Machine Learning week 3 video: Simplified Cost Function and Gradient Descent[[1]](#footnote-1)

So my purpose basically is to use gradient descent to get the best parameters for logistic regression

n. I have beforehand defined my histogram function using the sigmoid function for estimation of y, which is either 1 or 0 with binomial distribution.

Now, my purpose ist o use gradient descent to optimize h(x) parameters, e.g to minimize cost function. Minimizing the cost function is done by using gradient descent. The problem however lies with the convexness[[2]](#footnote-2) of the cost function. This is because we have estimated y with a sigmoid function. **To solve such problem, I have used the maximum likelihood estimation** which allows me to replace the cost function with some logs.

The intuition for such estimation derivation can be explained by using these three sources.:

1. Firstly a further explained source from Andrew Ng Machine learning course for real students of Stanford: <http://cs229.stanford.edu/notes/cs229-notes1.pdf>
2. Here is definition of likelihood: <http://stats.stackexchange.com/questions/2641/what-is-the-difference-between-likelihood-and-probability>
3. Maximum likelihood estimation
   1. <https://onlinecourses.science.psu.edu/stat414/node/191>
   2. <https://www.youtube.com/watch?v=Z582V53dfr8#t=427.246711>

# How to check for convergence ( e.g. whether a function is convex or not )

<http://mathworld.wolfram.com/ConvexFunction.html>

Condition Second derivative must be bigger than 0 for every x.

1. https://www.coursera.org/learn/machine-learning/lecture/MtEaZ/simplified-cost-function-and-gradient-descent [↑](#footnote-ref-1)
2. Convex definition: For every random y-s on the function, the line pulled between them is >= y. Therefore only one globaal optimum- in this case: minimum exists. [↑](#footnote-ref-2)