Perceptron Convergence

David Ricardo Pedraza Silva

April 10, 2021

Suppose that we have Θ^*, the ideal set of weights. If this is the case ${\Theta^*}^T x_n$ has the same sign as y_n for it is clasified appropriately, thus guaranteeing $y_n({\Theta^*}^T x_n) > 0$. If $r = min_{1 \leq n \leq N} y_n({\Theta^*}^T x_n)$, then $r > 0$.

We know that

$$\Theta(t+1) = \Theta(t) + y(t)x(t)$$

For intance:

$$\Theta^T(t+1)\Theta^* = \Theta^T(t)\Theta^* + y(t)({\Theta^*}^T x(t))$$

We know that $r \leq y(t)({\Theta^*}^T x(t))$, then:

$$\Theta^T(t+1)\Theta^* \geq \Theta^T(t)\Theta^* + r$$

As $\Theta^T(1)\Theta^* \geq \Theta^T(0)\Theta^* + r$ then $\Theta^T(1)\Theta^* \geq r$.

Suppose that $\Theta^T(t)\Theta^* \geq tr$. If that is the case, as

$$\Theta^T(t+1)\Theta^* = \Theta^T(t)\Theta^* + y(t)({\Theta^*}^T x(t))$$

thus

$$\Theta^T(t+1)\Theta^* \geq tr + y(t)({\Theta^*}^T x(t)) \geq tr + r$$

That is:

$$\Theta^T(t+1)\Theta^* \geq (t+1)r$$

Let us use $\Theta(t+1) = \Theta(t) + y(t)x(t)$ again. We see that:

$$|\Theta(t+1)|^2 = |\theta(t)|^2 + 2y(t)(\theta(t)x(t)) + y(t)^2|x(t)|^2$$

We know that $\Theta(t)$ missclassifies $y(t)$, then $2y(t)(\Theta(t)x(t)) \leq 0$. For instance:

$$|\Theta(t+1)|^2 \leq |\Theta(t)|^2 + |x(t)|^2$$

Let $R = max_{1 \leq n \leq N}|x_n|$. Then

$$|\Theta(t+1)|^2 \leq |\Theta(t)|^2 + R^2$$

As $\Theta(0) = 0$; $|\Theta(1)|^2 \leq R^2$. Suppose that for some t we get $|\Theta(t)|^2 \leq tR^2$, then as $\Theta(t+1) = \Theta(t) + y(t)x(t)$ and $|\Theta(t+1)|^2 = |\Theta(t)|^2 + |x(t)|^2$ we have $|\Theta(t+1)|^2 \leq tR^2 + R^2$. That is:

$$|\Theta(t+1)|^2 \leq (t+1)R^2$$

We now consider $\frac{\Theta^T(t)\Theta^*}{|\Theta(t)|}$. Note that

$$\frac{\Theta^T(t)\Theta^*}{|\Theta(t)|} \geq \frac{\Theta^T(t)\Theta^*}{\sqrt{t}R}$$

We conclude that:

$$\frac{\Theta^T(t)\Theta^*}{|\Theta(t)|} \geq \frac{\sqrt{t}r}{R}$$

We now simplify in order to isolate t

$$t \leq \frac{|\Theta^*|^2|\Theta(t)|^2R^2}{|\Theta(t)|^2r^2}$$

$$t \leq \frac{|\Theta^*|^2R^2}{r^2}$$

We have bounded t in terms of the data and Θ^* thus proving that this learning model converges, as long as we have a fine set of training points and said set is separable.