perceptron algorithm:

$y = +1 \rightarrow w^T x > 0$

$y = -1 \rightarrow w^T x < 0$

} Correct $\Leftrightarrow y w^T x > 0$

$y = +1 \rightarrow w = w + x$

$y = -1 \rightarrow w = w - x$

} Traing / wrong $\Leftrightarrow w = w + yx$

$$\exists w^* \text{ s.t. } \forall (x,y) \in D \quad y w^T x > 0$$

We are scaling w^* s.t. $\|w^*\| = 1$

and $\forall i \; \|x_i\| \leq 1$

We are assuming missclassification so

$$y w^T x \leq 0$$

Lets start with looking at how $w^T w^*$ changes for each update.

It is the same as $(w + yx)^T w^* = w^T w^* + y w^{*T} x \geq w^T w^* + \gamma$

Golden boy so > 0

$\geq \gamma$

Margin (length from boundry to point)

Defined $\gamma = \min_{(x,y) \in D} |x^T w^*| > 0$

what is the distance to the closest point

We have found that when we make an update to w in relation with w^*: $(w + yx)^T w^*$, it grows by gamma: $w^T w^* + \gamma$

Now let's look at $w^T w$

$$\rightarrow \quad (w + yx)^T (w + yx) = w^T w + 2\underbrace{yw^T x}_{<0} + \underbrace{y^2 x^T x}_{1 \leq}$$

$$\leq w^T w + 1.$$

After M updates:

cs inequality

$$M\gamma \leq w^T w^* = |w^T w^*| \leq \|w\| \|w^*\| = \|w\| = \sqrt{w^T w}$$

$$\leq \sqrt{M}$$

$$\rightarrow \quad M \leq \frac{1}{\gamma^2}$$

We have proved that the algorithm converges after a finite set of iterations "M" dependant on the distance to the closest point.