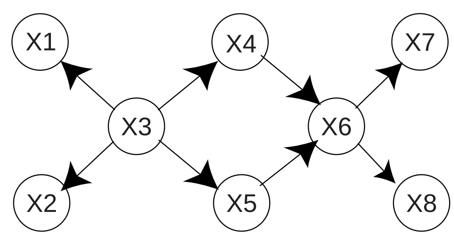
**COMS 6998-01 Advanced Machine Learning**
Assignment 1
February 5, 2002
Prof. Tony Jebara

The assignment is due on February 19th, before 5pm either in my office CEPSR 605 or via email to jebara@cs.columbia.edu. If you email me the assignment, please use convential formats, i.e. send me plain text, latex, postscript, pdf or word and keep the file size reasonable. The assignment will be evaluated not just on your ability to get the right result but also your ability to provide reasoning, derivations and discussion for your answer.

1. **Graphical Models and Conditional Independence**

   Write out a factorized version of the probability distribution $p(x_1, \ldots, x_8)$ which captures the conditional independency properties implied by the graph below.




   If the $x_i$ variables are binary, i.e. they can assume the values $\{0, 1\}$, how many entries would the probability table representing the factorized distribution contain in its most compact form? How does this compare with the number of table elements in the unfactored distribution $p(x_1, \ldots, x_8)$ with no independence properties. What happens to the sizes of the factorized and unfactorized forms if the variables $x_i$ are not binary but, rather, can assume one of three discrete states, eg. $x_i \in \{0, 1, 2\}$?

2. **ML Estimation for Exponential Distribution**

   Recall the exponential distribution over positive scalar values of $x$:

$$p(x|\lambda) = \lambda \, e^{-\lambda x}$$

   Given a data set of $n$ IID (independent identically distributed) samples: $x_1, \ldots, x_n$, derive the maximum likelihood (ML) estimate for $\lambda$.

3. **ML Estimation for Poisson Distribution**

   Recall the Poisson distribution over postive integer values of $x$:

$$p(x|\lambda) = \frac{\lambda^x \, e^{-\lambda}}{x!}$$

   Given a data set of $n$ IID (independent identically distributed) samples: $x_1, \ldots, x_n$, derive the maximum likelihood (ML) estimate for $\lambda$

4. **Jensen's Inequality**

Recall Jensen's inequality for concave functions $f(x)$ which states $f(E\{X\}) \geq E\{f(X)\}$ or, more explicitly:

$$f\left(\sum_i p_i x_i\right) \geq \sum_i p_i f(x_i)$$

where $p_i \geq 0$ and $\sum_i p_i = 1$. The inequality flips when $f(x)$ is convex.

The Kullback-Leibler divergence measures the 'distance' between two distributions and is defined as:

$$KL(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$$

Where $p_i \geq 0$ and $\sum_i p_i = 1$ and $q_i \geq 0$ and $\sum_i q_i = 1$. Use Jensen to prove that the KL-divergence is never negative.

5. **Exponential Family Form for Gaussian Covariance**

The exponential family has the following natural form:

$$p(X|\theta) = \exp(H(X) + \theta^T T(X) - A(\theta))$$

Where $T(X)$ is a vector function of the input $X$. Consider a zero-mean Gaussian with the covariance $\sigma^2$ as the free parameter:

$$p(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$$

Derive the exponential family natural form where a scalar $\theta$ parameterizes the exponential family model instead of the scalar covariance parameter $\sigma^2$ in the traditional Gaussian above.

6. **BONUS: Exponential Family Form for Multivariate Gaussian**

The multivariate Gaussian distribution over vectors $x$ in $\Re^D$ is traditionally expressed in the following form:

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma|}}\, e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

As a bonus question, derive the exponential family natural form of the multivariate Gaussian where both the mean vector $\mu$ and the covariance matrix $\Sigma$ are variable and are mapped into a $\theta$ natural parameter. HINT: the exponential family natural $\theta$ parameter and the $T(X)$ features can each be of size $D + D^2$.