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

Nuno Vasconcelos Fall 2013

1.a) The log-likelihood function is

$$\begin{aligned}\mathcal{L}(\mathcal{D})(\theta) &= \sum_{i=1}^{N} \left\{ (k-1)\log x_i - \frac{x_i}{\theta} - k\log(\theta) - \log((k-1)!) \right\} \\ &= (k-1)\sum_{i=1}^{N} \log x_i - \sum_{i=1}^{N} \frac{x_i}{\theta} - kN\log(\theta) - N\log((k-1)!)\end{aligned}$$

The derivative with respect to θ is zero when

$$\frac{\partial l(\mathcal{D})}{\partial \theta}(\theta) = \frac{1}{\theta^2}\sum_{i=1}^{N} x_i - \frac{kN}{\theta} = 0$$

or

$$\theta^* = \frac{1}{kN}\sum_{i=1}^{N} x_i.$$

The second derivative is

$$\frac{\partial^2 l(\mathcal{D})}{\partial \theta^2}(\theta) = -\frac{2}{\theta^3}\sum_{i=1}^{N} x_i + \frac{kN}{\theta^2} = \frac{1}{\theta^2}\left(-\frac{2}{\theta}\sum_{i=1}^{N} x_i + kN\right)$$

At the critical point this is

$$\frac{\partial^2 l(\mathcal{D})}{\partial \theta^2}(\theta^*) = \frac{1}{(\theta^*)^2}\left(-2kN + kN\right) = -\frac{kN}{(\theta^*)^2} < 0$$

and θ^* is an ML estimate.

b) Using the properties of the Gamma distribution,

$$\sum_{i=1}^{N} X_i \sim Gamma(x; kN, \theta)$$

and

$$\theta^* \sim Gamma\left(x; kN, \frac{\theta}{kN}\right).$$

It follows that

$$E_{X_1,\ldots,X_N}[\theta^*] = kN\frac{\theta}{kN} = \theta$$

and the estimator is unbiased. Also

$$var_{X_1,\ldots,X_N}[\theta^*] = kN\left(\frac{\theta}{kN}\right)^2 = \frac{\theta^2}{kN}.$$

Finally, since there is no bias, the MSE is the same as the variance.

c) In this case, the log-likelihood is a function of both k and θ

$$\mathcal{L}(\mathcal{D})(k,\theta) \quad = \quad (k-1)\sum_{i=1}^{N}\log x_i - \sum_{i=1}^{N}\frac{x_i}{\theta} - kN\log(\theta) - N\log((k-1)!).$$

We have seen that, for any k, this is maximized when

$$\theta = \frac{1}{kN}\sum_{i=1}^{N} x_i.$$

Plugging this into the log-likelihood, we obtain

$$\begin{aligned}
\mathcal{L}(\mathcal{D})(k) \quad &= \quad (k-1)\sum_{i=1}^{N}\log x_i - kN - kN\log\left(\sum_{i=1}^{N} x_i\right) + kN\log(kN) - N\log((k-1)!) \\
&= \quad k\left(\sum_{i=1}^{N}\log x_i - N - N\log\left(\sum_{i=1}^{N} x_i\right)\right) + kN\log(kN) - N\log((k-1)!) \\
&= \quad ak + N\log\frac{(kN)^k}{(k-1)!}
\end{aligned}$$

with

$$a = \sum_{i=1}^{N}\log x_i - N - N\log\left(\sum_{i=1}^{N} x_i\right).$$

Since k is an integer, the following procedure will find the ML solution

1. using the sample $\mathcal{D}$ compute a.
2. starting with $k = 1$, search for the k^* that maximizes

   $$ak + N\log\frac{(kN)^k}{(k-1)!}$$

   This can be done by simply trying all the values of k in $1 < k < k_{max}$, where k_{max} is some large integer, and picking the one that leads to the largest value of the function.
3. set

   $$\theta^* = \frac{1}{k^*N}\sum_{i=1}^{N} x_i.$$

4. return k^* and θ^*.

2. a) The BDR is

$$i^* = \arg\min_i ||\mathbf{x} - \mu_i||^2.$$

The decision boundary is given by

$$\begin{aligned}
||\mathbf{x} - \mu_1||^2 &= ||\mathbf{x} - \mu_{-1}||^2 \\
-2\mathbf{x}^T\mu_1 + ||\mu_1||^2 &= -2\mathbf{x}^T\mu_{-1} + ||\mu_{-1}||^2 \\
(\mu_1 - \mu_{-1})^T\mathbf{x} - \frac{1}{2}(||\mu_1||^2 - ||\mu_{-1}||^2) &= 0
\end{aligned}$$

Hence, the decision boundary as the desired form with

$$\begin{aligned}
\mathbf{w} &= \mu_1 - \mu_{-1} \\
b &= -\frac{1}{2}(||\mu_1||^2 - ||\mu_{-1}||^2).
\end{aligned}$$

b) The ML estimate for the mean is $\hat{\mu} = \frac{1}{n}\sum_i \mathbf{x}_i$. Hence, we have

$$\begin{aligned}
\hat{\mathbf{w}} &= \left(\frac{1}{n_1}\sum_{i|y_i=1}\mathbf{x}_i - \frac{1}{n_{-1}}\sum_{i|y_i=-1}\mathbf{x}_i\right) \\
\hat{b} &= -\left(\frac{1}{2n_1^2}\sum_{i,j|y_i=y_j=1}\mathbf{x}_i^T\mathbf{x}_j - \frac{1}{2n_{-1}^2}\sum_{i,j|y_i=y_j=-1}\mathbf{x}_i^T\mathbf{x}_j\right).
\end{aligned}$$

where

$$\begin{aligned}
n_1 &= \sum_i\left(\frac{1+y_i}{2}\right) \\
n_{-1} &= \sum_i\left(\frac{1-y_i}{2}\right)
\end{aligned}$$

are the numbers of positive and negative examples in Joe's sample.

c) The expressions above can be written as

$$\begin{aligned}
\hat{\mathbf{w}} &= \sum_i \alpha_i\mathbf{x}_i \\
\hat{b} &= -\frac{1}{2}\sum_{i,j|y_i=y_j} y_i\alpha_i^2\mathbf{x}_i^T\mathbf{x}_j.
\end{aligned}$$

where

$$\alpha_i = \begin{cases} 1/n_1, & \text{if } y_i = 1 \\ -1/n_{-1}, & \text{if } y_i = -1 \end{cases} \tag{1}$$

d) This problem is less trivial than it may seem initially. The reason is that the number of points in each class is itself a random variable (N_1 positives and N_1 negatives) since we don't know what the values of n_1 and n_{-1} are. We need to account for this when we consider the expected value of $\hat{\mathbf{w}}$. As usual, we have

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

Note, however, that α_i depends on the whole sample $Y_1, \ldots, Y_n$ and we cannot just get rid of the Y_i. We can use the fact, however, that α_i does not depend on the Y_i exactly, just on the number of positive examples N_1. This leads to

$$\begin{aligned}
\mu_{\mathbf{w}} &= \sum_i E_{\mathbf{X}_i, Y_1, \ldots, Y_n}[\alpha_i \mathbf{X}_i] = n E_{\mathbf{X}, Y, N_1}[\alpha \mathbf{X}] = n E_{N_1}[E_{\mathbf{X}, Y|N_1}[\alpha \mathbf{X} | N_1 = k]] \\
&= n E_{N_1}[P_{Y|N_1}(1|k) E_{\mathbf{X}|N_1, Y}[\alpha \mathbf{X} | N_1 = k, Y = 1] + P_{Y|N_1}(-1|k) E_{\mathbf{X}|N_1, Y}[\alpha \mathbf{X} | N_1 = k, Y = -1]] \\
&= n E_{N_1}\left[P_{Y|N_1}(1|k) \frac{1}{k} \mu_1 - P_{Y|N_1}(-1|k) \frac{1}{n-k} \mu_{-1}\right] \\
&= E_{N_1}\left[P_{Y|N_1}(1|k) \frac{n}{k}\right] \mu_1 - E_{N_1}\left[P_{Y|N_1}(-1|k) \frac{n}{n-k}\right] \mu_{-1} \\
&= E_{N_1}\left[\frac{P_{N_1|Y}(k|1) P_Y(1)}{P_{N_1}(k)} \frac{n}{k}\right] \mu_1 - E_{N_1}\left[\frac{P_{N_1|Y}(k|-1) P_Y(-1)}{P_{N_1}(k)} \frac{n}{n-k}\right] \mu_{-1} \\
&= E_{N_1}\left[\frac{\binom{n-1}{k-1} P_Y(1)^{k-1} P_Y(-1)^{n-k} P_Y(1)}{\binom{n}{k} P_Y(1)^k P_Y(-1)^{n-k}} \frac{n}{k}\right] \mu_1 - E_{N_1}\left[\frac{\binom{n-1}{k} P_Y(1)^k P_Y(-1)^{n-1-k} P_Y(-1)}{\binom{n}{k} P_Y(1)^k P_Y(-1)^{n-k}} \frac{n}{n-k}\right] \mu_{-1} \\
&= E_{N_1}\left[\frac{k}{n} \frac{n}{k}\right] \mu_1 - E_{N_1}\left[\frac{n-k}{n} \frac{n}{n-k}\right] \mu_{-1} \\
&= \mu_1 - \mu_{-1} \\
&= \mathbf{w}
\end{aligned}$$