CS281A/Stat241A Homework Assignment 4 (due October 29, 2015)

1. **(K-means)** Show that the K-means algorithm,

$$z_i^j := \mathbf{1}\left[j = \arg\min_k \|x_i - \mu_k\|\right]$$

$$\mu_j := \frac{\sum_i z_i^j x_i}{\sum_i z_i^j}$$

performs coordinate descent in the cost function

$$J(z, \mu) = \sum_{i=1}^{n} \sum_{j=1}^{K} z_i^j \|x_i - \mu_j\|^2,$$

where the x_i and μ_j are vectors in $\mathbb{R}^d$ and each vector $z_i = (z_i^1, \ldots, z_i^K)$ is constrained to have one component equal to 1 and the others zero.

2. **(IPF)** An independent polling agency is interested in determining the dependencies in election voting patterns between d local community members. For each proposition on the ballot for several years, the agency collects a response vector $x \in \{0,1\}^d$ with x_i meaning individual i says they will vote "yes" for that proposition. We can model the random vector $(X_1, X_2, ..., X_d)$ with an undirected graphical model of the form:

$$p(x_1, ..., x_d) = \frac{1}{Z} \prod_{(i,j) \in E} \psi_{i,j}(x_i, x_j),$$

where $x_1, ..., x_d$ are binary variables, and $\psi_{i,j}$ are non negative functions. (Note that we allow arbitrary clique potentials, and that the cliques are not necessarily maximal.)

   (a) For a set of $d = 7$ community members, the data in file **hw4data.data** contains a 7×500 matrix, with $n = 500$ different propositions voted on. Implement the IPF algorithm, and use your implementation on this data to estimate the model parameters for each of the following 3 graphs:

      (i) The model shown in Figure 1a.

      (ii) The model shown in Figure 1b.

      (iii) The fully connected graph.

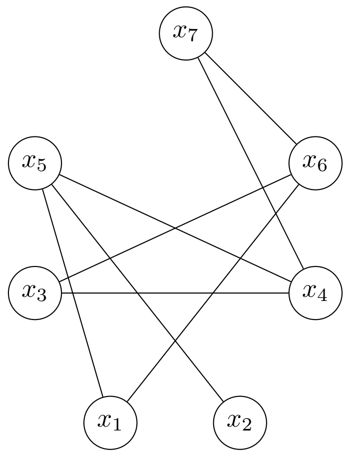
   In your solution, please report the values for $\psi_{3,4}(x_3, x_4)$ that are estimated by IPF for each of the three models.

   (b) What is the likelihood of each of the three models? Which is the highest? Do you think it is the "best" model fit for the data? Why or why not?

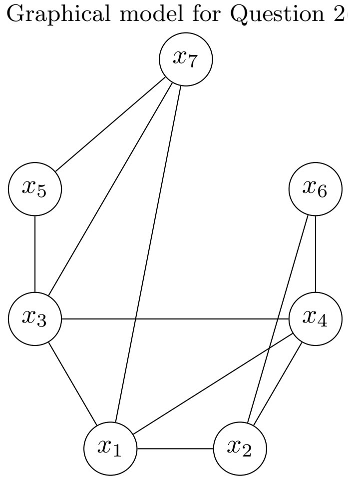
3. **(Max Likelihood Tree)** Consider an undirected tree $T = (V, E)$ and the corresponding model

$$p(x_V) = \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{(i,j) \in E} \psi_{i,j}(x_i, x_j),$$

where we assume that each x_i takes values in a finite set, and we allow arbitrary non-negative potentials ψ_i and $\psi_{i,j}$. Suppose we see a sample of size N with empirical distribution $\tilde{p}$, and we wish to choose the tree T and its potentials so as to maximize the likelihood. Every tree is chordal, and we have seen that, for a chordal undirected graphical model, we can construct a junction tree and then directly write down maximum likelihood potentials. Recall that an edge between cliques C_1, C_2 in a junction tree is labelled with the *separator set* $C_1 \cap C_2$.



(a) Graphical model for Question 2(a)i



(b) Graphical model for Question 2(a)ii

Figure 1: Social Networks for Question 2

(a) Consider a fixed tree T and a junction tree (E, S) for T. (Note that we include only the set E of maximal cliques in the junction tree.) Show that every node in T of degree d appears as the separator set of $d-1$ edges in the junction tree for T.

(b) Hence show that choosing

$$\psi_i(x_i) = \tilde{p}(x_i),$$
$$\psi_{i,j}(x_i, x_j) = \frac{\tilde{p}(x_i, x_j)}{\tilde{p}(x_i)\tilde{p}(x_j)},$$
$$Z = 1$$

maximizes the likelihood.

(c) Let $\tilde{p}_i$ and $\tilde{p}_{i,j}$ denote the marginal empirical distributions of x_i and (x_i, x_j) respectively, and define the marginal entropies

$$H(\tilde{p}_i) = -\sum_{x_i} \tilde{p}(x_i) \log \tilde{p}(x_i) \qquad \forall i \in V,$$
$$H(\tilde{p}_{i,j}) = -\sum_{x_i, x_j} \tilde{p}(x_i, x_j) \log \tilde{p}(x_i, x_j) \qquad \forall (i,j) \in E.$$

Show that the value of the log likelihood for the maximum likelihood estimates for a tree depends only on the marginal entropies.

(d) Consider the tree structure selection problem: we would like to select from the set of all possible trees on the vertices V a tree that maximizes the likelihood $l(\widehat{\theta}(T))$. Show that this problem is equivalent to finding a maximum weight spanning tree for the complete graph on V, where the weight of an edge between i and j is the empirical mutual information $I(i,j)$, defined as

$$I(i,j) = D(\tilde{p}_{i,j}; \tilde{p}_i \tilde{p}_j) = \sum_{x_i, x_j \in \mathcal{X}} \tilde{p}(x_i, x_j) \log \frac{\tilde{p}(x_i, x_j)}{\tilde{p}(x_i)\tilde{p}(x_j)}.$$

(e) Use the method from part 3d to choose the tree that maximizes the likelihood for the data in hw4data.data. (The set of trees on seven vertices is small enough to enumerate. Alternatively, you might use one of the many polynomial-time algorithms for this problem.)