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

Nuno Vasconcelos Fall 2017

1. a) According to bayes decision rule, the optimal decision function, under the "0/1" loss is "pick $i = 0$" if

$$\begin{aligned} P_{Y|X}(0|\mathcal{D}) &> P_{Y|X}(1|\mathcal{D}) \\ P_{X|Y}(\mathcal{D}|0)P_Y(0) &> P_{X|Y}(\mathcal{D}|1)P_Y(1) \\ P_{X|Y}(x_1,\ldots,x_n|0)P_Y(0) &> P_{X|Y}(x_1,\ldots,x_n|1)P_Y(1) \end{aligned}$$

Taking Log on both sides, and using the independence assumption,

$$\begin{aligned} \log\frac{P_{X|Y}(x_1,\ldots,x_n|0)}{P_{X|Y}(x_1,\ldots,x_n|1)} + \log\frac{P_Y(0)}{P_Y(1)} &> 0 \\ \log(\lambda_0^n e^{-\lambda_0(x_1+\ldots+x_n)}) - \log(\lambda_1^n e^{-\lambda_1(x_1+\ldots+x_n)}) + \log\frac{\pi_0}{\pi_1} &> 0 \\ n\log(\frac{\lambda_0}{\lambda_1}) + ns_n(\lambda_1-\lambda_0) + \log\frac{\pi_0}{\pi_1} &> 0 \\ s_n &> \frac{1}{(\lambda_1-\lambda_0)}\left(\log(\frac{\lambda_1}{\lambda_0}) + \frac{1}{n}\log\frac{\pi_1}{\pi_0}\right) \end{aligned}$$

Therefore, the threshold is, $T = \frac{1}{(\lambda_1-\lambda_0)}\left(\log(\frac{\lambda_1}{\lambda_0}) + \frac{1}{n}\log\frac{\pi_1}{\pi_0}\right)$

b) The ML estimate of the parameter b is given by,

$$b^* = \arg\max_b P_X(\mathcal{D})$$

where,

$$\begin{aligned} P_X(\mathcal{D}) &= \prod_1^n P_X(x_i) \\ \log P_X(\mathcal{D}) &= \sum_1^n \log P_X(x_i) \\ &= \sum_1^n (\log\lambda - \lambda x_i) \\ &= n\log\lambda - \lambda\sum_1^n \log x_i \end{aligned}$$

Taking gradient with respect to λ, and setting to zero

$$\frac{\partial \log P_X(\mathcal{D})}{\partial\lambda} = \frac{n}{\lambda} - \sum_1^n x_i = 0$$

It follows that,

$$\begin{aligned}\lambda^* &= \frac{n}{\sum_1^n x_i} \\ &= \frac{1}{s_n}\end{aligned}$$

Computing the second derivative,

$$\frac{\partial^2 \log P_X(\mathcal{D})}{\partial b^2} = -\frac{n}{\lambda^2}$$

is always negative, therefore we have a maximum at $\lambda^* = \frac{1}{s_n}$.

2. a) According to bayes decision rule, the optimal decision function, under the "0/1" loss is "pick i over j" if

$$\begin{aligned} P_{Y|X}(i|x) &> P_{Y|X}(j|x) \\ P_{X|Y}(x|i)P_Y(i) &> P_{X|Y}(x|j)P_Y(j) \\ P_{X|Y}(x|i) &> P_{X|Y}(x|j) \text{ (as the priors are equal)} \end{aligned}$$

Taking Log on both sides,

$$\begin{aligned} \log P_{X|Y}(x|i) - \log P_{X|Y}(x|j) &> 0 \\ (x-\mu_j)^T\Sigma^{-1}(x-\mu_j) - (x-\mu_i)^T\Sigma^{-1}(x-\mu_i) &> 0 \\ 2(\mu_i-\mu_j)^T\Sigma^{-1}x - (\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j) &> 0 \\ \Rightarrow (\mu_i-\mu_j)^T\Sigma^{-1}\left(x - \frac{1}{2}(\mu_i+\mu_j)\right) &> 0 \end{aligned}$$

So the decision boundary can be expressed as a hyperplane, $w_{ij}^T(x - x_0^{ij}) = 0$, where w_{ij} is the normal and x_0^{ij} is a point through the hyperplane, with

$$w_{ij} = \Sigma^{-1}(\mu_i - \mu_j) \quad (1)$$
$$x_0^{ij} = \frac{1}{2}(\mu_i + \mu_j) \quad (2)$$

b) Yes, we can determine w_{13} using the known quantities. Using 1, we can write

$$w_{12} = \Sigma^{-1}(\mu_1 - \mu_2) \quad (3)$$
$$w_{23} = \Sigma^{-1}(\mu_2 - \mu_3) \quad (4)$$
$$\Rightarrow w_{12} + w_{23} = \Sigma^{-1}(\mu_1 - \mu_3) \quad (5)$$
$$= w_{13} \quad (6)$$

c) No, we cannot determine the points $x_0^{12}, x_0^{23}, x_0^{13}$ using the known quantities. Using given quantities we can compute,

$$\begin{aligned} (\mu_1 - \mu_2) &= \Sigma w_{12} \\ (\mu_2 - \mu_3) &= \Sigma w_{23} \\ (\mu_1 - \mu_3) &= \Sigma w_{13} \end{aligned}$$

But as shown in part **b)** the three equations are not independent, and one relation can be obtained from the other two. So we cannot determine $x_0^{12}, x_0^{23}, x_0^{13}$, or equivalently, μ_1, μ_2, μ_3. However, we just need one of μ_1, μ_2, μ_3 to determine the other two. Geometrically, the given quantities fix the distance between the means. But an arbitrary translation to all three points will still preserve the distances. So we need atleast one point to determine the positions uniquely.

3. a)

$$E[\hat{\mu}] = \alpha \sum_i w_i E_{X_i}[x_i] \tag{7}$$
$$= \mu\alpha \sum_i w_i \tag{8}$$
$$\tag{9}$$

Hence the estimator is unbiased if

$$\alpha = \frac{1}{\sum_i w_i} \tag{10}$$

2. b) Using the fact that

$$E[Y] = \mathbf{v}^T E_{\mathbf{X}}[\mathbf{X}] = \mu \mathbf{v}^T \mathbf{1}, \tag{11}$$

where **1** is the vector with all entries 1,

$$\begin{aligned}
var(Y) &= E_Y[(Y - E_Y[Y])^2] = E_Y[(\mathbf{v}^T\mathbf{X} - \mu\mathbf{v}^T\mathbf{1})^2] \\
&= E_Y[(\mathbf{v}^T(\mathbf{X} - \mu\mathbf{1}))^2] = E_Y[(\mathbf{v}^T(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}]))^2] \\
&= E_Y[\mathbf{v}^T(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])^T\mathbf{v}] \\
&= \mathbf{v}^T E_Y[(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])^T]\mathbf{v} \\
&= \mathbf{v}^T \mathbf{\Sigma_X} \mathbf{v} = \sum_{ij} v_i v_j (\mathbf{\Sigma_X})_{ij} \\
&= \sum_i v_i^2 \sigma_i^2 + \sum_{i,j \neq i} v_i v_j \sigma_{ij} \\
&= \sum_i v_i^2 \sigma_i^2 + 2 \sum_{i,j<i} v_i v_j \sigma_{ij}
\end{aligned}$$

where we have used the fact that $\sigma_{ij} = \sigma_{ji}$ by symmetry of the covariance matrix.

c.1)

1. Using our result from **a)**,

$$\alpha = \frac{1}{n}. \tag{12}$$

2. the estimator is of the form of **b)**, with $v_i = \frac{1}{n}, \forall i$. Hence

$$var(Y) = \frac{1}{n^2} \sum_i \sigma_i^2 + 2\frac{1}{n^2} \sum_{i,j<i} \sigma_{ij} \tag{13}$$
$$= \frac{1}{n^2} n\sigma^2 + 2\frac{1}{n^2} \frac{m(m-1)}{2} \gamma \tag{14}$$
$$= \frac{\sigma^2}{n} + \frac{m(m-1)}{n^2} \gamma \tag{15}$$

c.2)

1. Using our result from **a)**,

$$\alpha = \frac{1}{n-m}. \tag{16}$$

2. the estimator is of the form of **b)**, with $v_i = 0, \forall i \leq m$, and $v_i = \frac{1}{n-m}, \forall i > m$. Hence

$$\begin{aligned} var(Y) &= \frac{1}{(n-m)^2}\sum_{i=m}^{n}\sigma_i^2 + 2\frac{1}{(n-m)^2}\sum_{i=m}^{n}\sum_{j=m}^{i-1}\sigma_{ij} && (17)\\ &= \frac{1}{(n-m)^2}(n-m)\sigma^2 + 0 && (18)\\ &= \frac{\sigma^2}{n-m} && (19)\end{aligned}$$

c.3) The bound is

$$\begin{aligned} \frac{\sigma^2}{n-m} &\leq \frac{\sigma^2}{n} + \frac{m(m-1)}{n^2}\gamma && (20)\\ \sigma^2\left(\frac{1}{n-m} - \frac{1}{n}\right) &\leq \frac{m(m-1)}{n^2}\gamma && (21)\\ \sigma^2\frac{m}{n(n-m)} &\leq \frac{m(m-1)}{n^2}\gamma && (22)\\ \frac{\sigma^2}{\gamma} &\leq \frac{(m-1)(n-m)}{n}. && (23)\end{aligned}$$

The plot below shows the bound for $n = 100$, as a function of m. Note that, as long as $\sigma^2/gamma$ is below the curve, it is a good idea to throw away the dependent samples. The ratio is largest when only half of the points are independent, in which case σ^2 can be as large as 25γ. This is large ratio, meaning that the dependencies are quite small. In summary, dependencies can be detrimental even when they are very small. This contradicts the rule we discussed in class, where we hinted that it is always good to add more data. Note that this is indeed true for absolutely independent samples. i.e. $m = 1$, in which case the bound is zero.