CS5350: Machine Learning *Due 4 Dec 2008*

HW9: Expectation Maximization

1 Written Exercises

1. We already saw the Poisson distribution in HW8. Recall that it is a distribution over positive count values; for a count k with parameter λ, the Poisson has the form $p(k \mid \lambda) = \frac{1}{e^\lambda}\frac{\lambda^k}{k!}$. We saw that the maximum likelihood estimat for λ given a sequence of counts $k_1, \dots, k_N$ was simply $\frac{1}{N}\sum_n k_n$ – the mean of the counts.

   Let's consider an generalization of this: the Poisson mixture model. Believe it or not, this is actually used in web server monitoring. The number of accesses to a web server in a minute typically follows a Poisson distribution.

   Suppose we have N web servers we are monitoring and we monitor each for M minutes. Thus, we have $N \times M$ counts; call $k_{n,m}$ the number of hits to web server n in minute m. Our goal is to *cluster* the web servers according to their hit frequency.

   Construct a Poisson mixture model for this problem and compute the expectations (pie charts) and maximization steps for this model.

   Hints. Suppose we want L clusters; let z_n be the latent variable telling us which cluster web server n belongs to (from one to L). Let λ_l denote the parameter for the Poisson for cluster l. Then, the complete data likelihood should look pretty close to the Gaussian case, but with a product of Poissons, rather than a Gaussian. This looks something like:

$$p(\boldsymbol{k}, \boldsymbol{z} \mid \boldsymbol{\lambda}) = \prod_n \prod_l \left[\prod_m \mathcal{P}oi(k_{n,m} \mid \lambda_k) \right]^{\mathbf{1}[z_n = l]} \tag{1}$$

   Here, $\mathbf{1}[z_n = l]$ is one if $z_n = l$ and zero otherwise.

   Next, go from the complete data likelihood to the incomplete data likelihood by summing over the unknowns $z_1, \dots, z_N$. Just like the Gaussian case, we can move the this sum from outside the $\prod_n$ to inside by observing that these are all mutually independent.

   Write down the incomplete data likelihood.

   Now, take the log of this, and apply Jensen's inequality. Produce the optimal choice for the mixing coefficients (the "qs" from the notes) and the maximization step. *That is, what do the "pie chart" probabilities look like and what does the "update the λs" step look like?*