Q1) Explain Logistic Regression in Detail

ANS. Logistic Regression is a statistical method used for Binary classification, predicting the probability of an event occurring (Yes/No), (0/1) based on a set of Independent variables.

- Logistic Regression is used for Binary classification where we use sigmoid function, that takes input as independent variables and produces a probability value between 0 and

1. TYPES OF LOGISTIC REGRESSION:-

① Binomial:-

In Binomial Logistic regression, there can be only two possible types of the dependent variables 0 or 1, Pass or Fail.

② Multinomial:-

In Multinomial Logistic Regression, there can be three (3) or more possible unordered types of the dependent variable, such as 'cat', 'dogs' or 'sheep'

③ Ordinal:-

In ordinal Logistic Regression, there can be three (3) or more possible ordered types of the dependent variable, such as 'Low', 'Medium' and 'High'.

In Logistic Regression instead of fitting Regression Line, we fit an "S" shaped logistic function which predicts two maximum values (0 or 1)

Q2) Differentiate between Linear and Logistic Regression.

ANS. Linear Regression is used for continuous outcome variables (e.g., days of hospitilization), and logistic regression is used for categorical variables, such as death.

NO:-	LINEAR REGRESSION:-	LOGISTIC REGRESSION:-
1.	Linear Regression is a supervised regression model	1. Logistic regression is a supervised classification model.
2.	Here, we predict value by an Integer Number	2. Here, we predict the value by 1 or 0.
3.	Here, No Activation Function is Used	3. Here, Activation function is used to convert a linear regression equation to the logistic regression equation
4.	Here, No threshold value is needed.	4. Here, a threshold value is added.
5.	It is based on the least Square estimation (loss Function)	5. It is based on maximum likelihood Estimation.
6.	Provides a continuous output	6. Provides a discrete output
7.	Purpose: To find the best Fitted Line	7. Purpose: To fit the line value to the sigmoid curve
8.	Used For buisness domain Forecasting tasks.	8. Used For classification, image processing, etc.

Q3.) Consider the Binary Classification task with two classes positive and negative. Find out TP, TN, FP, FN, Accuracy, Error rate, Precision, Recall.

→ Total Samples $N = 165$

1. True Positive (TP) = 150
2. False Negative (FN) = 10
3. False Positive (FP) = 20
4. True Negative (TN) = 100

ANS. SOLUTION:-

Metrits

(1.) $$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$ ....(Formula)

$$\text{Accuracy} = \frac{150+100}{150+100+10+20} = \frac{250}{280} = 0.89229$$

Accuracy = 0.89229

(2.) Error Rate (E.R.)

Error Rate = 1 - Accuracy ... (FORMULA)

$$E.R. = 1 - 0.89229$$

$$= 0.1071$$

Error Rate = 0.1071

(3.) Precision.-

Precision = $\frac{TP}{TP+FP}$ .... (FORMULA)

$$\text{Precision} = \frac{150}{150+20} = \frac{150}{170} = 0.8824$$

Precision = 0.8824

(4.) Recall

Recall = $\frac{TP}{TP+FN}$ ... (FORMULA)

$$\text{Recall} = \frac{150}{150+10} = \frac{150}{160} = 0.9375$$

Recall = 0.9375

Q4) Comment on whether the Model is a Best Fit or not based on the calculated values.

ANS. 1. High Accuracy i.e. (89.29%) means that the model is generally making correct predictions for both classes low error rate (10.71%) re-inforces the above metric.

2. Based on the Above performance Metrics. The Model performs well and can be considered a good fit for the Binary classification task.

3. Hence, Yes, the Model Appears to be a Best-Fit based on the given Data.

Q5) Write a Python code for the preprocessing mentioned in step 4 and explain every step in Detail.

ANS. PYTHON CODE:-

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split

df = pd.read_csv ('Social_Media_Adv.csv')
print (df.head())

df = df.dopna()
if 'Gender' in df.columns()
    df['Gender'] = df['Gender'].map ({'Male':0, 'Female': 1F})
x = df[['Age', 'Estimatedsalary']]
y = df['Clicked_on_Ad']
x_train, x_test, y_train, y_test = train_test_split (x,y,
                                    test_size = 0.2, random_state = 42)
scaler = Standard Scaler()
x_train = scaler.Fit_transform (x_train)
x_test = scaler.transform (x_test)
```

⟹ Explaination:- (STEP BY STEP)

1. Import Libraries : For Data handling, pandas, numpy and sklearn Libraries are imported.

2. Load Dataset : Reading the csv File into a Dataframe. (For Eg:- iris dataset)

3. Display Data : Checking what the data look like to plan.

4. Drop the Missing Value.

5. Encode Categorical Variable (Gender) convert string into number for ML Models.

6. Select Feature and Labels : Choose which column to use for prediction (X) and the column to predict (y)

7. Train-test Split : Divides data for training and evaluating the Model.