Proof of Perceptron Convergence Theorem

- $\theta^0 = 0$: Initial state of hypothesis
- $\theta^{(k)}$: Hypothesis after k mistakes
- θ^* : Correct hypothesis

- We want the angle between θ^k (current separator) and θ^* (good separator) to decrease
  ↳ Small angle ⇒ Big cosine
- Cosine of angle between them:

$$\cos\left(\theta^{(k)}, \theta^*\right) = \frac{\theta^{(k)} \cdot \theta^*}{\|\theta^{(k)}\| \|\theta^*\|} \quad \text{(Definition of dot product)}$$

$$\cos\left(\theta^{(k)}, \theta^*\right) = \frac{\theta^{(k)} \cdot \theta^*}{\|\theta^*\|} \cdot \frac{1}{\|\theta^k\|}$$

- Focusing on the first term, we assume that the k^{th} mistake occurs on the i^{th} example $\left(x^{(i)}, y^{(i)}\right)$.

$$\frac{\theta^{(k)} \cdot \theta^*}{\|\theta^*\|} = \frac{\left(\theta^{(k-1)} + y^{(i)} x^{(i)}\right) \cdot \theta^*}{\|\theta^*\|}$$

$$= \frac{\theta^{(k-1)} \cdot \theta^*}{\|\theta^*\|} + \frac{y^{(i)} x^{(i)} \cdot \theta^*}{\|\theta^*\|} \quad \text{Margin}$$

$$\geq \frac{\theta^{(k-1)} \cdot \theta^*}{\|\theta^*\|} + \gamma$$

Recursion!

$$\geq k\gamma$$ (started from 0 and iterated k times, each time there is a γ term)

- Focusing on the 2nd term, we look at $||\theta^{(k)}||^2$ for simplicity

$$||\theta^{(k)}||^2 = ||\theta^{(k-1)} + y^{(i)} x^{(i)}||^2$$

$$= ||\theta^{(k-1)}||^2 + \underbrace{2 y^{(i)} \theta^{(k-1)} x^{(i)}}_{\text{Classified incorrectly so negative}} + \underbrace{||x^{(i)}||^2}_{\text{Bounded by } R^2} \, y^{(i)2}$$

Always 1 bc $y^{(i)} \in [-1, 1]$

$$\leq ||\theta^{(k-1)}||^2 + R^2$$

Recursion!

$$\leq kR^2$$

$$\therefore \frac{1}{||\theta^{(k)}||} \geq \frac{1}{\sqrt{k}\,R}$$

- Then,

$$\cos(\theta^{(k)}, \theta^*) = \frac{\theta^{(k)} \cdot \theta^*}{||\theta^*||} \cdot \frac{1}{||\theta^k||}$$

$$\geq k\gamma \cdot \frac{1}{\sqrt{k}\,R}$$

$$= \sqrt{k} \cdot \frac{\gamma}{R}$$

Max value of cosine

$$1 \geq \sqrt{k}\,\frac{\gamma}{R}$$

$$k \leq \left(\frac{R}{\gamma}\right)^2$$