CSE 569 Homework #2

Total 3 points. Due Thursday, Oct. 10, 11:59pm.

Problem 1. Let p be the probability of "head" in a coin-tossing experiment. We repeat the experiment independently N times and let X record the number of "head" observations. Then X is a random variable that follows the binomial distribution with parameters N and p. That is, we have,

$$P(X = i|N, p) = \binom{N}{i} p^i (1-p)^{N-i}$$

where $\binom{N}{i}$ denotes "combination" (the number of ways to choose i elements out of N elements). Note: this is a discrete probability distribution.

Let N be given and fixed, and p be the only unknow parameter. Given n i.i.d. samples $\{X_1, X_2, \ldots, X_n\}$ from the above distribution, find the MLE for p.

Problem 2. Let X have an exponential density

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x}, & if\ x \geq 0 \\ 0, & otherwise \end{cases}$$

(1) Plot $p(x|\theta)$ versus x for $\theta = 1$ (i.e., $p(x|\theta)$ is viewed as a function of x).

(2) Plot $p(x|\theta)$ versus θ, $0 \leq \theta \leq 5$, for $x = 2$ ((i.e., $p(x|\theta)$ is viewed as a function of θ).

(3) Given a training set of n samples $D = \{x_1, x_2, \ldots, x_n\}$(i.i.d. samples drawn from the above distribution), find the MLE for θ.

(The plots in (1) and (2) can be approximately sketched by hand or drawn by a computer program.)

Problem 3. (from the textbook)

One measure of the difference between two distributions in the same space is the *Kullback-Leibler divergence* of Kullback-Leibler "distance":

$$D_{KL}(p_1(\mathbf{x}), p_2(\mathbf{x})) = \int p_1(\mathbf{x}) \ln \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x}.$$

(This "distance" does not obey the requisite symmetry and triangle inequalities for a metric.) Suppose we seek to approximate an arbitrary distribution $p_1(\mathbf{x})$ by a normal $p_2(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Show that the values that lead to the smallest Kullback-Leibler divergence are the obvious ones:

$$\boldsymbol{\mu} = \mathcal{E}_1[\mathbf{x}]$$
$$\boldsymbol{\Sigma} = \mathcal{E}_1[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t],$$

where the expectation taken is over the density $p_1(\mathbf{x})$.

Problem 4. Slide 29 of Notes #04 covers an example from the textbook, where we obtained the following general result: $p(\theta|D^n) \propto 1/\theta^n$ for $\max_x[D^n] \leq \theta \leq 10$.

Using all 4 samples from this example (i.e., given D^4=D={4, 7, 2, 8}) and the given prior for θ (uniform on (0, 10]), find the distribution $p(x|D^4)$ by the Bayesian estimation approach.

(Note: the second figure of the example already plots $p(x|D^4)$ for you, but you need to figure out the precise expression for $p(x|D^4)$.)

Problem 5. We plan to use the PCA algorithm on a data set of 2-dimensional feature vectors. Then we find out that all the feature vectors happen to fall onto the same straight line in the 2-d space. What will be the first principal vector $\mathbf{e}_1$ and the second largest eigen value λ_2 from the PCA algorithm?

$9 + 15 + 3$

$p(S^1 = S_1)$

Problem 6. Consider the simple HMM for a 3-state weather model as illustrated on Slide 62 of Notes #04. We further assume that there are only 5 different observation values in this model.

a_{ij} b_{jk} π_i

$p(S^1 = S_2)$

$p(S^1 = S_3)$

(1) How many parameters do we have for this HMM? List them explicitly.

(2) We consider HMM in the context of "parametric approaches" since the parameters in (1) fully define the model and if the parameters are known, we can draw sample observation sequences from the model, just like in the case of a parametric density model, where we could draw samples from the density if its parameters are known. Outline a procedure for drawing observation sequences from an HMM with given model parameters.

① $\pi_i = p(S^1 = S_i) = V^0$

$t = 0$

② For $t \leftarrow t+1$:

$V^t = a_{ij} \, b_{jk} \, V^{t-1}$

③ V^t