Introduction to Machine Learning **Fall Semester, 2013/14**

Homework 2: Oct 27th, 2013

Due: Nov 11th (See the submission guidelines in the course web site)

Theory Questions

1. Consider a collection of points, $x_1, \cdots x_m$, sampled i.i.d. according to an Exponential Distribution. (Recall that an exponential distribution has a parameter λ and the density of x is $\lambda e^{-\lambda x}$.)

   (a) What is the Maximum Likelihood (ML) estimate of λ.

   (b) What is the Maximum A Posteriori (MAP) value of λ given that its prior distribution is exponential with parameter 1.

   (c) Compute the Bayesian posterior distribution for an exponential distribution where λ is distributed according to a Gamma distribution, i.e.,

$$\text{Gamma}(\lambda\,;\,\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\,\lambda^{\alpha-1}\,\exp(-\lambda\,\beta).$$

   (Comment: when maximizing, remember to check also the second derivative.)

2. Show that the decision rule of a binary naive bayes classifier for binary attributes $\mathbf{x} = (x_1, \ldots x_d)$ is of the form "classify $\mathbf{x}$ as class 1 if $\mathbf{w}^t\mathbf{x} + b > 0$, otherwise classify as class 2" for some $\mathbf{w}$ and b. State explicitly $\mathbf{w}$ and b as a function of $\theta_i^1, \theta_i^0, p_1$ and p_0 for $i \in \{1 \ldots d\}$, where $\theta_i^1 = p(x_i = 1|class = 1)$, $\theta_i^0 = p(x_i = 1|class = 0)$, $p_1 = p(\text{class}= 1)$, and $p_0 = p(\text{class}= 0)$.

3. Consider the "three-coins" problem discussed in the recitation: A coin with λ probability of "H" is tossed first. If the result of the first toss is "H" a coin with p_1 probability of "H" is tossed. Otherwise, if the result of the fist toss is "T" a coin with p_2 probability of "H" is tossed. Only the result of the second toss is observed. This experiment is repeated independently several times.

   Show that if the observations are H, H, T, T and the initial guess is $p_1 = 0.5, p_2 = 0.5, \lambda = 0.3$ then the EM algorithm is "stuck" and would not improve the initial values.

Programming Assignment

Every line g_i of the accompanying "hw2.data" file was generated independently of the other lines as follows: First, a 5 bit binary string $s_r \in \{s_1 \ldots s_{32}\}$ was generated with probability p_r

(for example $s_{18} =$ "10010"), and then two random bits of s_r were masked (that is, replaced with "2", for example the result of masking s_{18} may be one of "22010", "20210", "20020", etc'. Each with probability $\binom{5}{2}^{-1} = 0.1$).

1. Derive the *EM* algorithm's iterative update rules (Specifically, define the E-step and M-step) for maximizing the likelyhood
   $L(\{p_r\}_{r=1}^{32} | \{g_i\}_{i=1}^{n})$
2. Using Matlab, implement the *EM* algorithm and indicate the resulting $\{p_r\}_{r=1}^{32}$ and the number of iterations used. Submit your implementation according to the guidelines in the course web site.