

ECE 271A: Homework #1

Due on October 16, 2023 at 11:59pm

Professor Vasconcelos

Ray Tsai

A16848188

Problem 1

In this problem we will consider the traditional probability scenario of coin tossing. However, we will consider two variations. First, the coin is not fair. Denoting by S the outcome of the coin toss we have

$$P_S(\text{heads}) = \alpha, \alpha \in [0, 1].$$

Second, you do not observe the coin directly but have to rely on a friend that reports the outcome of the toss. Unfortunately your friend is unreliable, he will sometimes report heads when the outcome was tails and vice-versa. Denoting the report by R we have

$$P_{R|S}(\text{tails}|\text{heads}) = \theta_1$$

$$P_{R|S}(\text{heads}|\text{tails}) = \theta_2$$

where $\theta_1, \theta_2 \in [0, 1]$. Your job is to, given the report from your friend, guess the outcome of the toss.

Part A

Given that your friend reports heads, what is the optimal decision function in the minimum probability of error sense. That is, when should you guess heads, and when should you guess tails?

Solution

We decide heads if

$$\begin{aligned} P_{S|R}(\text{heads}|\text{heads}) &> P_{S|R}(\text{tail}|\text{heads}) \\ P_{R|S}(\text{heads}|\text{heads})P_S(\text{heads}) &> P_{R|S}(\text{heads}|\text{tails})P_S(\text{tails}) \\ (1 - \theta_1)\alpha &> \theta_2(1 - \alpha) \\ \alpha &> \frac{\theta_2}{1 - \theta_1 + \theta_2}. \end{aligned}$$

Otherwise, we choose tails.

Part B

Consider the case $\theta_1 = \theta_2$. Can you give an intuitive interpretation to the rule derived in part a?

Solution

Suppose that $\theta_1 = \theta_2 = \theta$, then we should pick heads if $\alpha > \theta$. That is, given that the probability of my friend misreporting is independent to the outcome of the coin flip, I should pick heads if the probability of flipping heads is greater than the chances of my friend tripping me.

Part C

You figured out that if you ask your friend to report the outcome of the toss various times, he will produce reports that are statistically independent. You then decide to ask him to report the outcome n times, in the hope that this will reduce the uncertainty. (Note: there is still only one coin toss, but the outcome gets

reported n times). What is the new minimum probability of error decision rule?

Solution

Let R_i be the i -th report of the outcome. Suppose that there are k heads out of those n reports. Then, we should pick heads if

$$\begin{aligned} P_{S|R_1, R_2, \dots, R_n}(\text{heads}|r_1, r_2, \dots, r_n) &> P_{S|R_1, R_2, \dots, R_n}(\text{tails}|r_1, r_2, \dots, r_n) \\ P_{R_1, R_2, \dots, R_n|S}(r_1, r_2, \dots, r_n|\text{heads})P_S(\text{heads}) &> P_{R_1, R_2, \dots, R_n|S}(r_1, r_2, \dots, r_n|\text{tails})P_S(\text{tails}) \\ (1 - \theta_1)^k \theta_1^{n-k} \alpha &> \theta_2^k (1 - \theta_2)^{n-k} (1 - \alpha) \\ \alpha &> \frac{\theta_2^k (1 - \theta_2)^{n-k}}{(1 - \theta_1)^k \theta_1^{n-k} + \theta_2^k (1 - \theta_2)^{n-k}}. \end{aligned}$$

Part D

Consider the case $\theta_1 = \theta_2 = \theta$ and assume that the report sequence is all heads. Can you give an intuitive interpretation to the rule derived in part C?

Solution

Suppose that $\theta_1 = \theta_2 = \theta$ and all reports are head. From the result in part C, we should pick heads when

$$\alpha > \frac{\theta^n}{\theta^n + (1 - \theta)^n} = \frac{1}{\left(\frac{1-\theta}{\theta}\right)^n + 1}.$$

Let $T = \frac{1}{\left(\frac{1-\theta}{\theta}\right)^n + 1}$, and let $n \rightarrow \infty$. If $1 - \theta > \theta$, then $T \rightarrow 0$. If $1 - \theta < \theta$, then $T \rightarrow 1$. If $\theta = \frac{1}{2}$, then $T = \frac{1}{2}$.

This implies that if there's a one half chance that your friend will trip you, then you should ignore your friend's report and purely focus on your prior belief. Otherwise, the more report you get, the more you should base your decision on your friend's trustworthiness.

Problem 2

Suppose two equally probable one-dimensional densities are of the form $p(x|w_i) \propto e^{-|x-a_i|/b_i}$ for $i = 1, 2$ and $0 < b_i$.

Part A

Write an analytic expression for each density, that is, normalize each function for arbitrary a_i and positive b_i .

Solution

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-|x-a_i|/b_i} dx &= \int_{a_i}^{\infty} e^{(-x+a_i)/b_i} dx + \int_{-\infty}^{a_i} e^{(x-a_i)/b_i} dx \\ &= e^{a_i/b_i} \int_{a_i}^{\infty} e^{-x/b_i} dx + e^{-a_i/b_i} \int_{-\infty}^{a_i} e^{x/b_i} dx \\ &= 2b_i. \end{aligned}$$

Thus, we need to multiply each function by $\frac{1}{2b_i}$ to normalize them, that is,

$$p(x|w_i) = \frac{1}{2b_i} e^{-|x-a_i|/b_i},$$

for $i = 1, 2$.

Part B

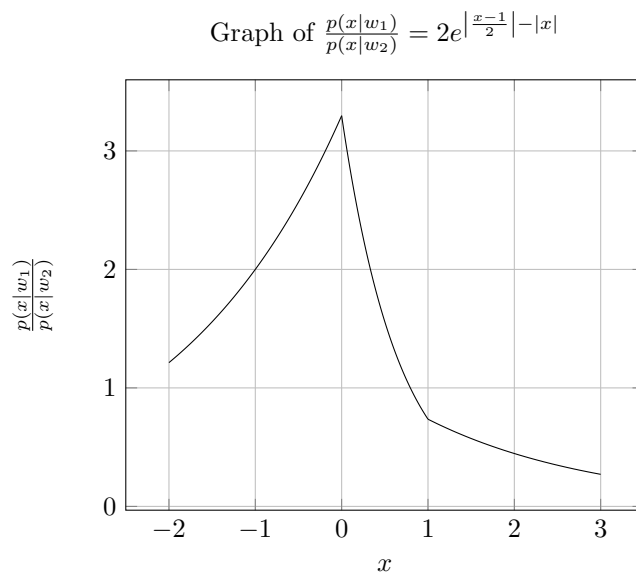
Calculate the likelihood ratio as a function of your four variables.

Solution

$$\frac{p(x|w_1)}{p(x|w_2)} = \frac{b_2}{b_1} e^{|x-a_2|/b_2 - |x-a_1|/b_1}.$$

Part C

Sketch a graph of the likelihood ratio $\frac{p(x|w_1)}{p(x|w_2)}$ for the case $a_1 = 0, b_1 = 1, a_2 = 1, b_2 = 2$.

Solution

Problem 3

Consider the three-dimensional normal distribution $p(x|\omega) \sim N(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}.$$

Part A

Find the probability density at the point $x_0 = (.5, 0, 1)^t$.

Solution

$$p(x_0|\omega) = 0.0082.$$

Part B

Construct the whitening transformation A_ω . Compute the matrices representing eigenvectors and eigenvalues, Φ and Λ . Next, convert the distribution to one centered on the origin with covariance matrix equal to the identity matrix, $p(x|\omega) \sim N(0, I)$.

Solution

We first normalize the mean by subtracting x by μ and get $z = x - \mu$. We know $z \sim N(0, \Sigma)$.

We then normalize the covariance matrix. We perform eigenvalue decomposition and get $\Sigma = \Phi \Lambda \Phi^T$, where

$$\Phi \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.7071 & 0.7071 \\ 0 & 0.7071 & 0.7071 \end{pmatrix} \text{ and } \Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

Let $A_w = \Phi \Lambda^{-\frac{1}{2}} \approx \begin{pmatrix} 1.0000 & 0 & 0 \\ 0 & -0.4082 & 0.2673 \\ 0 & 0.4082 & 0.2673 \end{pmatrix}$. Then $A_w^T \Sigma A_w = \Lambda^{-\frac{1}{2}} \Phi^T \Phi \Lambda \Phi^T \Phi \Lambda^{-\frac{1}{2}} = I$.

Thus, $y = A_w^T z = A_w^T (x - \mu)$ is the normalization transformation we are looking for.

Part C

Apply the same overall transformation to x_0 to yield a transformed point x_w .

Solution

$$x_w = A_w^T (x_0 - \mu) = (-0.5, 0.4082, -0.8018)^T.$$

Part D

By explicit calculation, confirm that the Mahalanobis distance from x_0 to the mean μ in the original distribution is the same as for x_w to 0 in the transformed distribution.

Solution

$$r^2 = (x_0 - \mu)^T \Sigma^{-1} (x_0 - \mu) = 1.0595 = x_w^T x_w.$$

Part E

Does the probability density remain unchanged under a general transformation? In other words, is $p(x_0|N(\mu, \Sigma)) = p(T^t x_0|N(T^t \mu, T^t \Sigma T))$ for some linear transform T ?

Solution

Let $y = T^t x_0$. Notice that the Mahalanobis distance of the transformed distribution is

$$\begin{aligned} (T^t x_0 - T^t \mu)^t (T^t \Sigma T)^{-1} (T^t x_0 - T^t \mu) &= (x_0 - \mu)^t T T^{-1} \Sigma^{-1} (T^t)^{-1} T^t (x_0 - \mu) \\ &= (x_0 - \mu)^t \Sigma^{-1} (x_0 - \mu), \end{aligned}$$

which remains the same. However, since the transformation changed the covariance matrix from Σ to $T^t \Sigma T$, the normalizing term of each distribution, $\frac{1}{|\Sigma|}$ and $\frac{1}{|T^t \Sigma T|}$, differs unless $|T| = \pm 1$. Since normalization is not preserved, the transformed probability density is not unchanged in general.

Part F

Prove that the whitening transform $A_w = \Phi \Lambda^{-1/2}$ when applied to a Gaussian distribution ensures that the final distribution has covariance proportional to the identity matrix I . Check whether normalization is preserved by the transformation.

Solution

Let there be a distribution $p(x|N(\mu, \Sigma))$, and let $y = A_w^t x$. Then,

$$\begin{aligned} \Sigma_y &= E[(y - \mu_y)(y - \mu_y)^t] \\ &= E[(A_w^t x - A_w^t \mu)(A_w^t x - A_w^t \mu)^t] \\ &= E[A_w^t (x - \mu)(x - \mu)^t A_w] \\ &= A_w^t \Sigma A_w \\ &= (\Lambda^{-1/2} \Phi^t)(\Phi \Lambda \Phi^t)(\Phi \Lambda^{-1/2}) \\ &= I. \end{aligned}$$

Thus, we have shown that the covariance after the whitening transformation is indeed I .

However, since

$$p(x|N(\mu, \Sigma)) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^t \Sigma^{-1} (x - \mu)\right)$$

but $p(y) =$

$$\begin{aligned}
 p(y) &= \frac{1}{\sqrt{(2\pi)^d |\Sigma_y|}} \exp \left(-\frac{1}{2} (y - \mu_y)^t \Sigma_y^{-1} (y - \mu_y) \right) \\
 &= \frac{1}{\sqrt{(2\pi)^d}} \exp \left(-\frac{1}{2} (A_w^t x - A_w^t \mu)^t \Sigma_y^{-1} (A_w^t x - A_w^t \mu) \right) \\
 &= \frac{1}{\sqrt{(2\pi)^d}} \exp \left(-\frac{1}{2} (x - \mu)^t A_w A_w^t (x - \mu) \right) \\
 &= \frac{1}{\sqrt{(2\pi)^d}} \exp \left(-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right),
 \end{aligned}$$

the distribution differs as the normalization fail to hold unless $|\Sigma| = \pm 1$.

Problem 4

Let the components of the vector $x = (x_1, \dots, x_d)^t$ be binary valued and $P(\omega_j)$ be the prior probability for the state of nature ω_j and $j = 1, \dots, c$. Now define

$$p_{ij} = \text{Prob}(x_i = 1 | \omega_j) \quad i = 1, \dots, d \quad j = 1, \dots, c,$$

with the components of x_i being statistically independent for all x in ω_j .

- (a) Interpret in words the meaning of p_{ij} .

Solution

p_{ij} is the probability that x_i is 1 given the state of nature is ω_j .

- (b) Show that the minimum probability of error is achieved by the following decision rule: Decide ω_k if $g_k(x) \geq g_j(x)$ for all j and k , where

$$g_j(x) = \sum_{i=1}^d x_i \ln \frac{p_{ij}}{1 - p_{ij}} + \sum_{i=1}^d \ln(1 - p_{ij}) + \ln P(\omega_j).$$

Solution

Let $g^*(x)$ be the decision rule of minimum probability error. We first note that

$$P_{X_i|Y}(x_i | \omega_j) = \begin{cases} p_{ij} & x_i = 1 \\ 1 - p_{ij} & x_i = 0 \end{cases} = p_{ij}^{x_i} (1 - p_{ij})^{1-x_i}.$$

Then,

$$\begin{aligned} g^*(x) &= \arg \max_{\omega_j} P_{Y|X}(\omega_j | x) \\ &= \arg \max_{\omega_j} P_{X|Y}(x | \omega_j) P(\omega_j) \\ &= \arg \max_{\omega_j} \ln P_{X|Y}(x | \omega_j) + \ln P(\omega_j) \\ &= \arg \max_{\omega_j} \ln(\prod_i P_{X_i|Y}(x_i | \omega_j)) + \ln P(\omega_j) \\ &= \arg \max_{\omega_j} \ln(\prod_i p_{ij}^{x_i} (1 - p_{ij})^{1-x_i}) + \ln P(\omega_j) \\ &= \arg \max_{\omega_j} \sum_i \ln(p_{ij}^{x_i} (1 - p_{ij})^{1-x_i}) + \ln P(\omega_j) \\ &= \arg \max_{\omega_j} \sum_i (x_i \ln(p_{ij}) + (1 - x_i) \ln(1 - p_{ij})) + \ln P(\omega_j) \\ &= \arg \max_{\omega_j} \sum_i \left(x_i \ln \left(\frac{p_{ij}}{1 - p_{ij}} \right) + \ln(1 - p_{ij}) \right) + \ln P(\omega_j) \\ &= \arg \max_{\omega_j} \sum_i x_i \ln \left(\frac{p_{ij}}{1 - p_{ij}} \right) + \sum_i \ln(1 - p_{ij}) + \ln P(\omega_j). \end{aligned}$$