Spring 2017 EECE 5644 Homework #2 Solutions

Prof: Jaume Coll-Font TA: Aziz Kocanaogullari, Zhiqiang Tao

2.1 Problem 2.12 from Duda's book

Solution:

(a)

$$\begin{aligned} 1 = \sum_{i=1}^{c} P(\omega_i|x) &\leq \sum_{i=1}^{c} P(\omega_{max}|x) \\ &= P(\omega_{max}|x) \sum_{i=1}^{c} 1 \\ &= P(\omega_{max}|x)c \end{aligned} \tag{1}$$

Therefore,

$$P(\omega_{max}|x) = \frac{1}{c} \tag{2}$$

(b)

$$\begin{aligned} p(error) =& 1 - P(correct) \\ &= 1 - \int P(correct|x)p(x)dx \\ &= 1 - \int \max_i P(\omega_i|x)p(x)dx \\ &= 1 - \int P(\omega_{max}|x)p(x)dx \end{aligned} \tag{3}$$

(c)

$$\begin{aligned} p(error) =& 1 - \int P(\omega_{max}|x)p(x)dx \\ &\leq 1 - \int \frac{1}{c}p(x)dx \\ &\leq 1 - \frac{1}{c}\int p(x)dx \\ &\leq 1 - \frac{1}{c} \\ &\leq \frac{c-1}{c} \end{aligned} \tag{4}$$

(d) Equality occurs when probabilities are equal

2.2 (10 pts) Problem 3.1 from Duda's book. Skip (c) (removed from the problem statements below)

Solution:

(a) From the formulas, we obtain figure 1

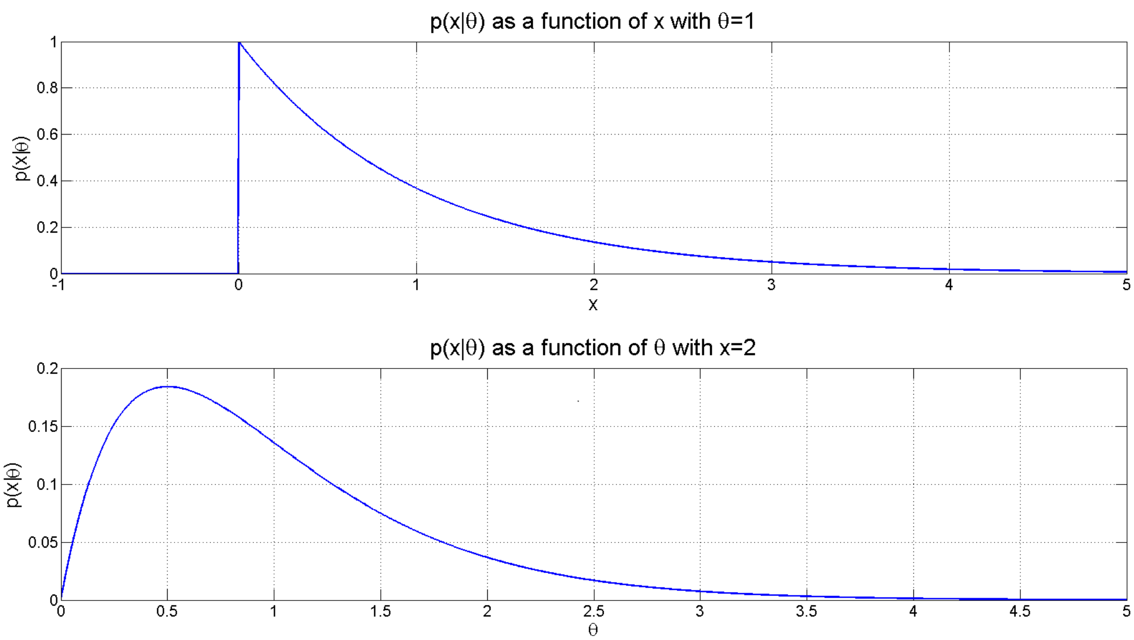

Figure 1: Problem 2.1a

(b) The MLE estimate is given by:

$$\hat{\theta}_{\text{mle}} = \arg\max_{\theta} p(D|\theta) = \arg\max_{\theta} \prod_{i=1}^{n} p(x_i|\theta) = \arg\max_{\theta} \prod_{i=1}^{n} \theta \ e^{-\theta x_i} = \arg\max_{\theta} \theta^n \ e^{-\sum_{i=1}^{n} \theta x_i}$$

$$= \arg\max_{\theta} n \ln\theta + \ln e^{-\sum_{i=1}^{n} \theta x_i} = \arg\max_{\theta} n \ln\theta - \theta \sum_{i=1}^{n} x_i = \arg\max_{\theta} f(\theta)$$

Taking the derivative and equating to 0, we obtain:

$$\left.\frac{\partial f(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{\text{mle}}} = n\frac{1}{\hat{\theta}_{\text{mle}}} - \sum_{i=1}^{n} x_i = 0$$

$$\hat{\theta}_{\text{mle}} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

2.3 (10 pts) Problem 3.2 from Duda's book. Skip (b) (removed from the problem statements below)

Solution:

(a) The MLE estimate is given by:

$$\hat{\theta}_{\text{mle}} = \arg\max_\theta p(D|\theta) = \arg\max_\theta \prod_{i=1}^{n} p(x_i|\theta) = \arg\max_\theta \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{1}\{0 \le x_i \le \theta\}$$
$$= \arg\max_\theta \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}\{0 \le x_i \le \theta\}$$
$$= \arg\max_\theta \frac{1}{\theta^n}\mathbb{1}\{\theta \ge \max_i x_i\}\ \mathbb{1}\{\min_i x_i \ge 0\}$$

This function decreases as θ increases. In addition, we notice that this function is 0 for $\theta < x_i$ for all i. Therefore, we need θ to be the smallest value that to makes the likelihood non-zero. We can achieve this by making $\hat{\theta}_{\text{mle}} = \max_i\{x_i\}$.

2.4 (20 pts) Problem 3.3 from Duda's book.

Solution:

(a) Since z_{ij} can only be 0 or 1, we can express the $P(z_{ij}|P(\omega_i))$ as:

$$P(z_{ij}|P(\omega_i)) = P(\omega_i)^{z_{ij}}(1 - P(\omega_i))^{1-z_{ij}}$$

Therefore, since the states are sampled independently,

$$P(z_{i1}, \cdots, z_{in}|P(\omega_i)) = \prod_{k=1}^{n} P(z_{ik}|P(\omega_i)) = \prod_{k=1}^{n} P(\omega_i)^{z_{ik}}(1 - P(\omega_i))^{1-z_{ik}}$$

(b) Let $\theta = P(\omega_i)$. Then the MLE estimate is given by

$$\hat{\theta}_{\text{mle}} = \arg\max_\theta \ln\left(\prod_{k=1}^{n} \theta^{z_{ik}}(1-\theta)^{1-z_{ik}}\right) = \arg\max_\theta \sum_{k=1}^{n} z_{ik}\ln\theta + (1 - z_{ik})\ln(1-\theta)$$

Taking the derivative and equating to 0, we get

$$\frac{1}{\hat{\theta}_{\text{mle}}}\sum_{k=1}^{n} z_{ik} - \frac{1}{1-\hat{\theta}_{\text{mle}}}\sum_{k=1}^{n} 1 - z_{ik} = 0$$
$$(1-\hat{\theta}_{\text{mle}})\sum_{k=1}^{n} z_{ik} = \hat{\theta}_{\text{mle}}\sum_{k=1}^{n} 1 - \hat{\theta}_{\text{mle}}\sum_{k=1}^{n} z_{ik}$$
$$\hat{\theta}_{\text{mle}} = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$$

This result can be interpreted as the Laplacian view of probability: we estimate the probability of an even by counting the number of occurrences and dividing by the total number of events.

2.5 (10 pts) Problem 3.4 from Duda's book.

Solution:

We know that $P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i}$. Therefore, we can obtain the MLE estimate by:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{\text{mle}} &= \arg\max_{\boldsymbol{\theta}} \ln P(D|\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \ln \prod_{j=1}^{n} \prod_{i=1}^{d} \theta_i^{x_{ji}}(1-\theta_i)^{1-x_{ji}} \\
&= \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^{n} \sum_{i=1}^{d} \ln \theta_i^{x_{ji}}(1-\theta_i)^{1-x_{ji}} \\
&= \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^{n} \sum_{i=1}^{d} x_{ji} \ln \theta_i + (1-x_{ji}) \ln(1-\theta_i)
\end{aligned}$$

Taking the partial derivative with respect to θ_k and equating to 0, we get

$$\begin{aligned}
\frac{1}{\hat{\theta}_{\text{mle},k}} \sum_{j=1}^{n} x_{jk} - \frac{1}{1-\hat{\theta}_{\text{mle},k}} \left(\sum_{j=1}^{n} 1 - \sum_{j=1}^{n} x_{jk} \right) &= 0 \\
\sum_{j=1}^{n} x_{jk} - \hat{\theta}_{\text{mle},k} \sum_{j=1}^{n} x_{jk} &= \hat{\theta}_{\text{mle},k} n - \hat{\theta}_{\text{mle},k} \sum_{j=1}^{n} x_{jk} \\
\hat{\theta}_{\text{mle},k} &= \frac{1}{n} \sum_{j=1}^{n} x_{jk}
\end{aligned}$$

Therefore,

$$\hat{\boldsymbol{\theta}}_{\text{mle}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$$

2.6 (40 pts) Problem 3.17 from Duda's book.

Solution:

(a) We know that $\sum_{k=1}^{n} x_{ji} = s_i$.

$$
\begin{aligned}
P(D|\theta) &= \prod_{k=1}^{n} P(\mathbf{x}_k|\boldsymbol{\theta}) \\
&= \prod_{k=1}^{n}\prod_{i=1}^{d} \theta_i^{x_{ki}}(1-\theta_i)^{1-x_{ki}} \\
&= \prod_{i=1}^{d}\prod_{k=1}^{n} \theta_i^{x_{ki}}(1-\theta_i)^{1-x_{ki}} \\
&= \prod_{i=1}^{d}(\theta_i^{x_{1i}}\cdot\theta_i^{x_{2i}}\cdots\theta_i^{x_{ni}})((1-\theta_i)^{1-x_{1i}}\cdot(1-\theta_i)^{1-x_{2i}}\cdots(1-\theta_i)^{1-x_{ni}}) \\
&= \prod_{i=1}^{d}\theta_i^{\sum_{k=1}^{n} x_{ki}}(1-\theta_i)^{\sum_{k=1}^{n} 1-x_{ki}} \\
&= \prod_{i=1}^{d}\theta_i^{\sum_{k=1}^{n} x_{ki}}(1-\theta_i)^{n-\sum_{k=1}^{n} x_{ki}} \\
&= \prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}
\end{aligned}
$$

(b) We know that $\theta_i \in [0,1]$ since it determines a probability. Thus, the prior is a uniform prior between 0 and 1.

$$
\begin{aligned}
p(\boldsymbol{\theta}|D) &= \frac{p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \\
&= \frac{p(D|\boldsymbol{\theta})}{\int p(D|\boldsymbol{\theta})d\boldsymbol{\theta}} \\
&= \frac{\prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}}{\int\prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}d\boldsymbol{\theta}} \\
&= \prod_{i=1}^{d}\frac{1}{\int\theta_i^{s_i}(1-\theta_i)^{n-s_i}d\boldsymbol{\theta}}\theta_i^{s_i}(1-\theta_i)^{n-s_i} \\
&= \prod_{i=1}^{d}\frac{1}{\frac{s_1!(n-s_i)!}{(s_i+(n-s_i)+1)!}}\theta_i^{s_i}(1-\theta_i)^{n-s_i} \\
&= \prod_{i=1}^{d}\frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}
\end{aligned}
$$

(c) Using the attached MATLAB script, we obtain figure 2

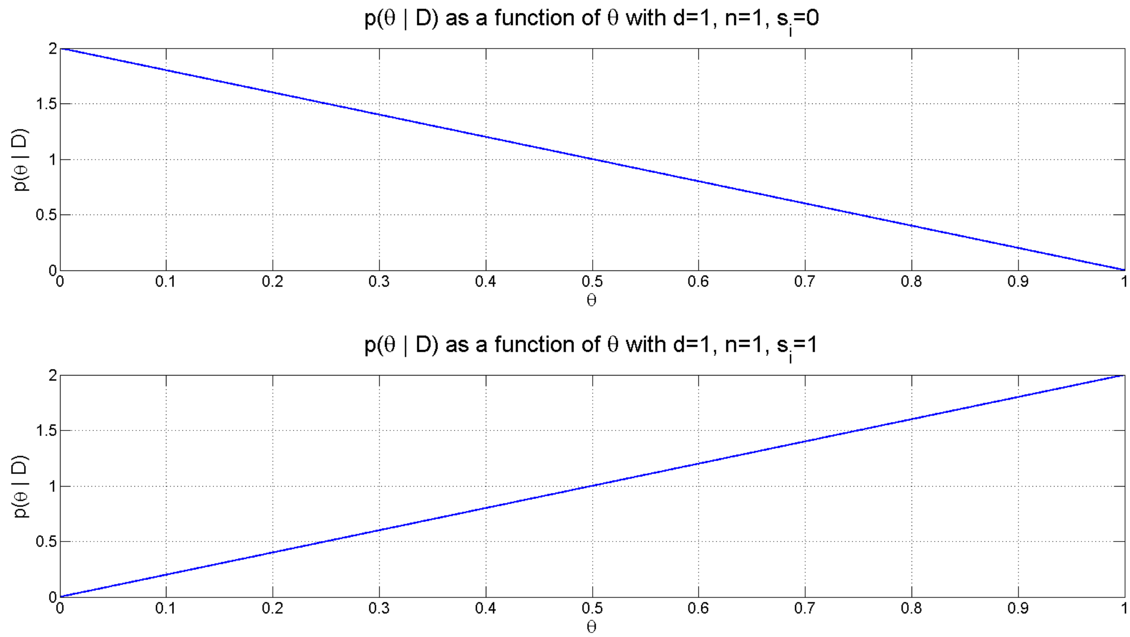

Figure 2: Problem 2.5c

(d)

$$
\begin{aligned}
\int P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|D)d\boldsymbol{\theta} &= \int \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i} \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}d\boldsymbol{\theta} \\
&= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \cdot \int \prod_{i=1}^{d} \boldsymbol{\theta}_i^{x_i}(1-\theta_i)^{1-x_i}\theta_i^{s_i}(1-\theta_i)^{n-s_i}d\boldsymbol{\theta} \\
&= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \cdot \prod_{i=1}^{d} \int \theta_i^{s_i+x_i}(1-\theta_i)^{n-s_i+1-x_i}d\boldsymbol{\theta} \\
&= \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \cdot \prod_{i=1}^{d} \frac{(s_i+x_i)!(n-s_i+1-x_i)!}{(s_i+x_i+n-s_i+1-x_i+1)!} \\
&= \prod_{i=1}^{d} \frac{(n+1)!(s_i+x_i)!(n-s_i+1-x_i)!}{s_i!(n-s_i)!(n+2)!} \\
&= \prod_{i=1}^{d} \frac{(s_i+x_i)!(n-s_i+1-x_i)!}{s_i!(n-s_i)!(n+2)}
\end{aligned}
$$

Since x_i can only be 0 or 1, we have that

$$
\int P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|D)d\boldsymbol{\theta} = \prod_{i=1}^{d} \left(\frac{(s_i+1)!(n-s_i)!}{s_i!(n-s_i)!(n+2)}\right)^{x_i} \left(\frac{s_i!(n-s_i+1)!}{s_i!(n-s_i)!(n+2)}\right)^{1-x_i}
$$

Simplifying the factorials,

$$\int P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|D)d\boldsymbol{\theta} = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(\frac{n-s_i+1}{n+2}\right)^{1-x_i}$$
$$= \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1-\frac{s_i+1}{n+2}\right)^{1-x_i}$$

(e) Looking at the formulas for the distributions, we can make $p(\mathbf{x}|D) = p(\mathbf{x}|\hat{\boldsymbol{\theta}}_{\text{bayes}})$ if we let $\hat{\theta}_{\text{bayes},i} = \frac{s_i+1}{n+2}$