UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544NA PATTERN RECOGNITION
Fall 2007

Exam 1

Monday, October 14, 2007

- This is a CLOSED BOOK exam, but you may use ONE PAGE, BOTH SIDES of hand-written notes
- Calculators are permitted
- You must SHOW YOUR WORK to get full credit.

Problem	Score
1	
2	
3	
4	
Total	

Name: SOLUTION

Problem 1 (40 points)

The planet xkblprq and its core rotate asynchronously, in a complex multiperiodic pattern, with the result that some mornings, the sun doesn't rise. You have the following model: on the ith day, the sun rises ($x_i = 1$) with probability θ, or fails to rise ($x_i = 0$) with probability $1-\theta$. In other words,

$$p_x(x_i) = \theta^{x_i}(1-\theta)^{(1-x_i)}$$

(a) What is $\hat{\theta}_{ML}$, the maximum likelihood estimate of θ given i.i.d. training data $\mathcal{D} = \{x_1, \ldots, x_n\}$?

$$\hat{\theta} = \arg\max \sum_{i=1}^{n} \left[x_i \ln\theta + (1-x_i)\ln(1-\theta)\right]$$

$$= \arg\max \left[\left(\sum_{i=1}^{n} x_i\right)\ln\theta + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-\theta)\right]$$

$$\frac{\partial \ln p(\mathcal{D}|\theta)}{\partial\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} = 0$$

$$\sum_{i=1}^{n} x_i\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \frac{n}{1-\theta}$$

$$\sum_{i=1}^{n} x_i\left(\frac{1-\theta}{\theta} + 1\right) = n \Rightarrow \boxed{\frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\theta}_{ML}}$$

(b) Prove that $s = \sum_{i=1}^{n} x_i$ is a sufficient statistic for θ.

Prove that $p(\theta|s,\mathcal{D}) = p(\theta|s)$ or $p(\mathcal{D}|s,\theta) = p(\mathcal{D}|s)$

or $p(\mathcal{D}|\theta) = g(s,\theta)\,h(\mathcal{D})$

$$p(\mathcal{D}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{(1-x_i)} = \theta^{s}(1-\theta)^{n-s}$$

$$= g(s,\theta)\,h(\mathcal{D})$$

$$g(s,\theta) = \theta^{s}(1-\theta)^{n-s}$$

$$h(\mathcal{D}) = 1$$

(c) You decide to regularize your estimate with a Dirichlet prior:

$$p(\theta) = \begin{cases} \frac{(m+1)!}{k!(m-k)!}\theta^k(1-\theta)^{(m-k)} & 0 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where "!" denotes factorial. Find $\hat{\theta}_{MAP}$.

$$\hat{\theta}_{MAP} = \arg\max \left[\ln p(\mathcal{D}|\theta) + \ln p(\theta) \right]$$

$$= \arg\max \left[s \ln\theta + (n-s)\ln(1-\theta) + k\ln\theta + (m-k)\ln(1-\theta) + \ln\left(\frac{(m+1)!}{k!(m-k)!}\right) \right]$$

$$= \arg\max \left[(s+k)\ln\theta + (n+m-s-k)\ln(1-\theta) + \ln\left(\frac{(m+1)!}{k!(m-k)!}\right) \right]$$

$$= \frac{s+k}{n+m} \quad \text{by analogy to } \boxed{a}$$

(d) Find a function which is proportional to the Bayesian estimate $\hat{p}(x_0|\mathcal{D})$, using the Dirichlet prior. In order to evaluate the integral, you may find it useful to note that the Dirichlet prior is a correctly normalized probability density.

$$\hat{p}(x_0|\mathcal{D}) = \int p(x_0|\theta)\, p(\theta|\mathcal{D})\, d\theta$$

$$\propto \int p(x_0|\theta)\, p(\theta, \mathcal{D})\, d\theta$$

$$= \int \theta^{x_0}(1-\theta)^{(1-x_0)}\, \theta^s (1-\theta)^{n-s} \frac{(m+1)!}{k!(m-k)!} \theta^k (1-\theta)^{m-k}\, d\theta$$

$$= \left(\frac{(m+1)!}{k!(m-k)!}\right) \int \theta^{(s+k+x_0)} (1-\theta)^{(m+n+1-s-k-x_0)}\, d\theta$$

$$= \frac{\left(\frac{(m+1)!}{k!(m-k)!}\right)}{\left(\frac{(m+n+2)!}{(s+k+x_0)!(m+n+1-s-k-x_0)!}\right)}$$

Problem 2 **(20 points)**

You have landed on a planet on which it is possible, with prior probability P_1, that the sun rises every day, but it is also possible, with prior probability $1 - P_1$, that the sun only rises 50% of the time.

(a) The sun has risen for the past three days. Your commander asks you to determine, with minimum probability of error, whether or not this is a planet on which the sun always rises. For what values of P_1 would you answer "yes"?

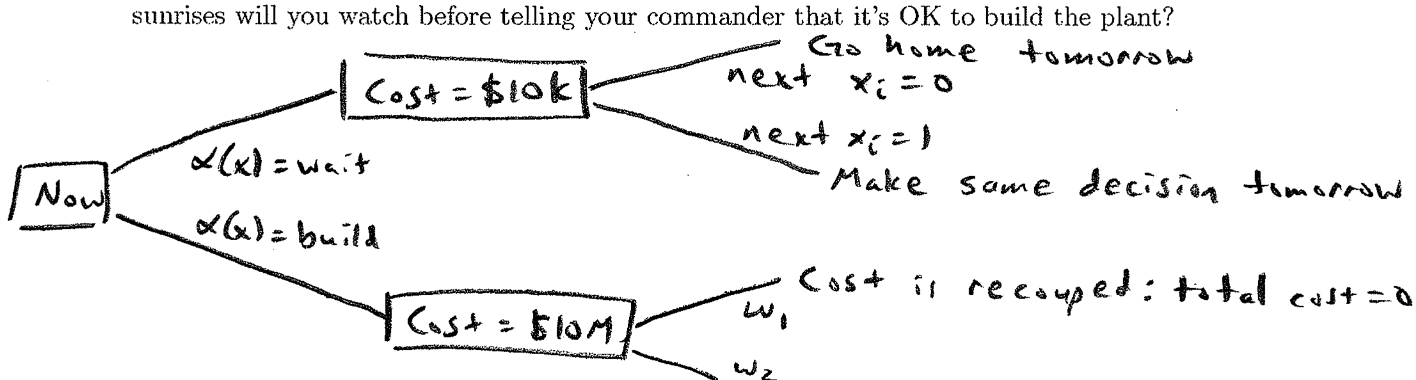
$$P(\omega_1, \mathcal{D}) = P_1 \prod_{i=1}^{3} p(x_i = 1 | \omega_1) = P_1$$

$$P(\omega_2, \mathcal{D}) = (1-P_1) \prod_{i=1}^{3} p(x_i = 1 | \omega_1) = (1-P_1)\left(\tfrac{1}{2}\right)^3$$

$$P(\omega_1, \mathcal{D}) > P(\omega_2, \mathcal{D}) \quad \text{if} \quad P_1 > (1-P_1)\left(\tfrac{1}{8}\right)$$

$$\Leftrightarrow \frac{9}{8} P_1 > \frac{1}{8} \quad \Leftrightarrow \boxed{P_1 > \frac{1}{9}}$$

(b) Your spaceship is expensive: each day that you sit on this planet costs \$10,000. Your goal is to determine whether or not the sun always rises, and if so, to build a \$10,000,000 solar power plant. If the sun ever fails to rise on any given day, chemicals in the solar cells will destroy the power plant, wasting \$10,000,000. As a function of P_1, how many consecutive sunrises will you watch before telling your commander that it's OK to build the plant?



$E[\lambda(\alpha = \text{wait})] = \$10\text{k}$, independent of truth.

$$E[\lambda(\alpha = \text{build})] = 10{,}000{,}000 \; P(\omega_2 | \mathcal{D}) = 10{,}000{,}000 \frac{(1-P_1)\left(\frac{1}{2}\right)^n}{P_1 + (1-P_1)\left(\frac{1}{2}\right)^n}$$

$$E[\lambda(\alpha = \text{wait})] < E[\lambda(\alpha = \text{build})] \text{ iff } P_1 + (1-P_1)\left(\tfrac{1}{2}\right)^n < 1000(1-P_1)\left(\tfrac{1}{2}\right)^n$$

$$\Rightarrow \boxed{n > \ln\left(P_1 / 999(1-P_1)\right) / \ln\left(\tfrac{1}{2}\right)}$$

Problem 3 (20 points)

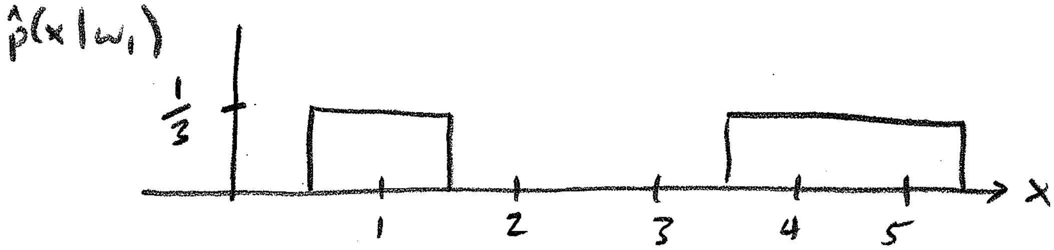
You are a microcontroller with a sensor that measures the light level, x_i, in arbitrary units. Your job is to determine the weather: is today ω_1 ="sunny", or ω_2 ="cloudy?" You have been programmed to compute Parzen window estimates of $p(x|\omega_1)$ and $p(x|\omega_2)$ using the rectangular window:

$$\hat{p}(x) = \frac{1}{nV}\sum_{i=1}^{n}\phi\left(\frac{x-x_i}{h}\right)$$

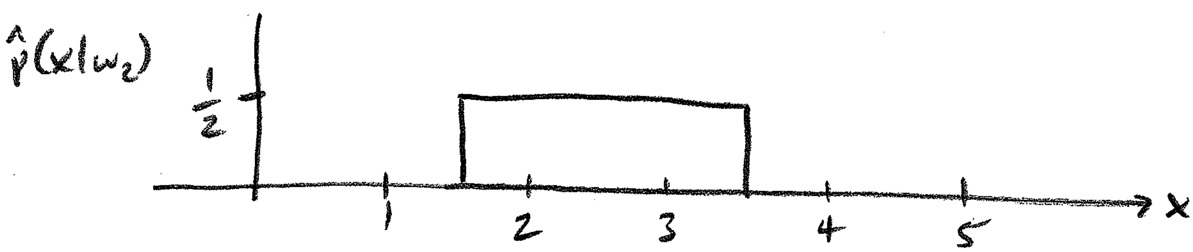
$$\phi\left(\frac{x}{h}\right) = \begin{cases} 1 & |x| < \frac{h}{2} \\ 0 & \text{otherwise} \end{cases}$$

You have five labeled training days. Three days were sunny, with light levels of $x_1 = 4$, $x_2 = 1$, and $x_3 = 5$ units, respectively. Two days were cloudy, with light levels of $x_4 = 3$ and $x_5 = 2$ units.

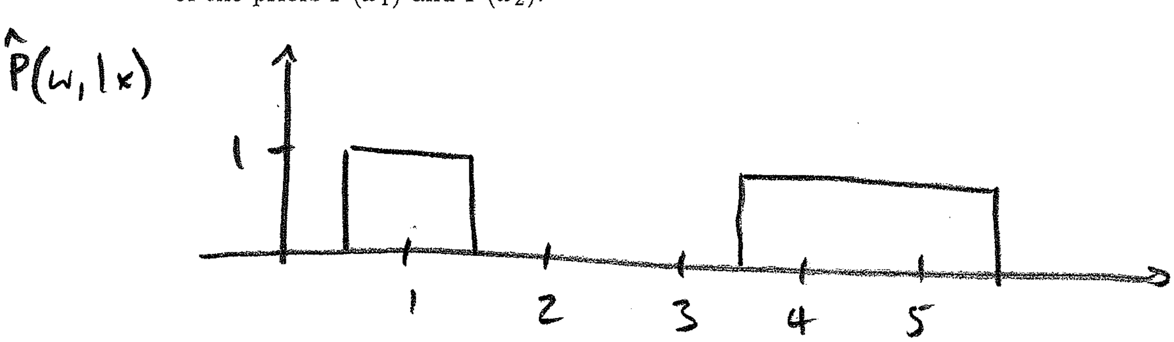
(a) Plot the Parzen window estimated likelihood $\hat{p}(x|\omega_1)$ as a function of x, using $h = 1$.



(b) Plot the Parzen window estimated likelihood $\hat{p}(x|\omega_2)$ as a function of x, using $h = 1$.



(c) Plot the posterior probability, $\hat{P}(\omega_1|x)$, implied by your choices in parts (a) and (b), for all values of x for which it is defined, using $h = 1$. Assume maximum likelihood estimates of the priors $\hat{P}(\omega_1)$ and $\hat{P}(\omega_2)$.



(d) In this part, you will choose a new value of h in order to control the bias of the estimator. Suppose that you don't know anything about the true likelihood, $p(x|\omega_1)$, except that it is everywhere continuous with bounded slope:

$$\left|\frac{\partial p(x|\omega_1)}{\partial x}\right| < a$$

As in the previous sections, assume that

$$\phi\left(\frac{x}{h}\right) = \begin{cases} 1 & |x| < \frac{h}{2} \\ 0 & \text{otherwise} \end{cases}$$

Choose h, as a function of a, so that

$$|p(x|\omega_1) - E_1[\hat{p}(x|\omega_1)]| < 0.01$$

where

$$E_1[f(x)] \equiv \int f(x)p(x|\omega_1)dx$$

$$E_1[\hat{p}(x|\omega_1)] = \int_{-h/2}^{h/2} p(x-y|\omega_1)\frac{1}{V}\phi\left(\frac{y}{h}\right)dy$$

$$= \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} p(x-y|\omega_1)\,dy$$

$$\leq p(x|\omega_1) + \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} a|y|\,dy$$

$$= p(x|\omega_1) + \frac{2a}{h}\left(\frac{y^2}{2}\right)\Big|_0^{h/2} = p(x|\omega_1) + \frac{ah}{4}$$

$$\left|p(x|\omega_1) - E_1[\hat{p}(x|\omega_1)]\right| < \frac{ah}{4} < 0.01$$

$$\text{iff } \boxed{h < \frac{0.04}{a}}$$

Problem 4 (20 points)

The dissimilarity between Gaussian distributions $p(x|\omega_1)$ and $p(x|\omega_2)$, with means μ_1 and μ_2 and with covariance matrices Σ_1 and Σ_2 respectively, may be measured as

$$d(\omega_1, \omega_2) = \ln\left(\text{trace}\left(\Sigma_1^{-1}\Sigma_2\right) E_1\left[(x-\mu_1)^T\Sigma_2^{-1}(x-\mu_1)\right] / d^2\right)$$

where

$$E_1[f(x)] \equiv \int f(x)p(x|\omega_1)dx$$

(a) Is $d(\omega_1, \omega_2)$ symmetric? (note: trace(AB)=trace(BA) for commensurate matrices.)

$$d(\omega_1,\omega_2) = \ln\left|\frac{1}{d^2}\text{trace}(\Sigma_1^{-1}\Sigma_2)\, E_1\left[\text{trace}(\Sigma_2^{-1}(x-\mu_1)(x-\mu_1)^T)\right]\right|$$

$$= \ln\left|\frac{1}{d}\text{trace}(\Sigma_1^{-1}\Sigma_2)\right| + \ln\left|\frac{1}{d}\text{trace}\left(\Sigma_2^{-1}E_1[(x-\mu_1)(x-\mu_1)^T]\right)\right|$$

$$= \ln\left|\frac{1}{d}\text{trace}(\Sigma_1^{-1}\Sigma_2)\right| + \ln\left|\frac{1}{d}\text{trace}(\Sigma_2^{-1}\Sigma_1)\right|$$

which is symmetric

(b) Is $d(\omega_1, \omega_2)$ reflexive?

- if $\Sigma_1 = \Sigma_2$, $d(\omega_1,\omega_2)=0$
- if $\Sigma_1 \neq \Sigma_2$, it is still possible that $d(\omega_1,\omega_2)=0$, for example, if $\Sigma_2 = aI$, $\Sigma_1 = \frac{1}{a}I$, then

$$d(\omega_1,\omega_2) = \ln\left|\frac{1}{d}\cdot d\left(\frac{a}{a}\right)\right| + \ln\left|\frac{1}{d}d\left(\frac{a}{a}\right)\right| = 0$$

$\therefore$ Not reflexive