CS 541: Machine Learning, Spring 2024
Name: SRIHARI TANMAY KARTHIK TADALA

PSU ID: 918597835
Graduate Student

(1)

MACHINE LEARNING ASSIGNMENT-3

A1.) Given us Monty Hall case of 4 doors (3 goats, 1 car, host opens 1 door after selection)

Host knows what is behind each door

Assumption:- uniform prior distribution, meaning each door has an equal chance of car initially ie (1/4)th for each door.

Prior probability:- $P(h)$

$P(h_1) = 1/4$, $P(h_2) = 1/4$, $P(h_3) = 1/4$, $P(h_4) = 1/4$

Calculate the conditional probabilities:-

Door 1 choosen monty opens door 2

$P(D/h_1) = 1/3$ (since monty has 3 choices)

Door 2 choosen monty opens door 1

$P(D/h_2) = 0$ (Host Cant open the choosen door).

Door 3 choosen; Monty opens door 4.

$P(D/h_3) = 1/2$ (Monty opens and has 2 doors left)

Door 4 choosen; Monty opens door 3

$P(D/h_4) = 1/2$ (Mounty has 2 choice)

Calculating marginal probability $P(D)$:-

$$P(D) = [P(D/h_1) \cdot P(h_1) + P(D/h_2) \cdot P(h_2) + P(D/h_3) \cdot P(h_3) + P(D/h_4) \cdot P(h_4)]$$

$$= 1/3(1/4) + 0 + (1/2)(1/4) + (1/2)(1/4)$$

$$= 1/12 + 1/8 = 1/3$$

$P(D) = 1/3$

Calculating posterior probability $P(h/D)$:

$$P(h_i/D) = \frac{P(D/h_i) \cdot P(h_i)}{P(D)}$$

$$P(h_1/D) = \frac{1/3 \cdot 1/4}{1/3} = 1/4 \qquad P(h_2/D) = 0$$

$$P(h_3/D) = \frac{1/2 \cdot 1/4}{1/3} = 3/8$$

$$P(h_4/D) = \frac{1/2 \cdot 1/4}{1/3} = 3/8$$

The probability of car behind the 2 doors, you switch is 3/8 which is greater than the probability of first door (1/4)

(2)

Therefore it is advantageous to switch initial selection.

2.) Example of bias: using the logistic regression, when the data is not linearly seperable.

Example of varience:- Using SVM; on high dimension feature space can exhibit high varience.

Example of Noise:- Models like Naive Bayes classifier are more sensitive to noisy data because naive baise classifiers assume independence between features & may struggle with noisy or correlated data.

3.a) To show how new patient P7 is classified (under which class) = ?

Using naive Bayes

1) Use training data to compute probabilistic model.

P(cough = True | False | Flu) = 2/4
P(cough = False | Flu) = 2/4
P(Fever = True | Flu) = 3/4

P(cough = True | cold) = 1/2
P(cough = False | cold) = 1/2
P(Fever = True | cold) = 1/2

P(Fever = False / Flu) = 1/4	P(Fever = False / cold) = 1/2
P(Chills = True / Flu) = 3/4	P(Chiles = True / cold) = 1/2
P(Chills = False / Flu) = 1/4	P(Chiles = False / cold) = 1/2.

P(congestion = True / False) = 1/4 P(congestion = True / cold) = 1/2

P(congestion = False / flu) = 3/4 P(congestion = False / cold) = 0

(2) Lets find the prior class of

P(Flu) = 4/6, P(cold) = 2/6.

(3) The probability of flu class & cold class for patient P7 is

Flu class(X) = argmax P(Flu) ≤ P(Xi / Flu).

* P(Fever = False / Flu) * P(Chills = False / Flu) * P(congestion = True / False)]

Flu[X] = 4/6 x 2/4 x 1/4 x 1/4 x 1/4

Flu[X] = 0.0005

cold class[X] = P(cold) x P(Cough = True / cold) *

(3) $P(\text{Fever} = \text{False} \mid \text{cold}) \times P(\text{Chills} = \text{False} \mid \text{cold}) \times P(\text{congestion} = \text{true} \mid \text{cold})$

$\text{cold}[X] = 2/6 \times 1/2 \times 1/2 \times 1/2 \times 1 = 0.041$

Hence the probability of cold class is > then probability of flu class ($0.041 > 0.005$)

∴ Therefore patient P7 comes under cold class

b) Estimating probability / smoothing

$$P(x_i = a_i / c) = \frac{nc^{x_i = a_k} + 1}{nc + K}$$

To find

$P(\text{congestion} = \text{True} \mid \text{cold}) = \frac{2+1}{2+2} = 3/4$

$P(\text{congestion} = \text{False} \mid \text{cold}) = \frac{0+1}{2+2} = 1/4$

4) logistic regression:-

Given $B_0 = -5.247$, $B_1 = 3.626$, also output

1 = No Bacteria, 0 = Bacteria present

$$P(Y=1|X) = \sigma(w \cdot x) = \frac{1}{1+e^{-(B_0+B_1 x)}}$$

$$P(Y=0|X) = 1 - \sigma(w \cdot x) = \frac{e^{-(B_0+B_1 x)}}{1+e^{-(B_0+B_1 x)}}$$

To find

i) P(no Bacteria | dosage = 1.0)

$P(Y=1 | X=1.0) \Rightarrow 0.165$

ii) P(no bacteria | dosage = 2.5)

$$P(Y=1|X=2.5) = \frac{1}{1+e^{-(-5.247+3.626(2.5))}}$$

$= P(Y=1|X=2.5) = 0.979.$

iii) P(bacteria | dosage $x = 2.0$)

$$P(Y=0|X=2.0) = \frac{e^{-(-5.247+3.626\times 2)}}{1+e^{-(-5.247+3.626\times 2)}}$$

$= 0.118$

$P(Y=0|X=2.0) = 0.118$

4) B) 80% percentile cutoff for dosage producing no bacteria (ie) to find the dosage (x) = ?

$P(Y=1 \mid \text{dosage}(x)) = 0.8$

no bacteria

$$P(Y=1 \mid \text{dosage}(x)) = \frac{1}{1+e^{-(B_0+B_1x)}} = 0.8$$

$$\left(\frac{1}{0.8}\right)^{-1} = e^{-(B_0+B_1x)}$$

$$\ln(0.25) = e^{-(B_0+B_1x)}$$

$$\ln(0.25) = \ln\left(e^{-(B_0+B_1x)}\right)$$

$$\ln(0.25) = -(B_0+B_1x)$$

$$-\ln(0.25) = B_0+B_1x \quad \text{Multiplying by } -1.$$

$$\frac{-\ln(0.25)+5.247}{3.626} = x$$

$$x = 1.828$$

Therefore $x = 1.828$ dosage, so 80th percentile cutoff no bacteria

5) Given $x+y=1$

$x+2y=2$ we got

$x+3y=2$

$x+4y=3$

$$\underset{A}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}} \underset{x}{\begin{bmatrix} m \\ y \end{bmatrix}} = \underset{b}{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}}$$

Using derivative formulae for OLS

$$x^* = (A^T A)^{-1} A^T B$$

Now to find $(A^T A)^{-1}$

$$(A^T A)^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix} - (1)$$

Now lets find $A^T b$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 8 \\ 23 \end{bmatrix} - (2)$$

combine eq (1) and (2)

$$x^* = (A^T)(A)^T A^T b$$

$$x^* = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}$$

$$y = 0.6 + 0.5x$$

(5)

6. Given $A = \begin{bmatrix} 0 & 1 \\ 1 & 3/2 \end{bmatrix}$

To find singular values of A, we should know λ_1, λ_2.

also σ_1 and $\sigma_2 = ?$, $\sigma = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}$

Now let us find A^TA;

$$A^TA = \begin{bmatrix} 0 & 1 \\ 1 & 3/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 3/2 \end{bmatrix} = \begin{bmatrix} 1 & 3/2 \\ 3/2 & 13/4 \end{bmatrix}$$

To find λ_1 and $\lambda_2 = ?$

Let $\lambda^2 - S_1\lambda + S_2 = 0$ be the characteristic equations

of $A^TA = 1 + 13/4 = 17/4$

S_2 = determinant of matrix (A^TA) (det A = ad - bc)

$$S_2 = 1 \times 13/4 - 9/4 = 4/4 = 1$$

Putting values of S_1 & S_2 in above equation

$$\lambda^2 - 17/4 + 1 = 0$$

From solving this equation we got value

of $\lambda = 4, 0.25$

$\lambda_1 = 4, \lambda_2 = 1/4$

also $\sigma_1 = \sqrt{\lambda_1}, \sigma = \sqrt{\lambda_2}$

$\sigma_1 = \sqrt{4}, \sigma_2 = \sqrt{1/4}$

We know that $S = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$

$$\Sigma = \begin{bmatrix} \sqrt{4} & 0 \\ 0 & \sqrt{1/4} \end{bmatrix} \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Here the singular value of A is $\sigma_1 = \sqrt{4}$, $\sigma_2 = \sqrt{1/4}$

Confirmation using matrix factorization multiplication to prove

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 3/2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 4/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix}$$

$$= \begin{bmatrix} 0.447 & -0.894 \\ 0.894 & 0.447 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \\ 0.894 & -0.447 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 \\ 1 & 1.5 \end{bmatrix}$$

Hence $A = U \Sigma U^T$ proved.

6) ii) Outer product form of SVD, by taking the rank-1 approximation for matrix A.

Rank $k = 1$ [We only consider largest singular value $\sigma_1 = \sqrt{4}$]

$$A_k = \sigma_1 U_1 U_1^T$$

$$A_k = \sqrt{4} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

(6)

$$A_k = \begin{bmatrix} 0.4 & 0.8 \\ 0.8 & 1.6 \end{bmatrix}$$

6) iii) To find SVD for A^{-1}

Formulae $A^{-1} = V \Sigma^{-1} U^T$

$$A^{-1} = \begin{bmatrix} -3/2 & 1 \\ 1 & 0 \end{bmatrix} \quad U^T = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

$$V = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \quad \Sigma^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & \sqrt{4} \end{bmatrix}$$

Putting in formulae $A^{-1} = V \Sigma^{-1} U^T$

$$A^{-1} = \begin{bmatrix} -3/2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{1/4} & 0 \\ 0 & \sqrt{4} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} -1.5 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.223 & 1.788 \\ 0.447 & -0.894 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \\ -0.894 & 0.447 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 1.5 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1.49 & 0.99 \\ 0.99 & 0 \end{bmatrix} \approx \begin{bmatrix} -1.5 & 1 \\ 1 & 0 \end{bmatrix}$$

Hence proved

$$\boxed{A^T = U^T \Sigma^{-1} V}$$

6

7) A) Given 2 cluster $m_1 m_2$ with center points

$m_1 = (1,1,1)$ and $m_2 = (1,2,0)$

Data points	$m_1 = 1,1,1$	distance $m_2 = 1\,2\,0$	Cluster
$x_1 = 1,1,1$	0	$\sqrt{2}$	m_1
$x_2 = 1,1,4$	3	4.12	m_1
$x_3 = 5,1,2$	4.12	4.58	m_1
$x_4 = 1,2,0$	$\sqrt{2}$	0	m_2

Formulae to find distance b/w 2 points for m_1:

$$d(p_1,p_2) = \sqrt{(m_{1,1} - x_{i,1})^2 + (m_{1,2} - x_{i,2})^2 + (m_{1,3} - x_{i,3})^2}$$

similarly for m_2

$$d(p_1,p_2) = \sqrt{(m_{2,1} - x_{i,1})^2 + (m_{2,} - x_{i,2})^2 + (m_{2,3} - x_{i,3})^2}$$

Let us find the distance ~~blow 2 points for~~ from m_1 (center) to all x points

$$= \sqrt{(1-1)^2 + (1-1)^2 + (1-1)^2} = 0$$

similarly distance from m_2 to all x_1 points

$$= \sqrt{(1-1)^2 + (2-1)^2 + (0-1)^2} = \sqrt{2}$$

for x_4: $\sqrt{(1-1)^2 + (1-2)^2 + (1-0)^2} = \sqrt{2}$

and $\sqrt{(1-1)^2 + (2-2)^2 + (0-0)^2} = 0$

Since we got $X_1 X_2 X_3$ comes under m_1 cluster & X_4 is m_2 cluster

So we have to find new centroid for $m_1 = ?$ & for m_2 the centroid will be same (since only 1 instance [X_4] comes under it)

New centroid for $m_1 = [7/3, 1, 7/3]$

ie $m_1 \simeq (2.3, 1, 2.33)$ remains same

$m_2 \simeq (1, 2, 0)$

7) b) mean entropy $(c) = \sum_{i=1}^{k} \frac{n_i}{n}$ Entropy (C_i)

where $(C_i) = \sum_{j=1} P_{ij} \log_2 P_{ij}$

Now to find entropy C_1 & C_2

Entropy $C_1 = (2/3 \log 2/3 + 1/3 \log 1/3)$

$= -(-0.929)$

Entropy $(C_1) = 0.929$, entropy $(C_2) = -(1 \log 1)' = 0$

Find mean entropy$(C) = 3/4$ Entropy $(C_1) + 1/4$ Entropy (C_2)

$= 3/4 * 0.929 + 0$

Entropy$(C) = 0.696$

Part 2 Programming assignment

Assignment 1 :

Given us

The initial starting points for the K cluster means can be K randomly selected data points. You should have an option to run the algorithm *r* times from *r* different randomly chosen initializations

```
LOWEST SUM OF SQUARE ERROR FOR RUN [10]:  946.6381021673415
SUM OF SQUARE ERROR RUN [10]:  1046.1569429493627

PLOTTING GRAPH FOR LOWEST SUM OF SQUARE ERROR:  946.6381021673415
```

The plot is included in a separate file called K-Means with screenshots pasted

Assignment 2 :

The main distinction between K-means and C-means is that in C-means, a data point can belong to multiple clusters. The user provides the number of clusters and interactions as input. Random membership weights are assigned to each data point. The membership grade and centroid for each cluster are computed using the fuzzifier parameter m = 2. The membership grade quantifies the degree to which an individual data point can belong to a cluster.

```
Epsilon reached, stopping early
New best square error: 2
All squareErrors sums:
[1196.8595182099502, 1177.3585820963945]
Best squareError sum:
1177.3585820963945
```

The plot is included in a separate file called C-Means with screenshots pasted