CSE 575: Homework #1

Amey Bhilegaonkar

09/24/2022

Problem 1

a)

Solution: Since X and Y are independent events, their joint probability will be equal to -
$\mathbb{P}(X \cap Y) = \mathbb{P}(X) * \mathbb{P}(Y)$ ...... (1)

and using conditional probability Equation -

$\mathbb{P}(X|Y) = \mathbb{P}(X \cap Y)/\mathbb{P}(Y)$ ...... (2)

Combining Equation 1 and 2, we can say, if X and Y are independent then,
$\mathbb{P}(X|Y) = \mathbb{P}(X)$

b)

Solution: Since X and Y are disjoint events, their joint probability will be equal to -
$\mathbb{P}(X \cap Y) = \mathbb{P}(X) * \mathbb{P}(Y) = 0$ ...... (1)

and using conditional probability Equation -

$\mathbb{P}(X|Y) = \mathbb{P}(X \cap Y)/\mathbb{P}(Y)$ ...... (2)

Combining Equation 1 and 2, we can say, if X and Y are disjoint then,
$\mathbb{P}(X|Y) = 0$

Solution: From the Given information we can derive:

1 $\mathbb{P}(C1 = H) = 0.6$
2 $\mathbb{P}(C1 = T) = 0.4$
3 $\mathbb{P}(C2 = H) = 0.4$
4 $\mathbb{P}(C2 = T) = 0.6$

For getting the sequence - HT,HT,TT,TT We need to calculate the probability of each of the event happening individually and multiply it to get the probability of the whole sequence happening at once.

Which means -

$\mathbb{P}(Sequence) = \mathbb{P}(HT) * \mathbb{P}(HT) * \mathbb{P}(TT) * \mathbb{P}(TT)$ .... (1)

$\mathbb{P}(HT) = \mathbb{P}(C1 = H) * \mathbb{P}(C2 = T)$ .... (2)

$\mathbb{P}(TT) = \mathbb{P}(C1 = T) * \mathbb{P}(C2 = T)$ .... (3)

Combining all the equations we get -

$\mathbb{P}(Sequence) = (0.6 * 0.6)^2 * (0.4 * 0.6)^2$

c) $\mathbb{P}(Sequence) = 0.00746$

Solution: The likelihood is
$\mathbb{P}(X|\theta) = \theta^{15} * (1 - \theta)^5$
d) $\theta_{MLE} = \frac{15}{15+5} = 0.75$

Problem 2

a) **Solution:** The error function can be defined as
$E(w) = \frac{1}{2} \times \sum_{n=1}^{N} (y(x_n) - t_n)^2$
$where\ y(x) = w^T.x$

here w is a 3×2 weight matrix with dummy weights and $x = (x_n^T\ 1)^T$
and x_n is the nth feature vector of dimension 2×1
and t_n can take values $[1\ 0]^T$ if it belongs to C_1 and $[0\ 1]^T$ if it belongs to C_2 depending on the output.

To find the weight vector, differentiate the Error function and equate it to 0 to find the minimum.

In matrix form, the weight vector W (calculated for all features)can be given as: $W = (X^T X)^{-1} X^T T$

Here, X is the matrix containing all the feature vectors $[x_1 x_2 ... x_n]^T$ where all feature vectors have

From the above equation we get the value of W as: $W = \begin{pmatrix} 0.166 \\ -0.44 \\ 0.94 \end{pmatrix}$

b) **Solution:** Assume projected means of for the classes $C_1\ and\ C_2$ equal to $m_1\ and\ m_2$ respectively.

Let $s_1^2\ and\ s_2^2$ be the projected variances for feature vectors for two classes $C_1 and C_2$ respectively

The Fishers' error function for Linear Discriminant is:

$J(w) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$ In matrix notation: $J(w) = \frac{w^T S_B w}{w^T S_w w}$

S_B is between class covariance matrix. $S_B = (m_2 - m_1)(m_2 - m_1)^T$

S_W is within class covariance matrix. $S_W = \sum_{n \in C_1} (x_n - m_1)(x_n - m_1)^T + \sum_{n \in C_2} (x_n - m_2)(x_n - m_2)^T$

The weight vector w can be given as:

$w = S_w^{-1}(m_1 - m_2)$

after calculation we get,

$w = \begin{pmatrix} -3 \\ 8 \end{pmatrix}$ Normalizing w, we get : $w = \begin{pmatrix} -0.3511 \\ 0.936 \end{pmatrix}$

The value for $w^T x$ for x$_1, x_2, x_3, and x_4 are 3, -2, -18 and -15 respectively. We can set the threshold for$ w$^T x$ to -8.

Solution: Below are the values of updated weights and the current parameter for each iteration.

Interation: 1

Value of x_n : $\begin{bmatrix}1\\0\\1\end{bmatrix}$

$Value of t_n : 1$

$Updated value of w :$

w = $\begin{bmatrix}2.5\\0.\\1.\end{bmatrix}$

Interation: 2

$Value of x_n : \begin{bmatrix}2\\1\\1\end{bmatrix}$

$Value of t_n : 1$

$Updated value of w :$

w = $\begin{bmatrix}4.5\\1.\\2.\end{bmatrix}$

Interation: 3

$Value of x_n : \begin{bmatrix}2\\3\\1\end{bmatrix}$

$Value of t_n : -1$

$Updated value of w :$

w = $\begin{bmatrix}2.5\\-2.\\1.\end{bmatrix}$

Interation: 4

$Value of x_n : \begin{bmatrix}3\\3\\1\end{bmatrix}$

$Value of t_n : -1$

$Updated value of w :$

w = $\begin{bmatrix}-0.5\\-5.\\0.\end{bmatrix}$

c)

Problem 3

a) **Solution:** $\mathbb{P}(y=2) = 1 - \mathbb{P}(y=1) = 1 - 0.4 = 0.6$

b) **Solution:** $\mathbb{P}(y=1|x) = \frac{\mathbb{P}(y=1|x) * \mathbb{P}(y=1)}{\mathbb{P}(x)}$

$\mathbb{P}(y=1|x) = \frac{\mathbb{P}(x|y=1) * \mathbb{P}(y=1)}{\mathbb{P}(x|y=1) * \mathbb{P}(y=1) + \mathbb{P}(x|y=2) * \mathbb{P}(y=2)}$

$\mathbb{P}(y=1|x) = \frac{0.5*0.4}{(0.5*0.4)+(0.125*0.6)}$

$\mathbb{P}(y=1|x) = \frac{0.5*0.4}{(0.5*0.4)+(0.125*0.6)}$

$\mathbb{P}(y=1|x) = \frac{0.2}{0.275}$

$\mathbb{P}(y=1|x) = 0.727$

for 2 ¡= x ¡= 8, the value of $\mathbb{P}$(x — y=1) = 0, so $\mathbb{P}$(y =1 — x) =0.

Therefore,

$\mathbb{P}$(y = 1 — x) = $\begin{cases} 0.727 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$

c) **Solution:** $\mathbb{P}(y=1|x=1) = 0.727$ $\mathbb{P}(y=2|x=1) = 0.272$

Since, $\mathbb{P}(y=1|x=1) > \mathbb{P}(y=2|x=1)$, Bayes classifier will assign the label y =1 for x =1.

The risk of this decision is $\mathbb{P}(y=2|x=1) = 0.272$

d) **Solution:** The decision function f(x) will be: f(x) = $\begin{cases} y=1 & 0 \leq x \leq 2 \\ y=2 & 2¡x \leq 8 \\ 1 or 2 & \text{otherwise} \end{cases}$

Problem 4

a) **Solution:** $\mathbb{P}(y=2) = 1 - \mathbb{P}(y=1) = 1 - 0.4 = 0.6$

b) **Solution:** $\mathbb{P}(y=1|x) = \frac{\mathbb{P}(x|y=1) * \mathbb{P}(y=1)}{\mathbb{P}(x|y=1) * \mathbb{P}(y=1) + \mathbb{P}(x|y=2) * \mathbb{P}(y=2)}$

Applying same to every other pair of x we get

(1) $\mathbb{P}(y=1|x_1=0, x_2=0) = \frac{0.4*0.6}{(0.4*0.6)+(0.3*0.4)} = 0.666$

(2) $\mathbb{P}(y=1|x_1=0, x_2=1) = \frac{0.3*0.6}{(0.3*0.6)+(0.1*0.4)} = 0.8181$

(3) $\mathbb{P}(y=1|x_1=1, x_2=0) = \frac{0.2*0.6}{(0.2*0.6)+(0.4*0.4)} = 0.4285$

(4) $\mathbb{P}(y=1|x_1=0, x_2=0) = \frac{0.1*0.6}{(0.1*0.6)+(0.2*0.4)} = 0.4285$

c) **Solution:** $\mathbb{P}(y=1|x_1=0, x_2=1) = 0.818$ $\mathbb{P}(y=2|x_1=0, x_2=1) = 0.182$

Since, $\mathbb{P}(y=1|x_1=0, x_2=1) > \mathbb{P}(y=2|x_1=0, x_2=1)$, Bayes classifier will assign the label y =1 for x =1.

The risk of this decision is $\mathbb{P}(y=2||x_1=0, x_2=1) = 0.182$

Problem 4 continued on next page…

d) **Solution:** The decision function f(x) will be: f(x) = $\begin{cases} y = 1 & x_1 = 0, x_2 = 0 \\ & x_1 = 0, x_2 = 1 \\ \\ y = 2 & x_1 = 1, x_2 = 0 \\ & x_1 = 1, x_2 = 1 \end{cases}$

Problem 5

a) **Solution:** There are a total of **4** independent parameters - Sky, Temp, Humid, Wind.
This is mainly because, if we calculate the co-variance matrix of parameters with each other, we can see that, the variance of the parameters with itself is maximum compared to others, so they can be called independent of other parameters.

Solution: Let's make some tables -

Table - **SKY**

Id	Yes	No	P(Y)	P(N)
Sunny	3	3	3/4	3/6
Rainy	1	3	1/4	3/6
Total	4	6	1	1

Table - **Humid**

Id	Yes	No	P(Y)	P(N)
Normal	4	2	4/4	2/6
High	0	4	0/4	4/6
Total	4	6	1	1

Table - **Temp**

Id	Yes	No	P(Y)	P(N)
Mild	2	2	2/4	2/6
Cold	1	2	1/4	2/6
Hot	1	2	1/4	2/6
Total	4	6	1	1

Table - **Wind**

Id	Yes	No	P(Y)	P(N)
Mild	2	2	2/4	2/6
Strong	2	4	2/4	4/6
Total	4	6	1	1

(1.1) $\mathbb{P}(x = sunny|y = Yes) = 3/4$
(1.2) $\mathbb{P}(x = sunny|y = No) = 3/6$
(1.2) $\mathbb{P}(x = Rainy|y = Yes) = 1/4$
(1.2) $\mathbb{P}(x = Rainy|y = No) = 3/6$

(2.1) $\mathbb{P}(x = mild|y = Yes) = 2/4$
(2.2) $\mathbb{P}(x = cold|y = Yes) = 1/4$
(2.3) $\mathbb{P}(x = hot|y = Yes) = 1/4$
(2.1) $\mathbb{P}(x = mild|y = No) = 2/6$
(2.2) $\mathbb{P}(x = cold|y = No) = 2/6$
(2.3) $\mathbb{P}(x = hot|y = No) = 2/6$

(3.1) $\mathbb{P}(x = Normal|y = Yes) = 4/4 = 1$
(3.2) $\mathbb{P}(x = Normal|y = No) = 2/6$
(3.3) $\mathbb{P}(x = High|y = Yes) = 0/4 = 0$
(3.4) $\mathbb{P}(x = High|y = No) = 4/6$

(1.1) $\mathbb{P}(x = Mild|y = Yes) = 2/4$
(1.2) $\mathbb{P}(x = Mild|y = No) = 2/6$
(1.1) $\mathbb{P}(x = Strong|y = Yes) = 2/4$
(1.2) $\mathbb{P}(x = Strong|y = No) = 4/6$

Problem 5 continued on next page. . .

c) **Solution:** Given input vector is - x = (sunny,cold,normal,strong)

Using Naiive Bayes assumption of class parameters being independent of each other and expanding Baye's theorem, we get -

$\mathbb{P}(Y = 1|x) = \frac{\mathbb{P}(x|Y=1)*\mathbb{P}(Y=1)}{\mathbb{P}(x|Y=1)*\mathbb{P}(Y=1)+\mathbb{P}(x|Y=0)*\mathbb{P}(Y=0)}$

$\mathbb{P}(Y = 1|x) = \frac{\mathbb{P}(x=sunny|Y=1)*\mathbb{P}(x=cold|Y=1)*\mathbb{P}(x=normal|Y=1)*\mathbb{P}(x=strong|Y=1)*\mathbb{P}(Y=1)}{\mathbb{P}(x=sunny|Y=1)*\mathbb{P}(x=cold|Y=1)*\mathbb{P}(x=normal|Y=1)*\mathbb{P}(x=strong|Y=1)*\mathbb{P}(Y=1)+\mathbb{P}(x=sunny|Y=0)*\mathbb{P}(x=cold|Y=0)*\mathbb{P}(x=n}$

$\mathbb{P}(Y = 1|x) = \frac{3/4*1/4*1*2/4*4/10}{3/4*1/4*1*2/4*4/10+3/6*2/6*2/6*4/6*6/10}$ $\mathbb{P}(Y = 1|x) = 27/43 = 0.628$

d) **Solution:** Given, parameter values from previous step: X = (sunny,cold,normal,missing)
Ignoring the missing x because, the possible values can be either of the two values in the 4 th feature. Therefore the probability of that feature becomes 1.

$\mathbb{P}(y = 1|x_1 = sunny, x_2 = cold, x_3 = normal, x_4 = miss)$ = $\mathbb{P}(x_1 = sunny, x_2 = cold, x_3 = normal, x_4 = miss|y = 1)p(y = 1)$
$\mathbb{P}(x1 = sunny, x2 = cold, x3 = normal, x4 = miss)$ =0.696

$\mathbb{P}(y = 1|x)$ = 0.696

As $\mathbb{P}(y = 1|x) > \mathbb{P}(y = 0|x)$ class label y=1 that is Yes label will be assigned by the Naive Bayes classifier