EECE 5698
Homework 3:   Due Wednesday, Feb. 5, 2014

Reading:

- Notes, Chapter 2.8, 3.1-3.5.

----

1. Given that **B** is a square symmetric matrix and **x** is a column vector. Show that

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = 2\,\mathbf{B}\,\mathbf{x}$$

Book Problems:

Chapter 2 problem #37. Chapter 3 problems #1 (a), (b), 3, 17, 19.

37. Consider a two-category classification problem in two dimensions with

$$p(\mathbf{x}|\omega_1) \sim N(\mathbf{0}, \mathbf{I}),\ p(\mathbf{x}|\omega_2) \sim N\left(\begin{pmatrix}1\\1\end{pmatrix}, \mathbf{I}\right),\ \text{and } P(\omega_1) = P(\omega_2) = 1/2.$$

(a) Calculate the Bayes decision boundary.

(b) Calculate the Bhattacharyya error bound.

(c) Repeat the above for the same prior probabilities, but

$$p(\mathbf{x}|\omega_1) \sim N\left(\mathbf{0}, \begin{pmatrix}2 & .5\\ .5 & 2\end{pmatrix}\right) \text{ and } p(\mathbf{x}|\omega_2) \sim N\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}5 & 4\\ 4 & 5\end{pmatrix}\right).$$

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, $(0 \leq \theta \leq 5)$, $x = 2$.

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

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

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.

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 $\boldsymbol{\theta}$?

19. Assume we have training data from a Gaussian distribution of known covariance $\boldsymbol{\Sigma}$ but unknown mean $\boldsymbol{\mu}$. Suppose further that this mean itself is random, and characterized by a Gaussian density having mean $\mathbf{m}_0$ and covariance $\boldsymbol{\Sigma}_0$.

(a) What is the MAP estimator for $\boldsymbol{\mu}$?

(b) Suppose we transform our coordinates by a linear transform $\mathbf{x}' = \mathbf{A}\mathbf{x}$, for non-singular matrix $\mathbf{A}$, and accordingly for other terms. Determine whether your MAP estimator gives the appropriate estimate for the transformed mean $\boldsymbol{\mu}'$. Explain.