Mid-term
ECE 271A
Electrical and Computer Engineering
University of California San Diego

Nuno Vasconcelos Fall 2013

1. The Gamma distribution is defined as

$$P_X(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k\Gamma(k)}, \qquad x > 0 \qquad k,\theta > 0. \tag{1}$$

where k and θ are parameters and $\Gamma(.)$ is the gamma function

$$\Gamma(k) = (k-1)!.$$

Note that k is an integer, θ a real, and k, θ, and x are all positive. This distribution has a number of interesting properties

1. if $X \sim Gamma(x;k,\theta)$

$$\begin{array}{rcl} E_X[X] & = & k\theta \\ var_X[X] & = & k\theta^2. \end{array}$$

2. if $X \sim Gamma(x;k,\theta)$ then, for any $c > 0$,

$$cX \sim Gamma\,(x;k,c\theta)$$

3. if $\{X_1,\ldots,X_N\}$ is a collection of independent random variables $X_i \sim Gamma(x;k_i,\theta)$ then

$$\sum_{i=1}^{N} X_i \sim Gamma\left(x;\sum_{i=1}^{N} k_i,\theta\right).$$

a) (15 points) Let $\mathcal{D} = \{x_1,\ldots,x_n\}$ be an iid sample from a Gamma random variable of parameters k,θ. Assume that k is known. Derive the ML estimator for θ.

b) (15 points) Derive expressions for the bias, variance, and mean squared error of the estimator of **a**).

c) (10 points) Let $\mathcal{D} = \{x_1,\ldots,x_n\}$ be an iid sample from a Gamma random variable of parameters k,θ. Assume that neither k nor θ are known. Describe a procedure (e.g. give an algorithm) to compute the ML estimates of k and θ.

2. Joe needs to classify a vector $\mathbf{x}$ of observations into two classes. The classes have labels $Y \in \{-1, 1\}$, i.e. an example $\mathbf{x}$ from the negative class has label $y = -1$. The two classes are Gaussian, i.e.

$$P_{\mathbf{X}|Y}(\mathbf{x}|i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)\right\}, \qquad i \in \{-1, 1\}$$

where d is the dimension of $\mathbf{x}$, μ_i the mean of class i, and Σ_i the covariance of class i. The two classes have equal probability, $P_Y(-1) = P_Y(1)$, and identity covariance, $\Sigma_i = \mathbf{I}$. Joe also collected a sample of n iid $(\mathbf{x}_i, y_i)$ pairs

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}.$$

Here[1], $y_i \in \{-1, 1\}$ is the class label for observation $\mathbf{x}_i$. In the following, whenever we mention the BDR, it is assumed that we use the "0-1" loss.

a) (10 points) Show that the decision boundary for the BDR or this problem has the form

$$\mathbf{w}^T \mathbf{x} + b = 0 \tag{2}$$

and derive expressions for $\mathbf{w}$ and b in terms of the parameters of the Gaussian distributions above.

b) (10 points) in practice, this rule cannot be implemented because some of the parameters are unknown. Joe used the maximum likelihood principle to estimate the parameters, and then plugged them in (2). Derive the expressions that he obtained for $\mathbf{w}$ and b as a function of $\mathbf{x}_i, y_i$ and n.

c) (20 points) Bob has been taking statistical learning classes on-line and learned about a classifier named the support vector machine (SVM). This is a two-class classifier with decision boundary

$$\sum_{k=1}^{K} \alpha_k \mathbf{v}_k^t \mathbf{x} + b = 0 \tag{3}$$

where α_k and b are scalars and $\mathbf{v}_k$ are vectors. Bob was told that this is the latest in learning technology, but is confused. How can Joe be learning something different at UCSD? Joe told him that this is not a problem, since the two solutions can be the same if $K = n$, $\mathbf{v}_k = \mathbf{x}_k$ for $k = 1, \ldots, K$. What are the values of α_k and b for which this is true?

d) (20 points) Bob liked this and went back to check if these conditions were met for the sample $\mathcal{D}$. However, this did not hold. Bob became even more confused. Joe told him that this is not surprising, since the two procedures are different estimators $\hat{\mathbf{w}}$ and $\hat{b}$ of $\mathbf{w}$ and b. Joe then wondered about the asymptotic behavior of his classifier and considered the expectation

$$\mu_{\mathbf{w}} \quad = \quad E_{\mathbf{X}_1, \ldots, \mathbf{X}_n, Y_1, \ldots, Y_n}[\hat{\mathbf{w}}]$$

for his method (based on the ML estimates). Compute this expectation.

[1] Note that the pairs are independent, but y_i is obviously not independent of $\mathbf{x}_i$.