MMDT0091
Mentor - Ma Nuwai Thet

**Question : What's the difference between polynomial/linear regression and logistic regression?**
**Explain when you would use each one and how their outputs are different.**

In this lab, the practical projects emphasize more on the implementation of Logistic regression. I have added two parts to the existing logistic regression code which are polynomial and linear regression. By doing that, I can interpret the results from these two models and determine where I can use each model.

In implementing the linear regression model, I drop out the columns of class and amount as we are focusing on predicting the amount. Then, I splitted the data to 40 % testing and 60 % training.

For the polynomial regression, I tested with poly degree 2.

According to my code, this is the summary of the results for each model.

```
----------------------------------------
Linear Regression    - MSE: 9,444.49, R²: 0.8495
Polynomial Regression - MSE: 11,933.45, R²: 0.8098
Logistic Regression  - AUC: 0.9829
```

The difference between these three models is that we worked on the linear and polynomial regression to predict the continuous numbers like house prediction. For logistic regression, it is more like predicting the outcome of the situation Probability like will it rain today or not. We did Fraud prediction, our main purpose is to predict the possibility of fraud. We started with numbers 0 and 1 and then converted them to categorical data of yes or no.This is because logistics regression predicts the amount between 0 and 1. **and then, showing off the results of this is fraud = 1 and not fraud = 0.**

**So their main difference would be predicting the  continuous value of the amount and predicting the category of the dataset.**

**For me, I would use linear regression like predicting the amount and then polynomial regression in cases that change over time, not constantly. For logistic regression, I would use it in the case of analyzing probability like is this transaction fraud or not.**

In this case, if we interpret the results from three different models, we can interpret like this.

In linear regression, we can see the average squared error in dollar prediction and explain 85 % of the variance which we can regard as a good prediction model.

In polynomial regression, we got a larger MSE error and also 81 % of variance which we can see is not a very good model with degree 2.

However in the logistic regression, we can see the accuracy of the model as 98% and all predictions are 0 0 0 0 and that can predict the fraud well. In this model, we will have probabilities first and then convert to 0 and 1.

I think their outputs are different based on each input in different models.

**Linear Regression - we have to predict the amount and then our target is the y_amount so we will have continuous money $ amount.**

**Regression Regression - our purpose is also to predict the amount but in curved lines. Then we will have continuous $ values.**

**Logistics Regression - Our purpose is different here. We want to know whether each transaction is fraud or not. So, that's why we have our target y_class of identifying 0 and 1 and then give the output of probabilities between 0 and 1 and then convert to the nearest value of 0 and 1.**