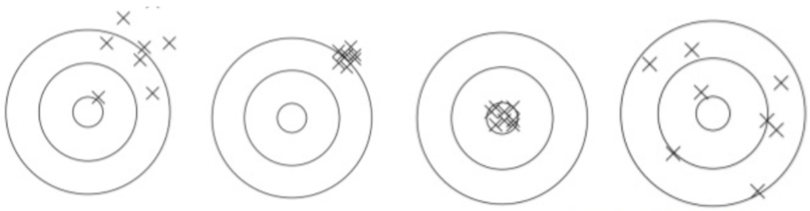
Problem 1: Bias/Variance Review

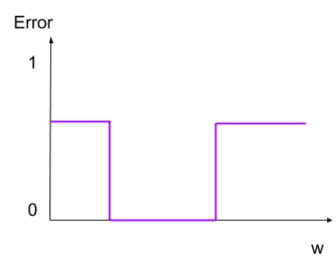


Answer: From left to right

- High Bias + High Variance. The x's are spread around a lot (high variance) and are not centered at the true value (high bias).
- High Bias + Low Variance. The x's are very concentrated (low variance) and are not centered at the true value (high bias).
- Low Bias + Low Variance. The x's are very concentrated (low variance) and are centered at the true value (low bias). Best predictor of true function.
- Low Bias + High Variance. The x's are spread around a lot (high variance) but are centered at the true value (low bias).

Problem 2: Classification Error

Answer:



This graph is similar because it has the general shape of decreasing and then increasing, meaning there is a optimal point. It's different because it has flat sections and is not continuous. We can't use gradient descent here the function is not differentiable everywhere and has 0 slope in most places; this means we would never take a step since there is no indication of up or down.

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?

Answer:

a. $[-\infty, \infty]$
b. $[0, \infty]$
c. $[0, 1]$

Problem 4: A Different Sigmoid

a. Does this function satisfy our probability assumption of $P(y = +1|x, w) = sigmoid(Score(x))$? Why or why not?

Answer: No. This function can output values in the range $[-1, 1]$ which has values that are not valid probabilities.

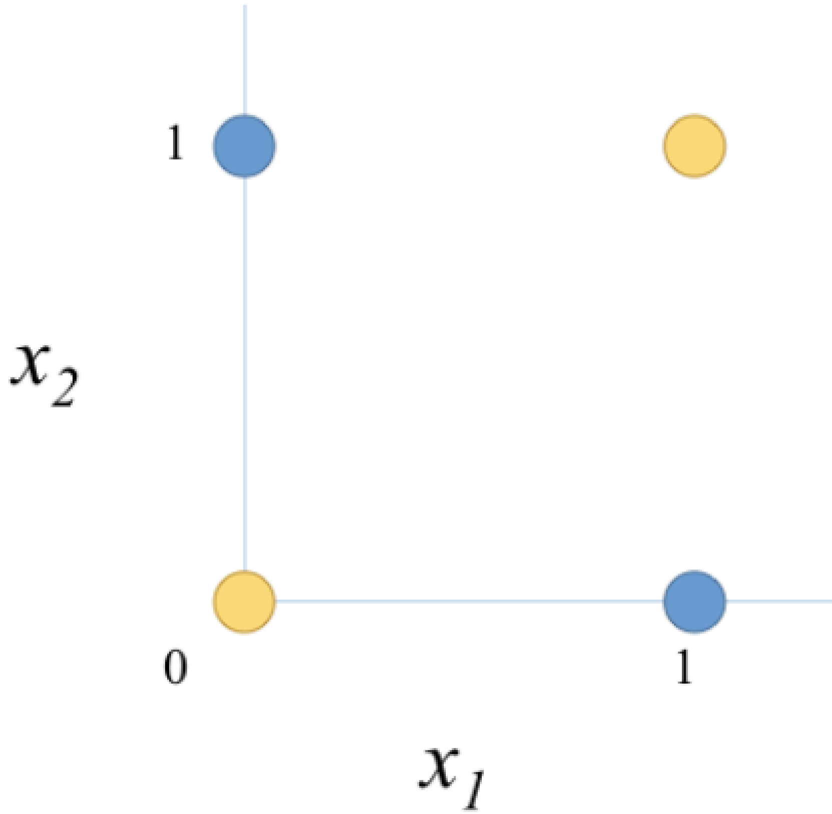
b. Using $P(y = +1|x, w) = sigmoid(Score(x))$ write out the log-likelihood using this sigmoid function. Don't worry whether or not values are valid probabilities, just write out the formula to practice seeing how the parts fit together. You don't have to simplify your answer any more than the slides,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

Answer:

$$\ln(\ell(w)) = \sum_{i=1:y_i=+1}^{n} \ln\left(\frac{w^T h(x_i)}{\sqrt{1 + (w^T h(x_i))^2}}\right) + \sum_{i=1:y_i=-1}^{n} \ln\left(1 - \frac{w^T h(x_i)}{\sqrt{1 + (w^T h(x_i))^2}}\right)$$

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.

Answer: No, this model can only learn linear decision boundaries and it is not possible to draw a line that separates the blue points and the yellow points.

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.

Answer: One possibel answer $w_0 = -0.5$, $w_1 = 1$, $w_2 = 1$, $w_3 = -2$.