Fall 2016, EECE 5644, Homework #2

Prof: Jennifer Dy, Deniz Erdogmus **TA: Sadegh Salehi, Hongfu Liu**

For your convenience, these problems from the textbook are provided in the following pages. Some sub-problems have been removed. Use the provided problem statements.

2.1 (10 pts) Problem 2.12 from Duda's book.

2.2 (10 pts) Problem 3.1 from Duda's book. Skip (c) (removed from the problem statements below)

2.3 (10 pts) Problem 3.2 from Duda's book. Skip (b) (removed from the problem statements below)

2.4 (20 pts) Problem 3.3 from Duda's book.

2.5 (10 pts) Problem 3.4 from Duda's book.

2.6 (40 pts) Problem 3.17 from Duda's book.

12. Let $\omega_{max}(\mathbf{x})$ be the state of nature for which $P(\omega_{max}|\mathbf{x}) \geq P(\omega_i|\mathbf{x})$ for all i, $i = 1, \ldots, c$.

(a) Show that $P(\omega_{max}|\mathbf{x}) \geq 1/c$.

(b) Show that for the minimum-error-rate decision rule the average probability of error is given by

$$P(error) = 1 - \int P(\omega_{max}|\mathbf{x})p(\mathbf{x})\, d\mathbf{x}.$$

(c) Use these two results to show that $P(error) \leq (c-1)/c$.

(d) Describe a situation for which $P(error) = (c-1)/c$.

1. Let x have an exponential density

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

(a) Plot $p(x|\theta)$ versus x for $\theta = 1$. Plot $p(x|\theta)$ versus θ, $(0 \leq \theta \leq 5)$, for $x = 2$.

(b) Suppose that n samples $x_1, \ldots, x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum-likelihood estimate for θ is given by

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}.$$

2. Let x have a uniform density

$$p(x|\theta) \sim U(0, \theta) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & \text{otherwise.} \end{cases}$$

(a) Suppose that n samples $\mathcal{D} = \{x_1, \ldots, x_n\}$ are drawn independently according to $p(x|\theta)$. Show that the maximum-likelihood estimate for θ is $\max[\mathcal{D}]$—that is, the value of the maximum element in $\mathcal{D}$.

3. Maximum-likelihood methods apply to estimates of prior probabilities as well. Let samples be drawn by successive, independent selections of a state of nature ω_i with unknown probability $P(\omega_i)$. Let $z_{ik} = 1$ if the state of nature for the kth sample is ω_i and $z_{ik} = 0$ otherwise.

(**a**) Show that

$$P(z_{i1}, \ldots, z_{in} | P(\omega_i)) = \prod_{k=1}^{n} P(\omega_i)^{z_{ik}} (1 - P(\omega_i))^{1-z_{ik}}.$$

(**b**) Show that the maximum-likelihood estimate for $P(\omega_i)$ is

$$\hat{P}(\omega_i) = \frac{1}{n} \sum_{k=1}^{n} z_{ik}.$$

Interpret your result in words.

4. Let $\mathbf{x}$ be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

$$P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i} (1 - \theta_i)^{1-x_i},$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^t$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum-likelihood estimate for $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k.$$

17. The purpose of this problem is to derive the Bayesian classifier for the d-dimensional multivariate Bernoulli case. As usual, work with each class separately, interpreting $P(\mathbf{x}|\mathcal{D})$ to mean $P(\mathbf{x}|\mathcal{D}_i, \omega_i)$. Let the conditional probability for a given category be given by

$$P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

and let $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of n samples independently drawn according to this probability density.

(a) If $\mathbf{s} = (s_1, \ldots, s_d)^t$ is the sum of the n samples, show that

$$P(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

(b) Assuming a uniform prior distribution for $\boldsymbol{\theta}$ and using the identity

$$\int_0^1 \theta^m(1-\theta)^n \, d\theta = \frac{m!n!}{(m+n+1)!},$$

show that

$$p(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

(c) Plot this density for the case $d = 1, n = 1$ and for the two resulting possibilities for s_1.

(d) Integrate the product $P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})$ over $\boldsymbol{\theta}$ to obtain the desired conditional probability

$$P(\mathbf{x}|\mathcal{D}) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}.$$

(e) If we think of obtaining $P(\mathbf{x}|\mathcal{D})$ by substituting an estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in $P(\mathbf{x}|\boldsymbol{\theta})$, what is the effective Bayesian estimate for $\boldsymbol{\theta}$?