If you choose zero initial weights, then the perceptron algorithm's learning rate η has no influence on a neuron's predicted class lable.

This is because the decision function used in the perceptron algorithm depends only on a sign of z:

$$\varphi(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

Consider a perceptron algorithm with a single neuron. (I will write $x.y$ for the vector product $x^T y$ to avoid double superscripts).

- pick an initial weight vector $w^{(0)}$
- plug in the first input vector $x^{(1)}$ and predict the class lable

$$\hat{y}^{(1)} = \varphi(w^{(0)} . x^{(1)})$$

This gives the weight update:

$$\Delta w^{(1)} = \eta (y^{(1)} - \hat{y}^{(1)}) x^{(1)},$$

where $\eta \in (0,1)$ is the learning rate and $y^{(1)}$ is the true class lable. the new weights are: $w^{(1)} = w^{(0)} + \Delta w^{(1)}$

- Similarly, plug in our the input vector $x^{(2)}$ and predict the class lable $\hat{y}^{(2)} = \varphi(w^{(1)} . x^{(2)})$

This gives the weight update: $\Delta w^{(2)} = \eta(y^{(2)} - \hat{y}^{(2)}) x^{(2)}$ where $y^{(2)}$ is the true class lable.

Notice that $\Delta w^{(2)}$ implicitly depends on $\Delta w^{(1)}$ which is turn implicitly depends on $w^{(0)}$. Now I will unravel these dependencies by plugging in:

$$\Delta w^{(2)} = \eta (y^{(2)} - \hat{y}^{(2)}) x^{(2)} = \eta (y^{(2)} - \varphi(w^{(1)} . x^{(2)})) x^{(2)}$$

$$= \eta (y^{(2)} - \varphi((w^{(0)} + \Delta w^{(1)}) . x^{(2)})) x^{(2)}$$

$$= \eta (y^{(2)} - \varphi((w^{(0)} + \eta (y^{(1)} - \hat{y}^{(1)}) x^{(1)}) . x^{(2)})) x^{(2)}$$

$$= \eta (y^{(2)} - \varphi((w^{(0)} + \eta (y^{(1)} - \varphi(w^{(0)} . x^{(1)}) . x^{(2)})) x^{(2)}$$

Scanned with CamScanner

Now suppose we have initialized with zero weights:

$w^{(0)} = 0$. Since $0 \cdot x^{(1)} = 0$ & $\varphi(0) = 1$ the last line in this calculation simplifies to $\Delta w^{(2)} = \eta(y^{(2)} - \varphi(\underbrace{\eta(y^{(1)} - 1)\, x^{(1)} \cdot x^{(2)}}))\, x^{(2)}$

Since $\eta > 0$, it does not change the sign of $(y^{(1)} - 1)\, x^{(1)} \cdot x^{(2)}$. But the sign is all that matters to the function φ. So we can simply remove η from the function argument without changing the result!

$\Delta w^{(2)} = \eta(y^{(2)} - \varphi(y^{(2)} - \varphi((y^{(1)} - 1)\, x^{(1)} \cdot x^{(2)})\, x^{(2)}$

The same arguments holds for $\Delta w^{(3)}, \Delta w^{(4)}, \ldots$

It follows that "$\Delta w^{(i)} = \eta * [$sth that does not depend on $\eta]$" for every $i = 1, 2, \ldots$ the parameter η affects only the scale of the weight vector, not the direction.

For nonzero initial weights this is different. Based on the last line of the calculation of $\Delta w^{(2)}$:

$$\varphi((w^{(0)} + \eta(y^{(1)} - \varphi(w^{(0)} \cdot x^{(1)}))x^{(1)})\, x^{(2)})$$

If $w^{(0)}$ is nonzero, then η may affect the sign of the function argument, and therefore the predicted class table.

Scanned with CamScanner