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

Nuno Vasconcelos Fall 2017

1. The exponential probability density function is defined for $x \geq 0$ and has the form

$$\mathcal{F}(x, \lambda) = \lambda e^{-\lambda x},$$

where $\lambda > 0$.

a) (10 points) Consider a classification problem with two exponential classes

$$P_{X|Y}(x|i) = \mathcal{F}(x, \lambda_i), \;\; i \in \{0, 1\}$$

and class probabilities $P_Y(i) = \pi_i$. A sample of independent measurements $\mathcal{D} = \{x_1, \ldots, x_n\}$ has been collected. It is known that they have all been drawn from the same class, and the goal is to determine that class. Show that the optimal decision function, under the "0/1" loss, for this problem is a threshold on the sample mean

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

What is this threshold? *(Note that the goal is to classify the entire sample, not one x_i at a time).*

b) (10 points) Consider that we have an independent sample $\mathcal{D} = \{x_1, \ldots, x_n\}$ from an exponential random variable X of probability density $\mathcal{F}(x, \lambda)$. What is the maximum likelihood estimate for the parameter λ?

2. Consider a classification problem with three Gaussian classes

$$P_{\mathbf{X}|Y}(\mathbf{x}|i) = \mathcal{G}(\mathbf{x}, \mu_i, \mathbf{\Sigma}), \; i \in \{1, 2, 3\},$$

equal class probabilities $P_Y(i) = 1/3, i \in \{1, 2, 3\}$, means μ_i, and a generic covariance matrix $\mathbf{\Sigma}$, which is equal for all classes. In class, we saw that the Bayes decision rule between classes i and j is a hyperplane. We denote its normal vector by $\mathbf{w}^{ij}$ and the point needed to specify the hyper-plane (in addition to the normal) by $\mathbf{x}_0^{ij}$.

a)(10 points) Starting from the Bayes decision rule, derive the expressions of the parameters of hyperplane ij as a function of the parameters of Gaussians i and j. *(Note: if you don't know how to derive the expressions, but remember them from class you can simply write them down. However, you will only receive partial credit.)*

b)(10 points) Assume that μ_1, μ_2, $\mathbf{w}^{12}$, and $\mathbf{w}^{23}$ are known, and all remaining variables are unknown. Is this sufficient information to determine $\mathbf{w}^{13}$? If so, provide an expression for $\mathbf{w}^{13}$ in terms of the known quantities. If not explain why (including what additional pieces of information would be needed).

c)(10 points) Assume that $\mathbf{w}^{12}$, $\mathbf{w}^{23}$, $\mathbf{w}^{13}$ and Σ are known, and all remaining variables are unknown. Is this sufficient information to determine $\mathbf{x}_0^{12}$, $\mathbf{x}_0^{23}$, and $\mathbf{x}_0^{13}$? If so, provide an expression for the variables that can be determined, in terms of the known quantities. If not explain why (including what additional pieces of information would be needed).

3. A common estimator of the mean μ of a Gaussian random variable X, from a sample $\mathcal{D} = \{x_1, \ldots, x_n\}$ is linear estimator

$$\hat{\mu} = \alpha \sum_i w_i x_i \tag{1}$$

This generalizes the sample mean, which is the special case where $\alpha = 1/n$ and $w_i = 1, \forall i$. In class, we have made various statements about this type of estimators for the case where the random variables X_i (from which the samples x_i are drawn) are independent. In this problem we investigate what happens when this is not the case. More precisely, we consider the case where the random variables X_i are *dependent,* have common mean μ, but unconstrained variances.

a) (10 points) We start by studying the bias. Assuming that the w_i are known, determine the value of α that makes the estimator unbiased.

b) (10 points) we next want to study the variance of $\hat{\mu}$. However, this is a little more complicated than what we did in class. To warm up to the problem show that if a random variable Y is defined as

$$Y = \mathbf{v}^T \mathbf{X} \tag{2}$$

where $\mathbf{v}$ is a constant vector and $\mathbf{X}$ a vector of *dependent* random variables, then

$$var(Y) = \sum_i v_i^2 \sigma_i^2 + 2 \sum_i \sum_{j:i<j} v_i v_j \sigma_{ij} \tag{3}$$

where

$$\begin{aligned} \sigma_i^2 &= var(X_i) = E_{X_i}[(X_i - \mu)^2] & (4) \\ \sigma_{ij} &= cov(X_i, X_j) = E_{X_i, X_j}[(X_i - \mu)(X_j - \mu)] & (5) \end{aligned}$$

(*Note: even if you cannot do this question, feel free to use the formula in the following questions.*)

c) In the remaining questions, we assume that $\mathbf{X}$ contains two sets of variables. The first m variables are dependent on each other, have variance $\sigma_i^2 = \sigma^2$ and covariance $\sigma_{i,j} = \gamma$. The remaining $n - m$ variables are independent of all other variables and have variance $\sigma_i^2 = \sigma^2$. We can summarize this as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_D \\ \mathbf{X}_I \end{bmatrix} \quad \mathbf{\Sigma_X} = \begin{bmatrix} \mathbf{\Sigma}_D & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_I \end{bmatrix} \tag{6}$$

where

- $\mathbf{X}_D$ is m dimensional and contains the dependent variables,
- $\mathbf{X}_I$ is $n - m$ dimensional and contains the independent variables,
- $\mathbf{\Sigma}_D$ is the covariance of $\mathbf{X}_D$. This is a full $m \times m$ covariance matrix with the value σ in the entries of the main diagonal and γ in the off-diagonal entries.
- $\mathbf{\Sigma}_I$ is the covariance of $\mathbf{X}_I$. This is a $(n - m) \times (n - m)$ diagonal matrix of the form $\mathbf{\Sigma}_I = \sigma^2 \mathbf{I}$.

c.1) (10 points) Consider the estimator of (1), assume that $w_i = 1$ for all i, and do the following

1. determine the value of α so that the estimator is unbiased

2. using this α, compute an expression for the variance of $\hat{\mu}$.

c.2) (10 points) Consider the estimator of (1), assume that $w_i = 0$ for all $i \leq m$ and $w_i = 1$ for all $i > m$, and do the following

1. determine the value of α so that the estimator is unbiased
2. using this α, compute an expression for the variance of $\hat{\mu}$.

c.3) (10 points) Denote the variance derived in **c.1)** as V_1 and that derived in **c.2)** as V_2. Derive a bound on $\frac{\sigma^2}{\gamma}$ as a function of n and m such that $V_2 \leq V_1$.

(Note: Note that the first estimator uses all the samples, while the second estimator throws away the dependent samples. What you have shown is that there are values of m and n such that you obtain an estimator of smaller variance by doing the latter. Since both estimators are unbiased, the latter is a better estimator. Note that this contradicts what we saw in class, where we said that "more data is always better". Although this is a good rule of thumb (and certainly true for *independent* observations) it may not necessarily hold when there are dependencies.)