Predicting Flight Delays

The Problem :

Is it possible to know if your flight will be delayed before it comes up on the departure boards?
In this project we will answer of this question with using Machine Learning (ML) Algorithms that will try to predict if a flight will be delayed or not.

The Data : source : Kaggle

The data contains many features that characterize a flight. We performed a comprehensive analysis of the data starting from One Hot Encoding and more. The full and comprehensive analysis of the data can be found here.

The Models:

With **Logistic regression** and **Neural network** we will try to solve this problem which will be defined as a classification problem where a "0" will correspond to a flight being on time, and a "1" to a flight being delayed.

Model function and Loss function:

Both models use the sigmoid function which maps any real value into another value between [0 ,1].
Our decision boundary will be defined as 0.5 such that :

$$\hat{y} = \begin{cases} 0 & \sigma(z) < 0 \\ 1 & \sigma(z) \geq 0.5 \end{cases}$$

The loss function of the model is the cross entropy function.
Gradient Descent chosen to be the optimization algorithm that finding a local minimum of this function.

The Results:

To illustrate the comparison between the models Neural Network with 2 hidden layers and logistic regression :
We performed plots of graphs that shows the level of accuracy of the models compared to the training and testing sets and in addition shows the minimum loss value obtained by Gradient Descent optimization algorithm.

Logistic Regression	Neural Network