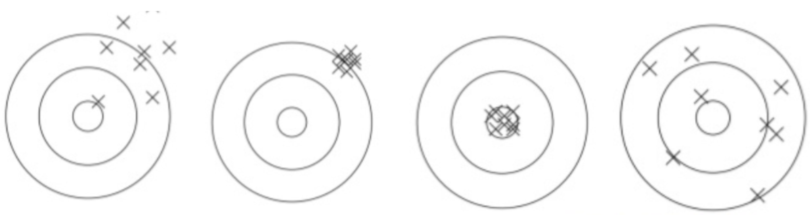
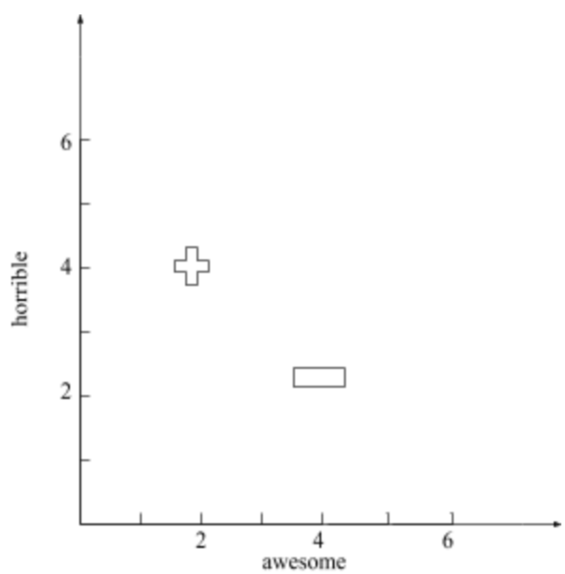
Problem 1: Bias/Variance Review

Included below are four different plots of bias-variance in the form of a bullseye. Consider the true function to be at the center of the target. Label each image's bias/variance as either low or high. Are any of them good predictors of the true function? Justify your response:



Problem 2: Classification Error

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 3: Functions

For binary variable y, and continuous variable x, consider the following model for regression:

$$P(y|x, w) = g(w_0 + w_1 x)$$

For some function g. For this problem, denote $p = P(y|x, w)$.

a. If g is the identity function ($g(v) = v$), what is the range of p?
b. If g is the absolute value function ($g(v) = |v|$), what is the range of p?
c. If g is the logistic function ($g(v) = \frac{1}{1+e^{-v}}$), what is the range of p?

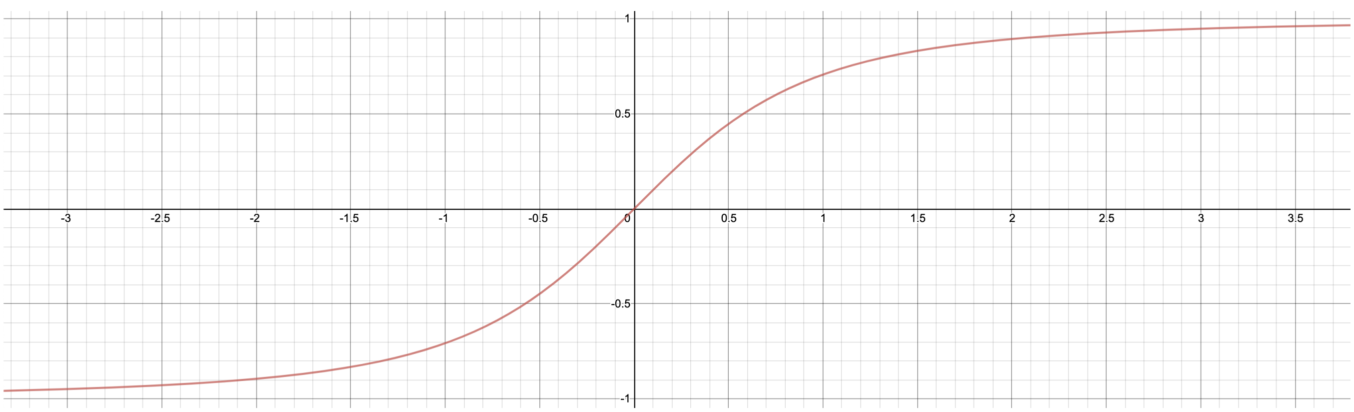
Problem 4: A Different Sigmoid

Suppose we wanted to use a different sigmoid function for our probability model. Suppose we used

$$sigmoid(Score(x)) = \frac{Score(x)}{\sqrt{1 + Score(x)^2}}$$

This function looks like the one shown below and has the following properties:

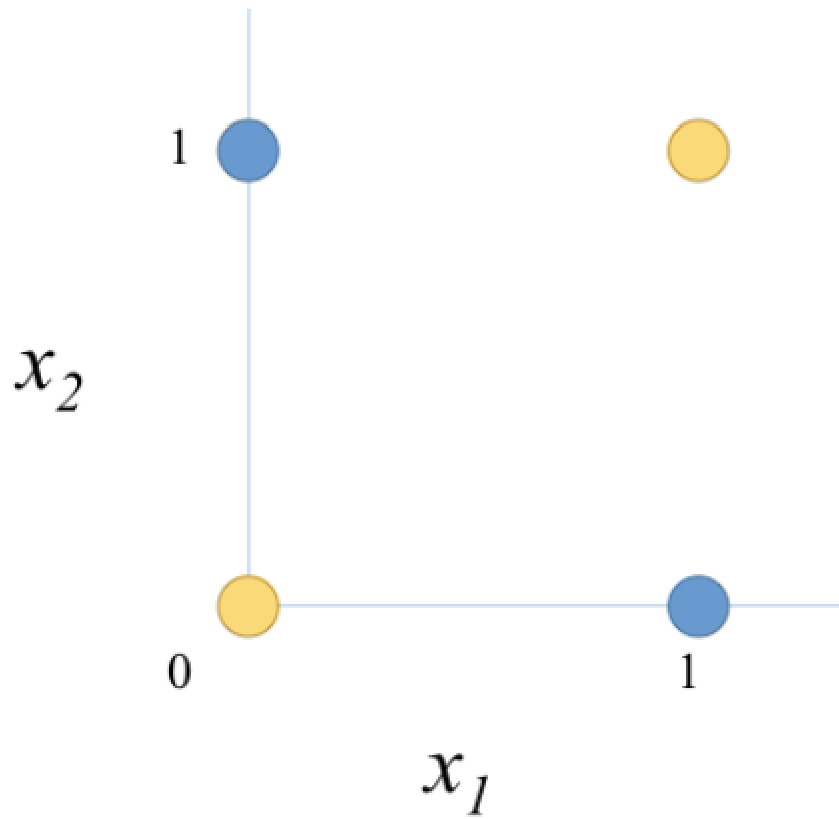
- As $Score(x) \to -\infty$, $sigmoid(Score(x)) \to -1$
- As $Score(x) \to \infty$, $sigmoid(Score(x)) \to 1$
- If $Score(x) = 0$, $sigmoid(Score(x)) = 0$



a. Does this function satisfy our probability assumption of $P(y = +1|x, w) = sigmoid(Score(x))$? Why or why not?
b. Using $P(y = +1|x, w) = sigmoid(Score(x))$ write out the log-likelihood using this sigmoid function. Don't worry whether or not values are valid probabilities, just write out the formula to practice seeing how the parts fit together. You don't have to simplify your answer any more than the slides, but your answer should not contain any $P(y = \ldots)$ and it should have the actual defintion of the sigmoid function instead of $sigmoid$.

Problem 5: XOR

Consider the dataset below with two data inputs x_1 and x_2 and labels $y = +1$ (blue) and $y = -1$ (yellow).



a. Can you use logistic regression with linear features $Score(x) = w_0 + w_1 x_1 + w_2 x_2$ 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 to the model $h_3(x) = x_1 x_2$? If it is possible, show which weights w can make the correct predictions and if it is not possible, explain why.