SAVEETHA ENGINEERING COLLEGE

AUTONOMOUS

NAAC A GRADE

Accredited by NBA

Affiliated to Anna University | Approved by AICTE

Sl. No.:

Marks	

CIA I Examination –April/May/Nov/Dec 2024

Register No.: 212223080032

Branch & Sem.: Mech (IInd sem)

Date: 10-4-24

Sign of the Invigilator: R. Sh 10/4/24

Subject Code & Subject Name : 19AI410- Introduction to Machine Learning

6. A)
```
Import numpy as np
Import matplotlib.pyplot as plt

x= np.aarray([1,2,3,4,5,6,7,8,9,10])
y= np.array([0,3,5,7,9,11,13,2,4,8])

X_Mean= np.mean(x)
Y_Mean= np.mean(Y)

num= 0
denom=0

For i in range(len(x)):

        num+= (X[i]- X_Mean] (Y[i]- Y_Mean)
        denom+=(X[i]-X-Mean]**2

m= num / denom

b = Y_Mean - X_Mean *m

print (m,b)
```

```
Y_pred = b + X*m

print (Y_pred)

plt.scatter(X, Y, color='red')
plt.plot (X, Y_pred, color='black')
plt.show ()
```

7 B) Logistic Regression is type binary classification statistical method used for classification problems. Based on nature of dependent variable. There are 3 types of Logistic Regression

i) Binary Logistic Regression

ii) Multinomial Logistic Regression.

iii) Ordinal Logistic Regression.

i) Binary Logistic Regression

Binary Logistic Regression is used when dependent variable has 2 categories which represent 2 possible outcomes

Example:

i) For a student to pass (1) or fail (0) no. of hours ~~k~~ he/she studied.

ii) To classify the email as spam (1) or not ~~&~~ spam(0). It depends on independent variable no. of spam words i~~a~~ ~~the~~ and length of email.

ii) Multinomial Logistic Regression.

Multnomi~~cal~~al Logistic Regression is used when dependent variable has two or more undoreded categories. The regression predicts the probablilities of classes. The class with highest probability is chosen as the outcome.

Exomples:

i) To classify cars ~~color~~ primary color such as red, black, blue. Here is no speific order in category. The outcome ~~as~~with color which sells the most.

ii) To classify animial as dog, cat, bird. Here also there is no specfic order.

can decide whether the animal is cat, or dog, or bird.

iii) Ordinal Logistic Regression:

Ordinal Logistic Regression is used when dependent variable has two or more unorded categories. It ~~cumulative~~ predicts the cumulative odds of class to occur. Normally it computes with natural order from lower to higher.

Examples:

i) Predicting ~~t~~ a student's Final grade. Based on independent variable attendance and test scores. We can predict a student's Final grade From low to high.

ii) Predicting a ~~s~~ customer's review. Based on independent variable of product quality and customer service. We can predict a customer's review From low to high.

In all Regression problems. We follow following steps.

i) Data collection; Gather the data with clear dependent and ~~&~~ one or more independent variable.

ii) Data preproccessing: Clean and prepocess the data, handle missing values, encoding categories, splitting data set into training and testing dataset.

iii) Model selection: chose the appropiate Logistic regression based on nature dependent variable

iv) Model training: ~~& model trainning~~ train the Logistic regression with traning data.

v) Model evaluation: Evaluate the model using metrics (F1, accuracy, ~~log~~-likelihood)

vi) Prediction. Use the trained model to predict the data.

To conclude. The various types of Logistic Regression models are used for classification. Based on the nature of dependent variable they might involve simple binary classification, multiple ordered or multiple unordered categories. The choice depends on specificity

8 B(A) From the given data lets calculate gini-index for the age, job_status, owns_house, credit_rating.

Gini index:

i) Age $\qquad G = 1 - \left(\frac{Yes}{N}\right)^2 - \left(\frac{No}{N}\right)^2$

Gini index for young:

$$G(Y) = 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2$$

$$= 1 - (0.4)^2 - (0.6)^2$$

$$= 1 - 0.16 - 0.36$$

$$= 0.48$$

Gini - index for middle:

$$G(M) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$$

$$= 1 - (0.6)^2 - (0.4)^2$$

$$= 1 - 0.36 - 0.16$$

$$= 0.48$$

Gini - index for old.

20/04

$G(O) = 1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2$

$= 1 - (0.8)^2 - (0.2)^2$

$= 1 - 0.64 - 0.04$

$= 0.32$

Total G.I for Age

$G.I(Age) = \frac{5}{15} \times G.I(Y) + \frac{5}{15} \times G(M) + \frac{5}{15} \times G(O)$

$= \frac{\cancel{5}^1}{\cancel{15}_3} \times 0.48 + \frac{\cancel{5}^1}{\cancel{15}_3} \times 0.48 + \frac{\cancel{5}^1}{\cancel{15}_3} \times 0.32$

$= 0.16 + 0.16 + 0.10666$

$= 0.4267$ approx.

ii) Job_status

For False Gini-index:

$G(F) = 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{6}{10}\right)^2$

$= 1 - (0.4)^2 - (0.6)^2$

$= 1 - 0.16 - 0.36)$

$= 0.48$

$G(T) = 1 - \left(\frac{5}{5}\right)^2 - \left(\frac{0}{0}\right)^2$

$= 1 - (1)^2$

$= 1 - 1$

$= 0$

The total G.I for job status.

$G.I(\text{Job\_status}) = \frac{10}{15} \times G(F) + \frac{5}{15} \times G(T)$

$= \frac{2}{3} \times 0.48 + \frac{1}{3} \times 0$

$= 0.32$

iii) OWN_HOUSE

For False Ginai index:

$G(F) = 1 - \left(\frac{3}{9}\right)^2 - \left(\frac{6}{9}\right)^2$

$= 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$

$= 1 - 0.111 - 0.444$

$= 1 - 0.04 - 0.64$

$= 0.32$

For Good Gini index

$G(\text{Good}) = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2$

$= 1 - (0.666)^2 - (0.333)^2$

$= 1 - 0.444 - 0.111$

$= 0.444$ approx.

For excellent Gini-index

$G(\text{Excellent}) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{0}\right)^2$

$= 1 - (1)^2$

$= 1 - 1$

$= 0$

Overall for credit_rating

$G.I = \frac{5}{15} \times GI(\text{Fair}) + \frac{6}{15} \times (G(\text{Good}))$

$+ \frac{4}{15} \times (G.I(\text{excellent}))$

$= 0.444$ approx.

For true Gini-index:

$$G(T) = 1 - \left(\frac{6}{6}\right)^2 - \left(\frac{0}{0}\right)^2$$

$$= 1 - (1)^2$$

$$= 1 - 1$$

$$= 0$$

The total G.I for OWN_HOUSE

$$G.I(OWN\_HOUSE) = \frac{9}{15} \times G(F) + \frac{6}{15} \times G(T)$$

$$= \frac{3}{5} \times 0.444 + \frac{2}{5} \times 0$$

$$= 0.2664$$

IV) On CREDIT-RATING

For Fair

$$G(Fair) = 1 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2$$

$$= 1 - (0.2)^2 - (0.8)^2$$

A Nayak

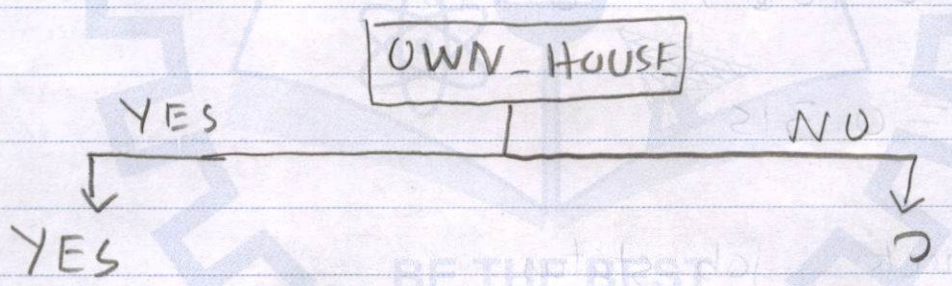
$$= \frac{1}{3} \times 0.32 + \frac{2}{5} \times 0.444 + 0 \times \frac{4}{15}$$

$$= 0.1067 + 0.1776$$

$$= 0.2843$$

We can select own_house since it has the lowest Gini-index

For Loan approval



Lets find gini-index for no cases alone:

Lets check age:

For young:

$$G.I(Y) = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2$$

$$= 1 - 0.0625 - 0.5625$$

$$= 0.375$$

$$G.I(M) = 1 - \left(\frac{0}{0}\right)^2 - \left(\frac{2}{2}\right)^2$$

$$= 1 - 0 - 1$$

$$= 0$$

For old

$$GI(O) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2$$

$$= 0.444 \text{ approx.}$$

Total G.I

$$= \frac{4}{9} \times 0.375 + \frac{2}{9} \times 0 + \frac{3}{9} \times 0.444$$

$$= 0.1667 + 0.148$$

$$= 0.315$$

Let check job-status.

For False Gini-index.

$$G(F) = 1 - \left(\frac{5}{5}\right)^2$$

$$= 0$$

For true Gini-inde

$$G(T) = 1 - \left([illegible]\right)^2$$

Total Gini index

$$= \frac{5}{9} \times G(F) + \frac{4}{9} G(T)$$

There cannot be a lesser value than 0
So we can job status as the
next criteria if person doesn't own house.

~~Total Gini index~~

So decision tree for loan approval
should be like.