Logistic Regression

What is the question we are asking this week?

- We want to predict which passenger of the titanic is going to survive the ship wrack of the titanic and which passenger is going to die immediately.
- In general, you want to build a model that predicts outcomes for datapoints that it has not seen before. The ultimate goal is to be able to do exactly that.
- You generally give the model a subset of the data for which the outcome is known. What it is supposed to learn

What are the values, y (Outcome variable) can have?

- Dead or Survived (O or 1)

What is information that we are going to use in order to make the predictions?

- Age
- Passenger Class
- All columns...

How does this information help us to predict the outcome?

- The logistic regression uses input data in order to predict a probability of survival for every passenger in the training dataset (involves the sigmoid function and the parameters b and w)
- It will make the predictions
- The loss function tells the model how far it is off from the real data
- The model optimizes the parameters as to minimze the loss function
- Intuitively, in each iteration of finding the parameters, the model looks at how it should tweak b and w in order to reduce the loss

Reference to the notebook!