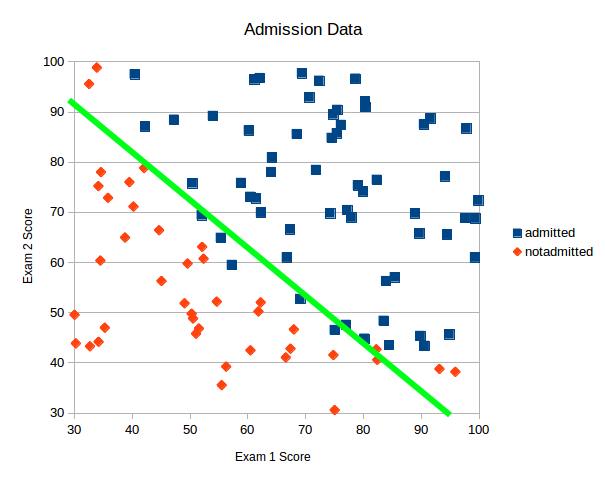
logistic regression

We want to make a boundary, not predictions. Solve a binary classification problem, not a regression problem. You use the features given to train parameters and classify new examples.

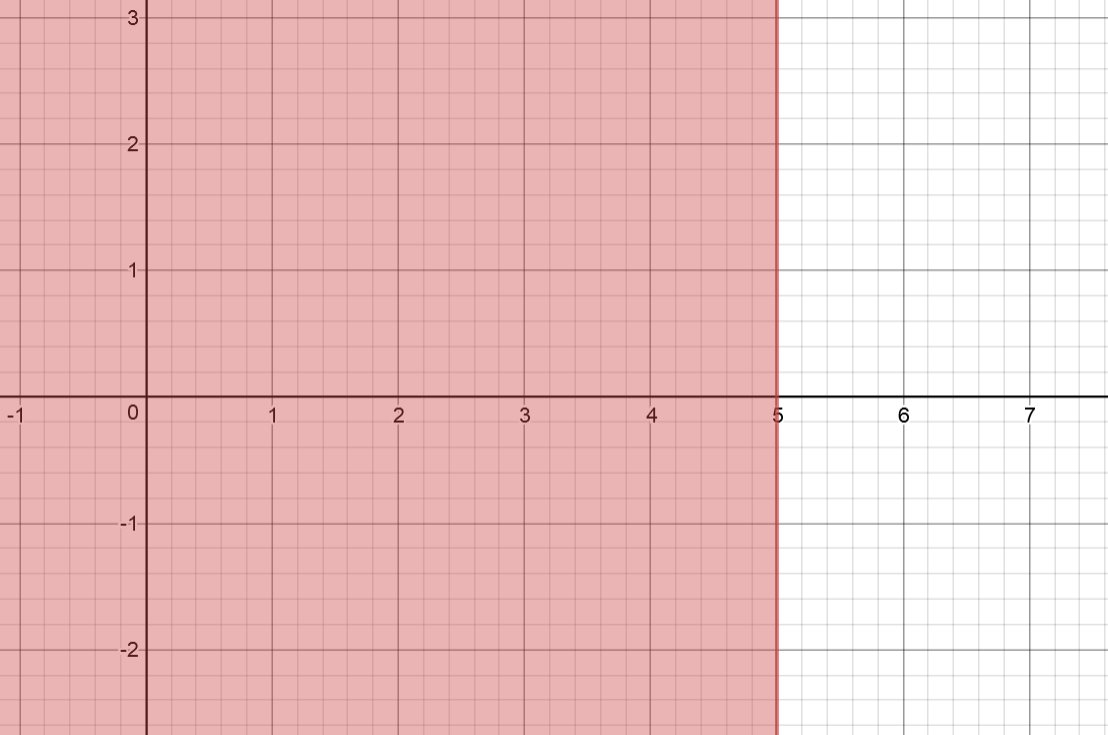
hypothesis function

Intuitively, it also doesn’t make sense for the output of the hypothesis function to take values larger than 1 or smaller than 0 when we know that y ∈ {0, 1}. To fix this, let’s change the form for our hypothesis.

Now the hypothesis function doesn’t output the prediction of y, but the probability that y = 1.

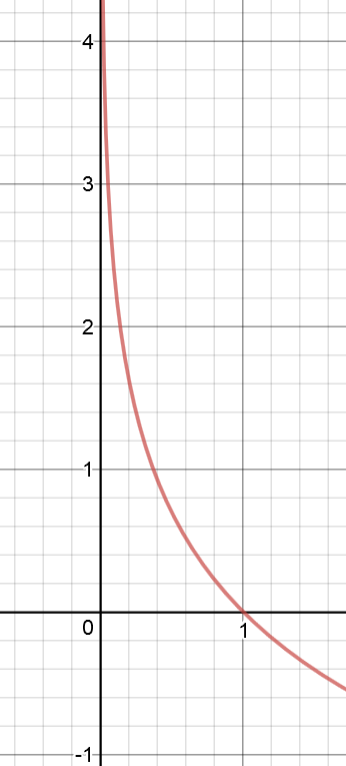
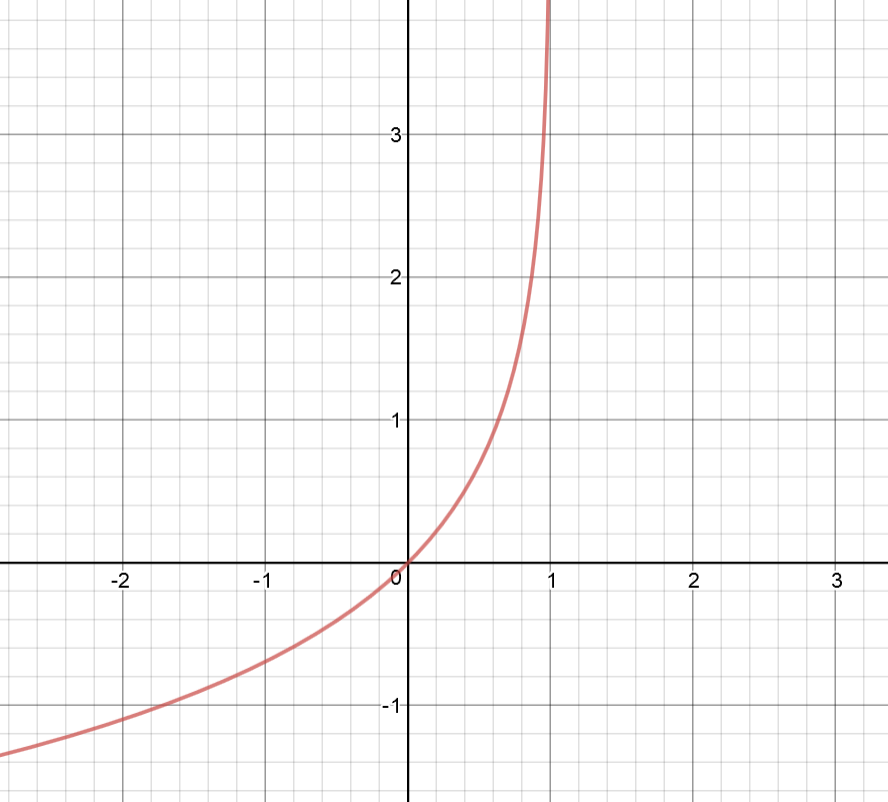
decision boundary

sigmoid function:

example:

cost function

If we use the cost function for linear regression, the cost function will have many local optima and we will not be able to run gradient descent. Therefore, instead of using mean squared error, we will use a method called mean cross-entropy.

if y = 1 if y = 0

As you can see, in both graphs, the cost decreases as the prediction is near the value of y and there is only 1 local optimum, so this cost function is suitable for logistic regression.

For ease of implementation, we can write the cost function in the following way:

if y = 1, the second term cancels out, and otherwise, the first term cancels out.

gradient descent

This is the algorithm for gradient descent. Let’s calculate the partial derivative.

It looks the same with linear regression, but the hypothesis function is different, so it is not the same.

vectorized implementation: