CS731 Homework 2

Due 4/4/2011 before class

What to hand in: the .pdf file, the .tex source file, and any other files required in the questions. You do not need to handin any code. See course webpage for hand-in instructions. This homework involves programming – you are free to choose the language, but Matlab or R is recommended. Please use the class mailing list to ask questions (and answer them if you can).

1. (10 points) Suppose we run a ridge regression with regularization parameter λ on a single variable X, and get coefficient β. We now include an exact copy $X^* = X$, and refit our ridge regression. Show that both coefficients are identical, and derive their value.

2. (30 points) Consider the 1D Gaussian distribution $p(x) = N(x; \mu, \sigma^2)$ with mean μ and variance σ^2. Define sufficient statistics $\phi(x) = (x, x^2)^\top$. Show key derivation steps when answering the following questions.

   (a) Express $p(x)$ in exponential form with parameters $\theta = (\theta_1, \theta_2)^\top$. In particular, derive $\theta_1, \theta_2, A(\theta)$ in terms of μ, σ.

   (b) Define Ω.

   (c) Given parameter θ, derive the corresponding mean parameter (suggestion: call the mean parameter $m = (m_1, m_2)^\top$ to avoid confusion with the mean of the Gaussian).

   (d) Derive the set of realizable mean parameters $\mathcal{M}$.

   (e) Derive the conjugate dual function A^*.

3. (30 points) You will implement a Metropolis-Hastings sampler in this question. The target distribution $p(\theta)$ is a Mixture-of-Dirichlet:

$$p(\theta) = \frac{1}{2} Dir(\theta; \alpha_1 \dots \alpha_d) + \frac{1}{2} Dir(\theta; \beta_1 \dots \beta_d) \qquad (1)$$

   where d is the dimensionality, and the α's and β's are positive Dirichlet parameters.

   (a) Derive $\mathbb{E}_p[\theta]$ in the general case.

   (b) Let $d = 3$. Describe how you generate a particular set of α's and β's in $(0, +\infty)$. These are going to specify your target distribution $p(\theta)$. Show your α's and β's, as well as the value of $\mathbb{E}_p[\theta]$.

(c) Your Metropolis-Hastings sampler can evaluate $p(\theta)$ (or an unnormalized version of it) for any θ. However, do not give the proposal distribution any knowledge of your α's and β's (i.e., it should not know where the true modes are, etc.). Instead, use a d-dimensional Gaussian distribution centered on the previous sample:

$$q(\theta' \mid \theta) = N(\theta'; \theta, \frac{1}{10} I) \tag{2}$$

where I is the d-dimensional identity matrix.

Explain the mismatch between the domains of the Mixture-of-Dirichlet distribution, and the Gaussian proposal distribution. Explain how you handle the mismatch.

(d) Generate 5,000 samples with your Metropolis-Hastings sampler, discard the first 1,000 for burn-in. Plot the remaining 4,000 samples as a 2D scatter plot: Each $\theta = (\theta_1, \theta_2, \theta_3)$ can be plotted as a 2D point (θ_1, θ_2).

(e) Estimate $\mathbb{E}_p[\theta]$ with your 4,000 samples.

(f) Repeat the question but with $d = 10$. Specifically, show the α's and β's, the true $\mathbb{E}_p[\theta]$, and your estimated $\mathbb{E}_p[\theta]$ from 4,000 samples. You do *not* need to visualize them.

4. (30 point)

MendotaIce.txt lists $x \in [0, 1]$ the fraction of time (per year) that Lake Mendota was covered by ice. Each row is for a single year (not necessarily in the correct order). This data comes from the Wisconsin State Climatology Office and dates back to 1855. The article DETERMINING THE ICE COVER ON MADISON LAKES at
`http://www.aos.wisc.edu/~sco/lakes/msn-lakes_instruc.html`
serves as a fine example of the Wisconsin tradition to integrate scientific research and beer.

Let us estimate the density distribution $p(x)$ with a kernel density estimator. Use the kernel $K = N(0, 1)$. Do not worry about probability mass "leaking" outside [0,1].

(a) Plot the cross-validation estimator of risk $\hat{J}(h)$ on a dense grid of h. Give the value of the optimal h you use.

(b) Plot the density using your optimal h.

(c) Plot the densities using $h/10$ (under-smoothing) and $10h$ (over-smoothing), respectively.