk} = \frac{\pi_k N(X_n|\mu_k, \Sigma_k)}{\Sigma_j \pi_j N(X_n|\mu_j, \Sigma_j)}$ and $N_k = \sum_n \gamma_{nk}$ is the solution to $\nabla_{\mu_k} L = 0$.

As we have seen in class, setting the gradient of L with respect to μ_k to zero we obtain

$$\nabla_{\mu_k} L = 0 \rightarrow \sum_n \frac{\pi_k N(X_n|\mu_k, \Sigma_k)}{\Sigma_j \pi_j N(X_n|\mu_j, \Sigma_j)} \Sigma_k^{-1}(X_n - \mu_k) = 0.$$

Multiply this equation with Σ_k to obtain

$$N_k \mu_k = \sum_n \gamma_{nk} X_n$$

(b) Show that $\Sigma_k = \frac{1}{N_k} \sum_n \gamma_{nk}(X_n - \mu_k)(X_n - \mu_k)^\intercal$ is the solution of $\nabla_{\Sigma_k} L = 0$.

Hint: Use $\frac{\partial \log|\det(X)|}{\partial X} = (X^{-1})^\intercal = (X^\intercal)^{-1}$ and $\frac{\partial \mathrm{Tr}(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})^\intercal$.

Setting the derivatives of L with respect to Σ_k to zero we obtain

$$\frac{\partial L}{\partial \Sigma_k} = 0 \rightarrow \sum_n \frac{\pi_k N(X_n|\mu_k, \Sigma_k)}{\Sigma_j \pi_j N(X_n|\mu_j, \Sigma_j)} \left(\Sigma_k^{-1} - \Sigma_k^{-1}(X_n - \mu_k)(X_n - \mu_k)^\intercal \Sigma_k^{-1}\right) = 0.$$

Multiply the equation by Σ_k from left and right to obtain

$$\sum_n \frac{\pi_k N(X_n|\mu_k, \Sigma_k)}{\Sigma_j \pi_j N(X_n|\mu_j, \Sigma_j)} \left(\Sigma_k - (X_n - \mu_k)(X_n - \mu_k)^\intercal\right) = 0.$$

Therefore

$$N_k \Sigma_k = \gamma_{nk}(X_n - \mu_k)(X_n - \mu_k)^\intercal.$$

(c) Consider what happens when we fit a GMM to a set of data points in such a way that one cluster (mixture component) "focuses" on only one of the points. What happens to Σ_k for this cluster, and how does the likelihood evolve?

This may happen with an unlucky initialization: the cluster zeroeing in on a single point has $\det \Sigma_k \to 0$ (with iterations of the EM algorithm), and the likelihood $L \to \infty$.

Problem 2

(a) Recall that the modularity of a graph is given by

$$Q = \frac{1}{2m} \sum_{i=1}^{|C|} \sum_{u,v \in c_i} \left(\mathbf{1}_{u,v} - \frac{d_u d_v}{2m} \right),$$

where $\mathbf{1}_{u,v} = 1$ if there is an edge between nodes u and v, d_u is the degree of node u, m is the number of edges in the graph, and c_i is the set of nodes in community i.

Show that the above definition of modularity can also be expressed more compactly via the following useful formula:

$$Q = \sum_{i=1}^{|C|} \left[\frac{a_i}{m} - \left(\frac{b_i}{2m} \right)^2 \right].$$

Explain what the values of a_i and b_i are.

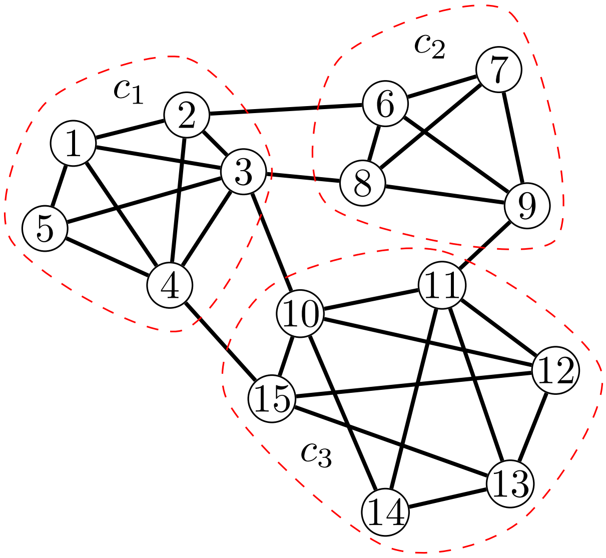
Hint: Recall that $(\sum_i x_i)^2 = \sum_i \sum_j x_i x_j$.

We can recover the second equation from the first one by plugging the constant and inner-most sum inside the brackets. Namely,

$$\begin{aligned}
Q &= \frac{1}{2m} \sum_{i=1}^{|C|} \sum_{u,v \in c_i} \left(\mathbf{1}_{u,v} - \frac{d_u d_v}{2m} \right) \\
&= \sum_{i=1}^{|C|} \sum_{u,v \in c_i} \left(\frac{\mathbf{1}_{u,v}}{2m} - \frac{d_u d_v}{(2m)^2} \right) \\
&= \sum_{i=1}^{|C|} \left(\frac{\sum_{u,v \in c_i} \mathbf{1}_{u,v}}{2m} - \frac{\sum_{u,v \in c_i} d_u d_v}{(2m)^2} \right) \\
&= \sum_{i=1}^{|C|} \left(\frac{\frac{1}{2} \sum_{u,v \in c_i} \mathbf{1}_{u,v}}{m} - \left(\frac{\sum_{u \in c_i} d_u}{2m} \right)^2 \right).
\end{aligned}$$

So $a_i = \frac{1}{2} \sum_{u,v \in c_i} \mathbf{1}_{u,v}$ is the number of edges in community i, and $b_i = \sum_{u \in c_i} d_u$ is the sum of degrees of nodes in community i.

(b) Find the partitioning which maximizes the modularity of the graph G in Figure 1. What is the maximum modularity value Q?



Graph G

Three partitions c_1, c_2 and c_3 which maximize the modularity are shown in Figure 1. The corresponding modularity is

$$Q = 0.4917.$$

(c) We want to increase the value of modularity by removing one edge from graph G. Guess the edge whose deletion results in the largest increase, and compute the new Q.

Removing an inter–partition edge between partitions c_1 and c_3, for example edge $(3, 10)$, results in the highest increase of modularity. The modularity after this edge deletion is

$$Q = 0.5217.$$