Perceptron Why not initializing weights to be zeros

Tuesday, December 13, 2022 2:42 PM

$$w_j \leftarrow w_j + \Delta w_j$$

$$\Delta w_j = \eta\,(y^{(i)} - \hat{y}^{(i)})\,x_j^{(i)}$$

If all w_j's are set to 0 in the beginning, then after the 1st update:

$$w_j = w_j^{(old)} + \Delta w_j = 0 + \Delta w_j = \eta\,(y^{(1)} - \hat{y}^{(1)})\,x_j^{(1)}$$

After the 2nd update:

$$w_j = \eta\,(y^{(1)} - \hat{y}^{(1)})\,x_j^{(1)} + \eta\,(y^{(2)} - \hat{y}^{(2)})\,x_j^{(2)}$$

$$\vdots$$

Finally: $$w_j = \eta \sum_{i=1}^{N} (y^{(i)} - \hat{y}^{(i)})\,x_j^{(i)} = \eta\, w_j^*$$

Similarly, $$b = \eta \sum_{i=1}^{N} (y^{(i)} - \hat{y}^{(i)}) = \eta\, b^*$$

Then $$z = w^T x + b = \eta\, w^{*T} x + \eta\, b = \eta\,(w^{*T} x + b)$$

$\Rightarrow$ η does not affect the sign of z

$\Rightarrow$ No matter what η we use, we get the same decision boundary. ($z = 0$)