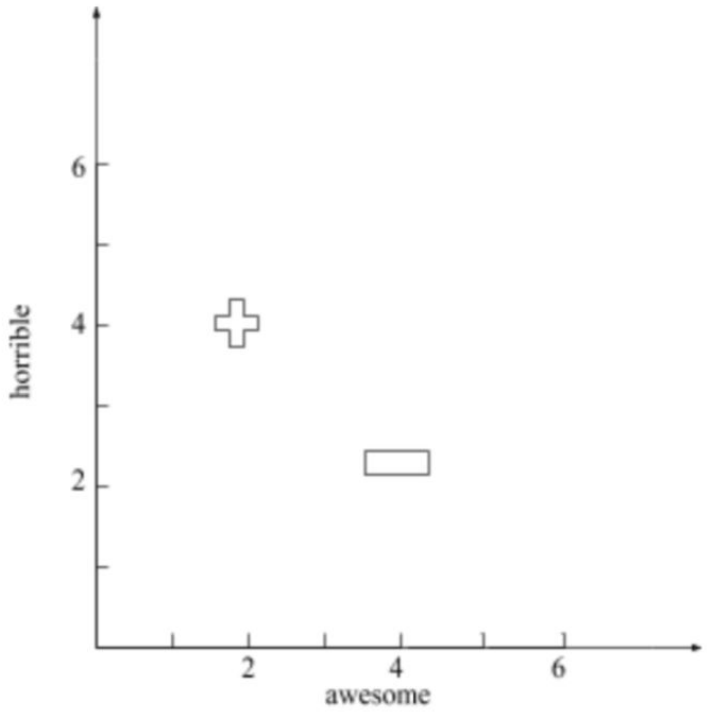
Problem 1

Suppose your work for Amazon and are tasked with classifying product reviews as positive or negative. You aren't very familiar with "big data", so your boss assigns you a small dataset with only two points. Your job is to create a classifier that correctly classifies the two points.

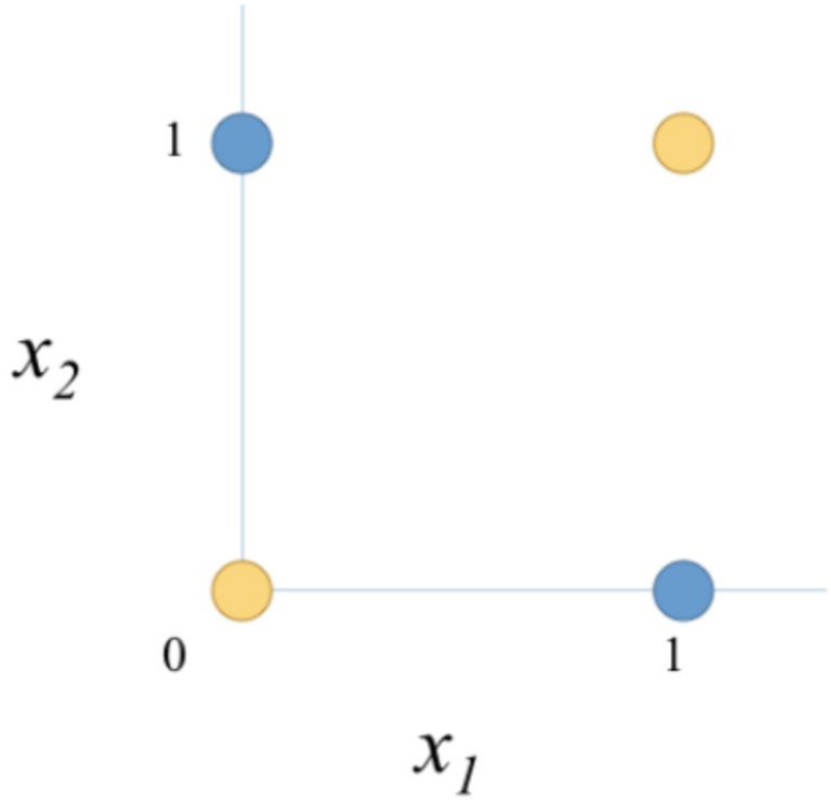


You develop the following procedure: you fix the weight of each count of "horrible" to be 1 and decide to vary the weight of "awesome". Create a graph below that shows how the classification error changes as you move through different weights for "awesome". The graph does not need to have correct values for the coefficients when the error changes; just show how the error changes as you change the weights for "awesome".

How is this graph similar or different to error graphs we have seen so far in this class (i.e. regression)? Can we use the same techniques to optimize classification error as we can RSS? Why or why not?

Problem 2

Consider the dataset below with two data inputs x_1 and x_2 and labels (blue) and (yellow)



A. Can you use logistic regression with linear features Score(x) = $w_0 + w_1x_1 + w_2x_2$ to perfectly classify this dataset? If it is possible, show which weights w can make the correct predictions and if it is not possible, explain why.
B. What if we added a third feature x_1x_2 to the model? If it is possible, show which weights w can make the correct predictions and if it is not possible, explain why.

Problem 3

Consider the dataset in the notebook. Where do we expect the decision boundary to be? What do we expect β_0 and β_1 to be?