Homework #1

COEN 240 Winter 2021

Q1. (20 pts) Consider a two-class, one-dimensional problem where $P(w_1) = P(w_2)$ and $p(x|w_i) \sim N(\mu_i, \sigma_i^2)$. Let $\mu_1 = 0, \sigma_1^2 = 1$, and $\mu_2 = \mu, \sigma_2^2 = \sigma^2$.

(a) Derive a general expression for the location of the Bayes optimal decision boundary as a function of μ and σ^2.

(b) With $\mu = 1$ and $\sigma^2 = 2$, make two plots (with Python or MATLAB): one for the class conditional pdfs $p(x|w_i)$ and one for the posterior probabilities $p(w_i|x)$ with the location of the optimal decision regions. Make sure the plots are correctly labeled (axis, titles, legend, etc) and that the fonts are legible when printed.

(c) Estimate the Bayes error rate p_e conditioning on (b).

Q2. (15 pts) Problem 3.1 (a-b) from the book [1] as shown in Fig. 1.

1. Let x have an exponential density

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Plot $p(x|\theta)$ versus x for $\theta = 1$. Plot $p(x|\theta)$ versus θ, $(0 \leq \theta \leq 5)$, for $x = 2$.

(b) Suppose that n samples $x_1, ..., x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum likelihood estimate for θ is given by

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum\limits_{k=1}^{n} x_k}.$$

Figure 1: Details of **Q2**. Figure adapted from [1].

Q3. (10 pts) Problem 3.2 (a) from the book [1] as shown in Fig. 2.

Q4. (10 pts) Problem 3.3 from the book [1] as shown in Fig. 3.

Q5. (10 pts) Problem 3.4 from the book [1] as shown in Fig. 4.

Q6. (35 pts) Problem 3.17 from the book [1] as shown in Fig. 5.

References

[1] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley, 2 edition, 2001.

2. Let x have a uniform density

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{otherwise.} \end{cases}$$

(a) Suppose that n samples $\mathcal{D} = \{x_1, ..., x_n\}$ are drawn independently according to $p(x|\theta)$. Show that the maximum likelihood estimate for θ is $\max[\mathcal{D}]$, i.e., the value of the maximum element in $\mathcal{D}$.

Figure 2: Details of **Q3**. Figure adapted from [1].

3. Maximum likelihood methods apply to estimates of prior probabilities as well. Let samples be drawn by successive, independent selections of a state of nature ω_i with unknown probability $P(\omega_i)$. Let $z_{ik} = 1$ if the state of nature for the kth sample is ω_i and $z_{ik} = 0$ otherwise.

(a) Show that

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

(b) Show that the maximum likelihood estimate for $P(\omega_i)$ is

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik}.$$

Interpret your result in words.

Figure 3: Details of **Q4**. Figure adapted from [1].

4. Let $\mathbf{x}$ be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

$$P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1 - \theta_i)^{1-x_i},$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)^t$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum likelihood estimate for $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k.$$

Figure 4: Details of **Q5**. Figure adapted from [1].

17. The purpose of this problem is to derive the Bayesian classifier for the d-dimensional multivariate Bernoulli case. As usual, work with each class separately, interpreting $P(\mathbf{x}|\mathcal{D})$ to mean $P(\mathbf{x}|\mathcal{D}_i, \omega_i)$. Let the conditional probability for a given category be given by

$$P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

and let $\mathcal{D} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be a set of n samples independently drawn according to this probability density.

(a) If $\mathbf{s} = (s_1, ..., s_d)^t$ is the sum of the n samples, show that

$$P(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

(b) Assuming a uniform a priori distribution for $\boldsymbol{\theta}$ and using the identity

$$\int_0^1 \theta^m(1-\theta)^n \; d\theta = \frac{m!n!}{(m+n+1)!},$$

show that

$$p(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

(c) Plot this density for the case $d = 1, n = 1$, and for the two resulting possibilities for s_1.

(d) Integrate the product $P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})$ over $\boldsymbol{\theta}$ to obtain the desired conditional probability

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

(e) If we think of obtaining $P(\mathbf{x}|\mathcal{D})$ by substituting an estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in $P(\mathbf{x}|\boldsymbol{\theta})$, what is the effective Bayesian estimate for $\boldsymbol{\theta}$?

Figure 5: Details of **Q6**. Figure adapted from [1].