Lab2_LogisticRegression_Report

What's the difference between polynomial/linear regression and logistic regression?

Explain when you would use each one and how their outputs are different.

Linear and Polynomial Regression

Linear regression models a **linear** relationship between a dependent variable (the output) and one or more independent variables (the inputs). The goal is to find the best-fit straight line that minimizes the distance between the actual data points and the predicted values.

Polynomial regression is an extension of linear regression that models a **non-linear** relationship. Instead of a straight line, it uses a curved line (a polynomial function) to fit the data, which can capture more complex patterns. Both are a type of **regression** analysis, which means it's used to predict a **continuous numerical value**. We can use it when the output we are trying to predict is a continuous value, but not a category or probability.

Logistic Regression

Logistic regression is a statistical method for **classification**, not regression. The primary purpose is to predict the probability of a **categorical outcome**. It models the relationship between the independent variables and the probability of a specific event occurring. It uses a **sigmoid function** to transform the output into a value **between 0 and 1**, which can be interpreted as a **probability**.

1. **Linear Regression**

   The goal is to predict the continuous numeric value (Amount).

   Predict the transaction amount(Amount) from transaction details (V1-V28)

   Dependent variable(output): Amount (The value of the transaction.)

   Independent variables(inputs): All columns **except** `Amount`, `Unnamed: 0`, Class

   Output: [18.49 31.55 41.59 ... -132.56 11.12 20.91] (Single number)

   Metrics: Use **MSE (lower is better)** and **R^2 score (higher is better)**.
   Linear Regression: MSE = 9444.49 , R^2 = 0.85

2. **Polynomial Regression**

   The goal is to predict the continuous numeric value (Amount) with curves.

   Predict the transaction amount(Amount) from transaction details (V1-V28)

   Dependent variable(output): Amount (The value of the transaction.)

   Independent variables(inputs): All columns **except** `Amount`, `Unnamed: 0`, Class

   Output: [7.46 37.13 35.92 ... -27.80 -0.214   15.50] (Single number)

   Metrics: Use **MSE (lower is better)** and **R^2 score (higher is better)**.
   Polynomial Regression: MSE = 13766905.35 , R^2 = -218.44

3. **Logistic Regression**

   The goal is to predict the **probability of a categorical outcome** (Class).

   Predict whether a transaction is **fraud (Class = 1)** or **not fraud (Class = 0)** from transaction details.

   Dependent variable(target): Class 0 = Not Fraud, 1 = Fraud

   Independent variables(features): All columns **except** Class, Unnamed：0

   Output**:** A probability between 0 and 1.

   Metric: confusion_matrix, classification_report, and roc_auc_score

```
[[8535    6]
 [  27  110]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      8541
           1       0.95      0.80      0.87       137

    accuracy                           1.00      8678
   macro avg       0.97      0.90      0.93      8678
weighted avg       1.00      1.00      1.00      8678

0.9828717982902565
```