CS559-B HW1 Solution

Due: Sep. 25th, 2019

Problem 1 (10pt): Independence and un-correlation

(1) Suppose X and Y are two continuous random variables, show that if X and Y are independent, then they are uncorrelated.

Solution: suppose X has density function $f_X(x)$, Y has density $f_Y(y)$. If X and Y independent, then $f(x,y) = f_X(x)f_Y(y)$. Therefore:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \int_x\int_y (x-\mu_X)(y-\mu_Y)f(x,y)dxdy \\
&= \int_x\int_y (x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y)dxdy \\
&= \int_x (x-\mu_X)f_X(x)dx \int_y (y-\mu_Y)f_Y(y)dy \\
&= \left(\int_x xf_X(x)dx - \mu_X\right)\left(\int_y yf_Y(y)dy - \mu_Y\right) = 0.
\end{aligned}$$

(2) Suppose X and Y are uncorrelated, can we conclude X and Y are independent? If so, prove it, otherwise, give one counterexample. (Hint: consider $X \sim Uniform[-1,1]$ and $Y = X^2$)

Solution: consider $X \sim Uniform[-1,1]$ and $Y = X^2$. Then X and Y are not independent. However, $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$ and $\mathbb{E}(X) = 0$, thus, $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$

Problem 2 (15pt): [Minimum Error Rate Decision] Let $\omega_{max}(x)$ be state of nature for which $P(\omega_{max}|x) \geq P(\omega_i|x)$ for all $i = 1, \ldots, c$.

(1) Show that $P(\omega_{max}|x) \geq \frac{1}{c}$

Solution: since $P(\omega_{max}|x) \geq P(\omega_i|x)$, then:

$$\sum_{i=1}^{c} P(\omega_{max}|x) \geq \sum_{i=1}^{c} P(\omega_i|x) = 1$$

So $P(\omega_{max}|x) \geq \frac{1}{c}$.

(2) Show that for minimum-error-rate decision rule, the average probability of error is given by

$$P(error) = 1 - \int P(\omega_{max}|x)p(x)dx$$

Solution: by definition of averaged probability of error, we have:

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

(3) Show that $P(error) \leq \frac{c-1}{c}$

Solution: from (2) and (1) we have:

$$P(error) = 1 - \int P(\omega_{max}|x)p(x)dx \leq 1 - \int \frac{1}{c}p(x)dx = 1 - \frac{1}{c} = \frac{c-1}{c}$$

Problem 3 (10pt): [Likelihood Ratio] Suppose we consider two category classification, the class conditionals are assumed to be Gaussian, i.e., $p(x|\omega_1) = \mathrm{N}(4,1)$ and $p(x|\omega_2) = \mathrm{N}(8,1)$, based on prior knowledge, we have $P(\omega_2) = \frac{1}{4}$. We do not penalize for correct classification, while for misclassification, we put 1 unit penalty for misclassifying ω_1 to ω_2 and put 3 unit for misclassifying ω_2 to ω_1. Derive the bayesian decision rule using likelihood ratio.

Solution: we have the likelihood distribution: $p(x|\omega_1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-4)^2}{2}}$, $p(x|\omega_2) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-8)^2}{2}}$, and we could also easily obtain the prior $p(\omega_2) = \frac{1}{4}$, $p(\omega_1) = 1 - p(\omega_2) = \frac{3}{4}$. Based on the problem, we could get the loss matrix as:

$$\lambda = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

For bayesian decision based on likelihood ratio, we decide ω_1 if the following:

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

For RHS, we have $RHS = \frac{3-0}{1-0} * \frac{1}{3} = 1$, For LHS, we have $\frac{p(x|\omega_1)}{p(x|\omega_2)} = e^{-\frac{(x-4)^2}{2} + \frac{(x-8)^2}{2}}$. Therefore we decide ω_1 if:

$$e^{-\frac{(x-4)^2}{2} + \frac{(x-8)^2}{2}} > 1$$

Take ln and solve the equation, we get bayesian decision rule which is decide ω_1 if $x < 6$, decide ω_2 otherwise.

Problem 4 (15pt): [Minimum Risk, Reject Option] In many machine learning applications, one has the option either to assign the pattern to one of c classes, or to reject it as being unrecognizable.

If the cost for reject is not too high, rejection may be a desirable action. Let

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0, \ i = j \text{ and } i, j = 1, \dots, c \\ \lambda_r, \ i = c+1 \\ \lambda_s, \ \text{otherwise} \end{cases}$$

where λ_r is the loss incurred for choosing the $(c+1)$-th action, rejection, and λ_s is the loss incurred for making any substitution error.

(1) Derive the decision rule with minimum risk.

Solution: for $i = 1, \dots, c$:

$$\begin{aligned} R(\alpha_i|x) &= \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|x) \\ &= \lambda_s \sum_{j=1, j \neq i} P(\omega_j|x) \\ &= \lambda_s [1 - P(\omega_j|x)] \end{aligned}$$

For $i = c+1$:

$$R(\alpha_{c+1}|x) = \lambda_r$$

Thus, decision rule based on minimum risk would be: decide ω_i if $R(\alpha_i|x) \leq R(\alpha_{c+1}|x)$, i.e., $P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$, and reject otherwise.

(2) What happens if $\lambda_r = 0$?

Solution: if $\lambda_r = 0$, we always reject.

(3) What happens if $\lambda_r > \lambda_s$?

Solution: if $\lambda_r > \lambda_s$, we never reject.

Problem 5 (10pt): [Maximum Likelihood Estimation (MLE)] Suppose we have training samples $\{x_1, x_2, \dots, x_n\}$. Consider the following distributions:

(1) Exponential density: $f(x; \theta) = \theta e^{-\theta x}$, $x \geq 0$, $\theta > 0$, find MLE for θ

Solution: the likelihood function is:

$$L(\theta; x_1, \dots, x_n) = \theta^n \exp(-\theta \sum_{i=1}^{n} x_i)$$

The log-likelihood function would be:

$$l(\theta; x_1, \dots, x_n) = n \ln(\theta) - \theta \sum_{i=1}^{n} x_i$$

Let $\frac{dl}{d\theta} = 0$, we have the MLE for θ:

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

(2) Uniform distribution $Unif[\theta_1, \theta_2]$, find the MLE for θ_1 and θ_2.

Solution: we do not need to use derivative calculus to find the MLE in this case. The density for $Unif[\theta_1, \theta_2]$ is $\frac{1}{\theta_2 - \theta_1}$ on $[\theta_1, \theta_2]$. Therefore, the likelihood function is:

$$L(\theta_1, \theta_2 | x_1, \ldots, x_n) = \begin{cases} (\frac{1}{\theta_2 - \theta_1})^n, & \text{if all } x_i \text{ in the interval } [\theta_1, \theta_2] \\ 0, & \text{otherwise} \end{cases}$$

We maximize the likelihood function by making $\theta_2 - \theta_1$ as small as possible. The only constraint is that the interval $[\theta_1, \theta_2]$ should include all data. Therefore, the MLE for θ_1 and θ_2 is:

$$\begin{aligned} \hat{\theta_1} &= \min(x_1, \ldots, x_n) \\ \hat{\theta_2} &= \max(x_1, \ldots, x_n) \end{aligned}$$