CS281 Practice Midterm

Fall 2013

1. **Fitting Via KL Divergence [?? Points]**

   Let $p(k)$ be a one-dimensional discrete distribution that we wish to approximate, with support on nonnegative integers. One way to fit an approximating distribution $q(k)$ is to minimize the Kullback-Leibler divergence:

$$\mathrm{KL}(p\,||\,q) = \sum_{k=0}^{\infty} p(k)\, \ln \frac{p(k)}{q(k)}$$

   Show that when $q(k)$ is a Poisson distribution, this KL divergence is minimized by setting λ to the mean of $p(k)$.

2. **Combining Gaussians [?? Points]** Let $\boldsymbol{r}$ and $\boldsymbol{s}$ both be K-dimensional Gaussian random variates with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$. Show that

$$\boldsymbol{u} = (\boldsymbol{r} - \boldsymbol{\mu}) \sin\theta + (\boldsymbol{s} - \boldsymbol{\mu}) \cos\theta + \boldsymbol{\mu}$$

is marginally Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$ for any θ.

3. **Linear Gaussian Models [?? Points]** Suppose we have the following model:

$$x_1 \sim \mathcal{N}(0, \sigma^2)$$
$$x_n \sim \mathcal{N}(x_{n-1}, \sigma^2)$$

for $n = 2, \ldots, N$.

(a) Write the joint distribution $p(\boldsymbol{x}) = p([x_1, \ldots, x_N])$ as a multivariate Gaussian parameterized by its mean and *inverse* covariance matrix.

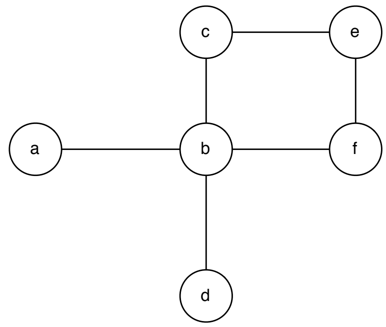
(b) Now let us reason about the covariance structure. Write down a recurrence equation for the variance of x_n, and then solve this equation to derive an analytical expression for $\mathrm{var}[x_n]$.

(c) Suppose instead that we had the model

$$x_n \sim \mathcal{N}(a x_{n-1}, \sigma^2)$$

for $a \in \mathbb{R}$. What conditions on a would guarantee $\lim_{n\to\infty} \mathrm{var}[x_n] < \infty$?

4. **Undirected Graphical Models [?? Points]** Use the following graphical model answer the following questions:



(a) Circle the maximal cliques.

(b) What is the tree width of this graph?

(c) Suppose we observe the value of node e, does the treewidth of the graph change? What is it?

(d) Write down a factorization of the joint probability distribution over a, b, c, d, e in terms of potentials ψ that is consistent with this graph.

(e) Draw a factor graph representation of this distribution that is consistent with the choices you made in part (d).

5. **Expectation-Maximization [?? Points]** Suppose we have a box of six-sided dice, which we know to consist of K types of dice. Each die type has a weight vector $\bar{w}_k$ associated with it. Let $\bar{c} = (c_1, \ldots, c_M), \sum_{i=1}^{6} c_i = R$ be the count vector associated with rolling a die R times, i.e., c_1 is the number of times that the die came up with a 1 when rolled R times, etc. So we have that

$$\bar{c} \mid k, R \sim \text{Multinomial}(\bar{w}_k, R)$$

(a) Write down the pmf for observing a count vector $\bar{c}$ given R rolls and the knowledge that the observations came from a die of type k.

$p(\bar{c} \mid k, R, \bar{w}_k) =$

Suppose we know that the fraction of the dice that are of type $k \in 1, \ldots, K$ is π_k. Further we receive observation count vectors $\bar{c}^{(1)}, \ldots, \bar{c}^{(N)}$, each of which resulted from choose a die randomly with replacement from the box and rolling it R times. Let z_{nk} be 1 if the nth die is of type k. You will derive the EM updates for the parameters $\bar{w}_1, \ldots, \bar{w}_k$ and π for this model.

(b) Write down the likelihood observing a the set of count vectors $\bar{c}^{(1)}, \ldots, \bar{c}^{(N)}$ specified above.

$p(\bar{c}^{(1)}, \ldots, \bar{c}^{(N)} | \bar{w}_1, \ldots, \bar{w}_k, \bar{\pi}) =$

(c) Write down the complete data log likelihood this model, where the complete data is both the count vector observations and the $\{z_{nk}\}$.

(c) Write down the expected complete data log likelihood for this model.

(d) Derive the EM updates for this model by maximizing the expected complete data log likelihood with respect to $\bar{w}_1, \ldots, \bar{w}_k$ and π.