Calculate Perceptron Update Including Bias

Not all data sets that are linearly separable can be separated with a hyperplane that passes through the origin. In order for the Perceptron to converge in this scenario, you must account for an offset value (or bias) that will adjust the position of your hyperplane in either a positive or negative direction.

The exercise below is identical to the one on the previous page, only this time we are going to account for our bias term ("**b**") and see how that changes the calculations.

Exercise 1

Consider the following two-point 2D data set:

- Positive class (+1): $(1, 3)$
- Negative class (-1): $(-1, 4)$

To account for bias, we are going to "absorb" our **b** value into the data by adding a dimension to each data point and setting the value of that dimension to 1. Thus, our 2D data set becomes a 3D data set as follows:

- Positive class (+1): $(1, 3, 1)$
- Negative class (-1): $(-1, 4, 1)$

Because our data is now 3 dimensional, our weight vector must be as well, and this new dimension in **w** will represent **b.** When we report the final **w** and **b** values at the end of our algorithm, we will report **w** as our final weight vector excluding the last dimension, and we will report **b** as just the value of the last dimension in our weight vector. Since positive points are added to **w** and negative points are subtracted from **w**, our **b** adjustment will equal either +1 or -1 based on the label.

Starting with $\mathbf{w}_{0+b} = (0, 0, 0)$, how many updates will you have to perform until convergence? Write down the sequence of each updated $\mathbf{w}_t$ $([\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n])$ by iterating the data points in the order:$[(1, 3, 1), (-1, 4, 1), (1, 3, 1), (-1, 4, 1), \ldots]$.

When you think you know the answers, click the images to reveal the solutions

Click to reveal the answer

Answer: As before, the perceptron algorithm stops when for all points $(\mathbf{x}_i, y_i)$ in the dataset D,
$$y_i \left(\mathbf{w}^\top \mathbf{x}_i\right) > 0.$$

Iteration	Value of	Misclassified?	If yes, update

1	$(0,0,0)$	$1 \cdot \left((0,0,0)^\top (1,3,1)\right) = 0 \leq 0$	Yes	$\mathbf{w_1} \leftarrow (0,0,0) +$
2	$(1,3,1)$	$-1 \cdot \left((1,3,1)^\top (-1,4,1)\right) = -12 \leq 0$	Yes	$\mathbf{w_2} \leftarrow (1,3,1) - ($
3	$(2,-1,0)$	$1 \cdot \left((2,-1,0)^\top (1,3,1)\right) = -1 \leq 0$	Yes	$\mathbf{w_3} \leftarrow (2,-1,0)$
4	$(3,2,1)$	$-1 \cdot \left((3,2,1)^\top (-1,4,1)\right) = -6 \leq 0$	Yes	$\mathbf{w_4} \leftarrow (3,2,1) - ($
5	$(4,-2,0)$	$1 \cdot \left((4,-2,0)^\top (1,3,1)\right) = -2 \leq 0$	Yes	$\mathbf{w_5} \leftarrow (4,-2,0)$
6	$(5,1,1)$	$-1 \cdot \left((5,1,1)^\top (-1,4,1)\right) = 0 \leq 0$	**YES***	$\mathbf{w_6} \leftarrow (5,1,1) - ($
7	$(6,-3,0)$	$1 \cdot \left((6,-3,0)^\top (1,3,1)\right) = -3 \leq 0$	Yes	$\mathbf{w_7} \leftarrow (6,-3,0) +$
8	$(7,0,1)$	$-1 \cdot \left((7,0,1)^\top (-1,4,1)\right) = 6 > 0$	No	
9	$(7,0,1)$	$1 \cdot \left((7,0,1)^\top (1,3,1)\right) = 8 > 0$	No	

Note: This is where the previous exercise converged, whereas using the same data but accounting for bias forces additional steps.

Therefore, when we accounted for bias there were **7** updates to $\mathbf{w}$:
$$[(0,0,0),(1,3,1),(2,-1,0),(3,2,1),(4,-2,0),(5,1,1),(6,-3,0),(7,0,1)]$$

Our final weight vector **w** = (7,0) and our final **b** value = 1.