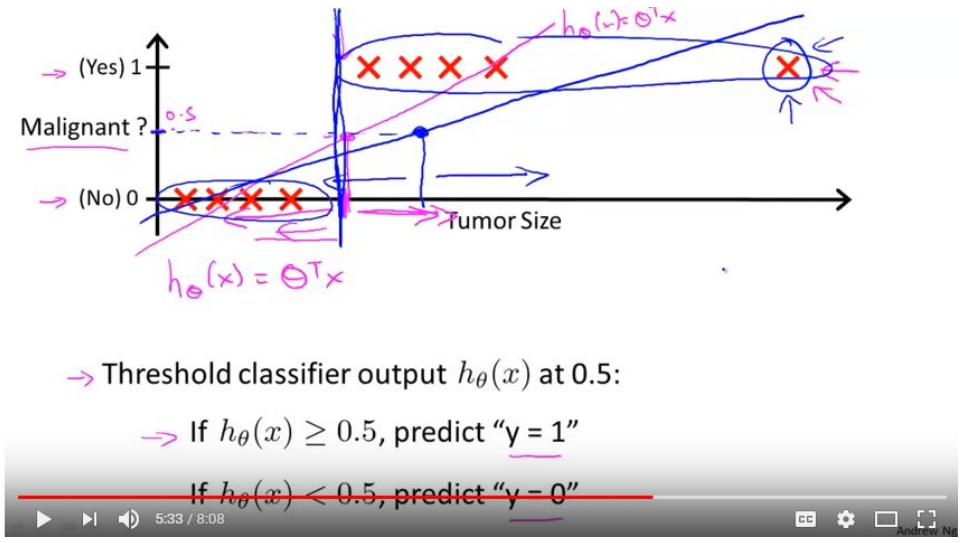
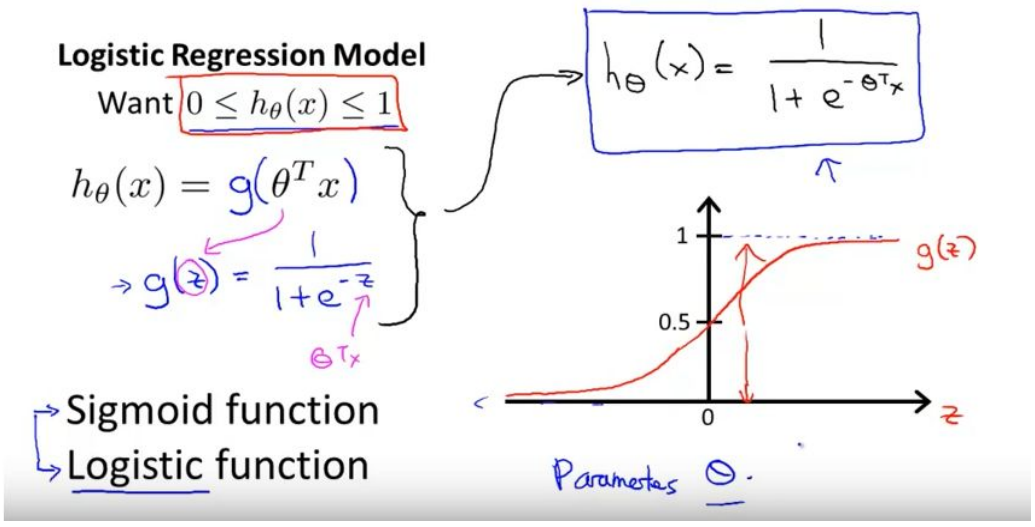
Logistic Regression

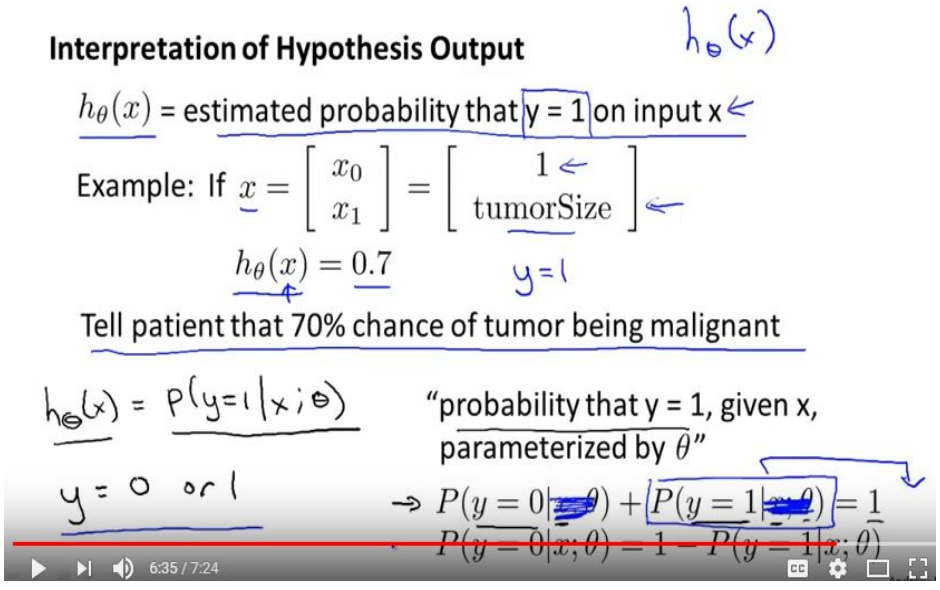
Logistic regression is a classification algorithm that provides the estimated probability of a class(y=1) given a set of features(X).

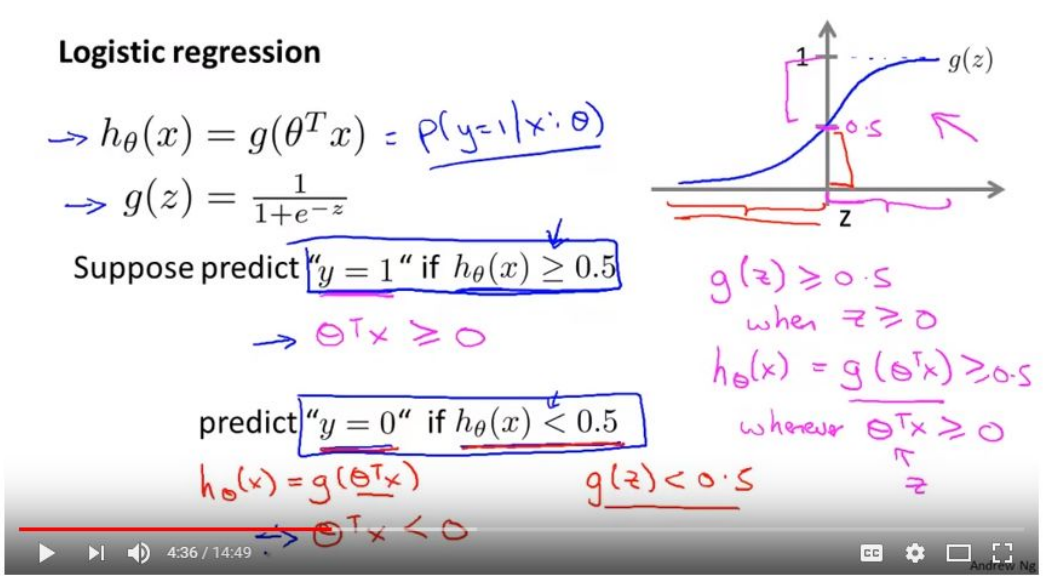
Linear regression can also be used to classify using a threshold value but can fail easily.





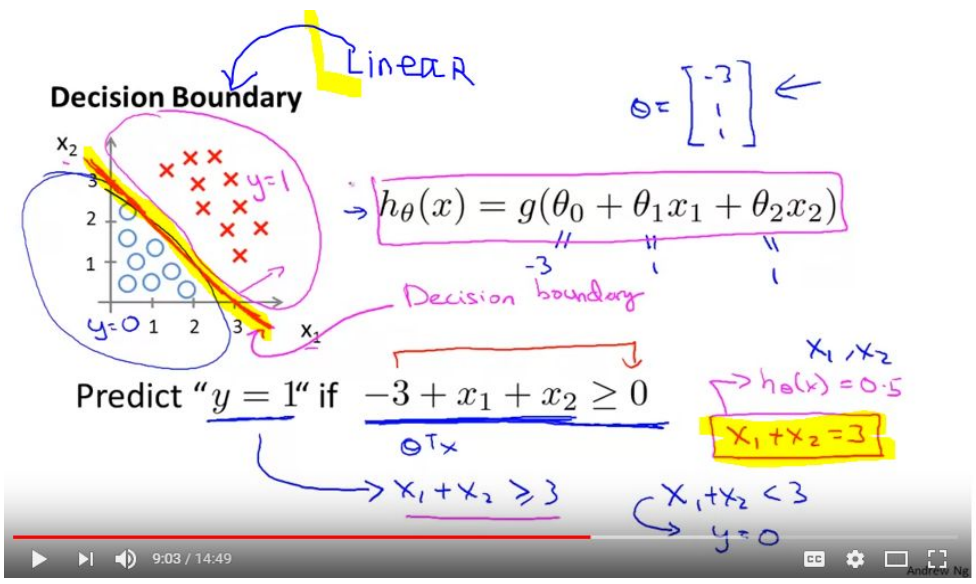
In Logistic regression a nonlinear function g(z) = 1/(1 + e^(-z)), is added to linear regression so that the output of the function is squashed between 0 and 1.

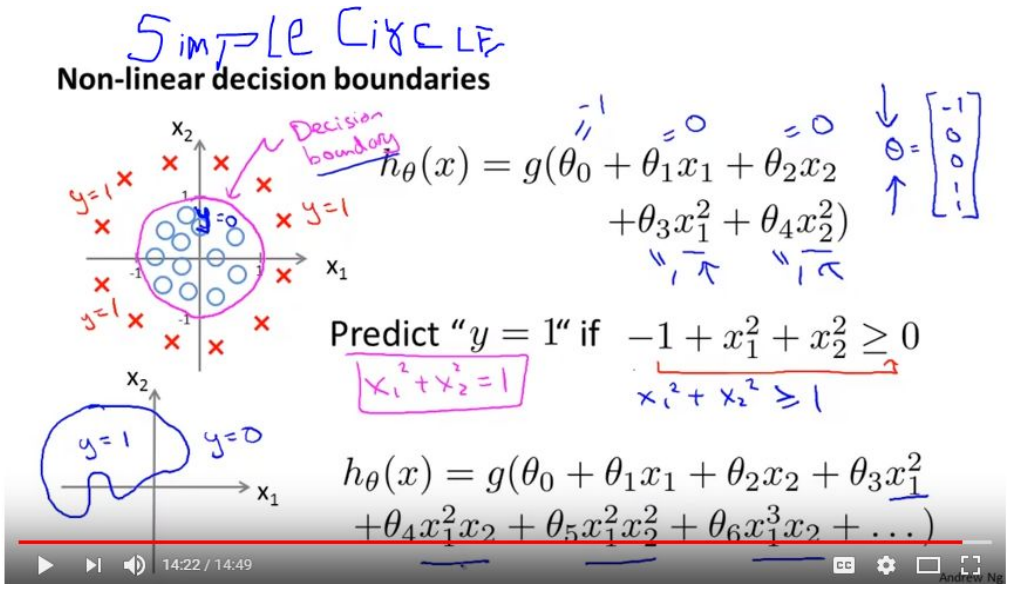




How threshold is selected in logistic regression? So decision boundary LINEAR is easily an outcome of the parameters of the regression model. As we modified Linear regression for complex boundaries like polynomial regression, quadratic and interaction terms we could do the same for the Logistic regression because Logistic regression = g(Linear regression) and here Linear regression can represent GLR model.

Also by looking at the scatter plot or the residuals plot we can say what kind of features transformations and interaction terms can be added so that we can simplify the task for the linear regression. Eg: If there is a circle add square transformation of the features, if hyperbola add square and interaction terms and soon.





Convexity of logistic regression, We can not use Least squares estimate because then we may end up with a non convex cost function with no guarantee of reaching global optima because of many local optimas. An alternative approach is to use log of the expected probability for nonlinear function g.

Because of the smooth convex property of negative log this transformation helps to achieve the convexity for the overall cost function.

Proof: https://math.stackexchange.com/questions/477207/derivative-of-cost-function-for-logistic-regression
Direct Proof: https://stats.stackexchange.com/questions/278771/how-is-the-cost-function-from-logistic-regression-derivated
Sigmoid Proof: https://math.stackexchange.com/questions/78575/derivative-of-sigmoid-function-sigma-x-frac11e-x

One vs All Logistic regression

The logistic function will always produce an S-shaped curve of this form, and so regardless of the value of X, we will obtain a sensible prediction.

The odds ratio: The Logits:
In a logistic regression model, increasing X by one unit changes the log odds by β1.

Maximum Likelihood Estimate:

Logistic Regression is actually a very simple Neural Network