Question 1

Describe one problem in your study area that can be solved by machine learning techniques. Classify your problem in terms of supervised or un- supervised, classification or regression? Explain the unique challenges to the standard machine learning methods, such as logistic regression?

I'm currently studying Computer Science as an undergraduate, but I am interested in machine learning's applications in the stock market. If output is simplified to buy/sell a stock or 1/0, the problem becomes supervised classification. However, even with this simplification, it is still difficult because it is a time series and not a classic plot. Finding the right regression model to fit the time series is also difficult because different attributes of a stock can affect its price such as seasonality and the business cycle. Log loss could function but may not prove to be the right fit for the data due to these constraints.

Question 2

Regarding the squared loss function we covered in the lecture, we sum up the squares of differences between the actual value and the estimated value from model output. This squared loss function is the one most frequently used for continuous output, but it can also be used for classification task. Since it sums up the squares of the differences, it is not robust to outliers. Propose a more robust loss function for classification task? Please define all mathematical notations clearly.

A more robust solution for binary classification is the CE Loss or Cross Entropy Loss. The function is less prone to outliers because rather than squaring the difference and magnifying the prediction, it keeps the difference in perspective with log. Log loss in binary classification can be defined as:

$$L_{CE} = -(y \log(p) + (1 - y) \log(1 - p))$$

and its derivative with respect to weights to produce the gradient can be defined as:

$$\frac{\partial L_{CE}}{\partial w_j} = [\sigma(wx + b) - y]x_j$$

or in a simpler form:

$$\frac{\partial L_{CE}}{\partial w_j} = (\hat{y} - y)X_j$$

The delta function remains simple for calculations, while the loss function becomes robust to outliers.