Machine Learning Homework 3 Solutions

Written Problems

1. Machine learning methods can be viewed as function estimators. Consider the logical functions AND, OR, and XOR. Using a signed representation for Boolean variables, where input and output variables are in $\{+1, -1\}$, these functions are defined as

$$\text{AND}(x_1, x_2) = \begin{cases} +1 & \text{if } x_1 = +1 \wedge x_2 = +1 \\ -1 & \text{otherwise} \end{cases} \tag{1}$$

$$\text{OR}(x_1, x_2) = \begin{cases} +1 & \text{if } x_1 = +1 \\ +1 & \text{if } x_2 = +1 \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

$$\text{XOR}(x_1, x_2) = \begin{cases} +1 & \text{if } x_1 = +1 \wedge x_2 = -1 \\ +1 & \text{if } x_2 = +1 \wedge x_1 = -1 \\ -1 & \text{otherwise} \end{cases} \tag{3}$$

(a) (6 points) Which of these three logical functions can be expressed as a linear classifier of the form

$$f(\boldsymbol{x}; \boldsymbol{w}) = \text{sign}(w_1 x_1 + w_2 x_2 + b), \tag{4}$$

and show weights w_1, w_2, and bias values b that mimic these logical functions. Which of these functions cannot be expressed as a linear classifier, and why not?

Solution: One strategy for determining whether these functions are linearly classifiable is to draw them. Each function has four possible input points at $(1, 1), (1, -1), (-1, 1), (-1, -1)$ with different target classes. From plots, one can imagine whether a line can separate the $+1$ outputs from the -1 outputs.

The AND function can be mimicked with the weights $w_1 = w_2 = 1$ and bias $b = -1$ (or any $b \in (-2, 0)$). With these weights, only when both x_1 and x_2 are 1 does the weight score cancel out the negative bias.

The OR function can be mimicked with the weights $w_1 = w_2 = 1$ and the bias $b = 1$ (or any $b \in (0, 2)$). If either of the inputs is 1, then their sum is at least 0. Only when both inputs are -1 will their sum be less than 1.

The XOR function cannot be mimicked with a linear classifier. The output of XOR is positive in the upper left and bottom right quadrants and negative elsewhere, so no single line can isolate the positive cases.

(b) (6 points) For any logical functions that cannot be expressed as a linear classifier, show how and with what weights it can be expressed as a two-layered perceptron of the form

$$f(\boldsymbol{x}; \boldsymbol{w}) = \text{sign}(\boldsymbol{w}^{\text{out}\top} \boldsymbol{h} + b^{\text{out}}) \tag{5}$$

$$\boldsymbol{h} = [h_1, h_2]^\top \tag{6}$$

$$h_1 = \text{sign}(\boldsymbol{w}^{(1)\top} \boldsymbol{x} + b^{(1)}) \tag{7}$$

$$h_2 = \text{sign}(\boldsymbol{w}^{(2)\top} \boldsymbol{x} + b^{(2)}) \tag{8}$$

For this problem, you must specify three weight vectors $(\boldsymbol{w}^{\text{out}}, \boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)})$ and three biases $(b^{\text{out}}, b^{(1)}, b^{(2)})$.

Solution: The XOR function is positive when the number of positive inputs is exactly 1 and the number of negative inputs is exactly 1. Conversely, XOR is negative when either the number of positive inputs is 2 and when the number of negative inputs is 2. We can check for these conditions in the first layer and use the second layer to check that both conditions are false:

$$w_1^{(1)} = w_2^{(1)} = 1, \;\; b^{(1)} = 1 \tag{9}$$

$$w_1^{(2)} = w_2^{(2)} = -1, \;\; b^{(2)} = 1 \tag{10}$$

$$w_1^{\text{out}} = w_2^{\text{out}} = -1, \;\; b^{\text{out}} = 1 \tag{11}$$

With these parameters, h_1 is equivalent to $\text{AND}(x_1, x_2)$ and h_2 is equivalent to $\text{AND}(\neg x_1, \neg x_2)$. The final output checks that neither of these cases is true, which means there is exactly one +1 input and one -1 input, just as in the XOR function.