Perceptron Convergence Proof:

1) If dataset is separable, there is a margin.

margin: γ

best direction: w^*

2) $\langle w^*, w^{(k)} \rangle = \langle w^*, w^{(k-1)} + yx \rangle$

$= \langle w^*, w^{(k-1)} \rangle + y \langle w^*, x \rangle$

$\geq \langle w^*, w^{(k-1)} \rangle + \gamma$

3) $\|w^{(k)}\|^2 = \|w^{(k-1)} + yx\|^2$ $\quad (\|x\|^2 = \langle x, x \rangle)$

$= \|w^{(k-1)}\|^2 + y^2 \|x\|^2 + 2y \langle w^{(k-1)}, x \rangle$

$\leq \|w^{(k-1)}\|^2 + 1 + 0$

4) $\langle w^*, w^{(k)} \rangle \geq k\gamma$ ⓐ

$\|w^{(k)}\|^2 \leq k$ ⓑ

$$\sqrt{k} \geq \|w^{(k)}\| \geq \langle w^*, w^{(k)} \rangle \geq k\gamma$$

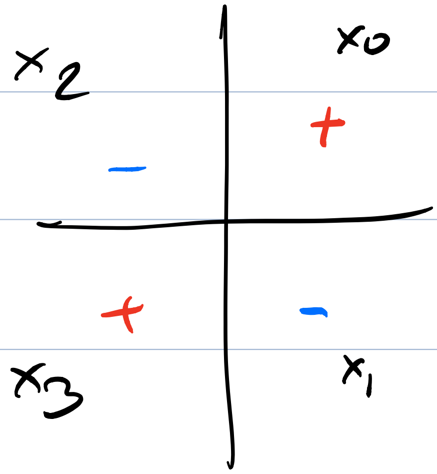
ⓑ ⓐ

$$\|w^*\| = 1$$

$$\sqrt{k} \geq k\gamma$$

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

$$k \leq \frac{1}{\gamma^2} \qquad \square$$



$w: 0, 0, 2$ $b: 2$

x		$x^0 x^1$	y	a
1	1	1	1	4
1	-1	-1	-1	0
-1	1	-1	-1	0
-1	-1	1	1	4

(What about 3 dimensions? D?)

(what happens to the efficiency in vetting the perceptron updates?)

(what about K-NN?)

(what about deep networks)

You can solve the XOR problem in <u>2</u> layers!

Averaged Perceptron

(Assume bias term in xs)

We want:

$$\sum_k c^k w^k \qquad \left(\text{dropping } \frac{1}{c} \text{ term}\right)$$

We have:

$$cw - \left(c^1 x^1 + (c^1 + c^2) x^2 + \dots + \left(\sum c^k\right) x^k\right) :=$$

$$cw - u$$

$$w^k = \sum_i^k x^i$$

$$\sum_{K} c^K w^K = c^1 x^1 + c^2 x^1 + \cdots + c^K x^1$$
$$+ c^2 x^2 + \cdots + c^K x^2$$
$$\ddots$$
$$c^K x^K$$

$$u = c^1 x^1 + c^1 x^2 + \cdots + c^1 x^K$$
$$c^2 x^2 + \cdots + c^2 x^K$$
$$\ddots$$
$$+ c^K x^K$$

$$c\,w = \left(\sum_i^K c^i\right)\left(\sum_i^K x^i\right)$$

$$\begin{aligned}
= {} & c^1 x^1 + c^1 x^2 + \cdots + c^1 x^K \\
+ {} & c^2 x^1 + c^2 x^2 + \cdots + c^2 x^K \\
+ {} & \quad \vdots \qquad\quad \vdots \qquad\quad \ddots \qquad\quad \vdots \\
+ {} & c^K x^1 + c^K x^2 + \cdots + c^K x^K
\end{aligned}$$

$$\left(\langle (1,2), (3,4) \rangle + 1\right)^2 =$$

$$(1\cdot 3 + 2\cdot 4 + 1)^2 =$$

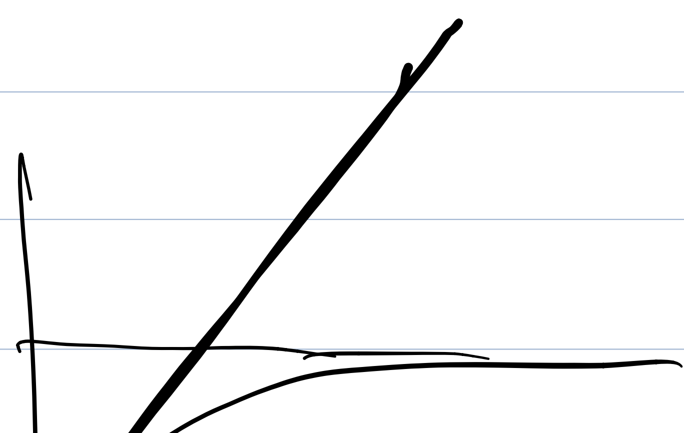
$$1\cdot 3\cdot 1 + 2\cdot 4\cdot 1 + 1\cdot 1$$

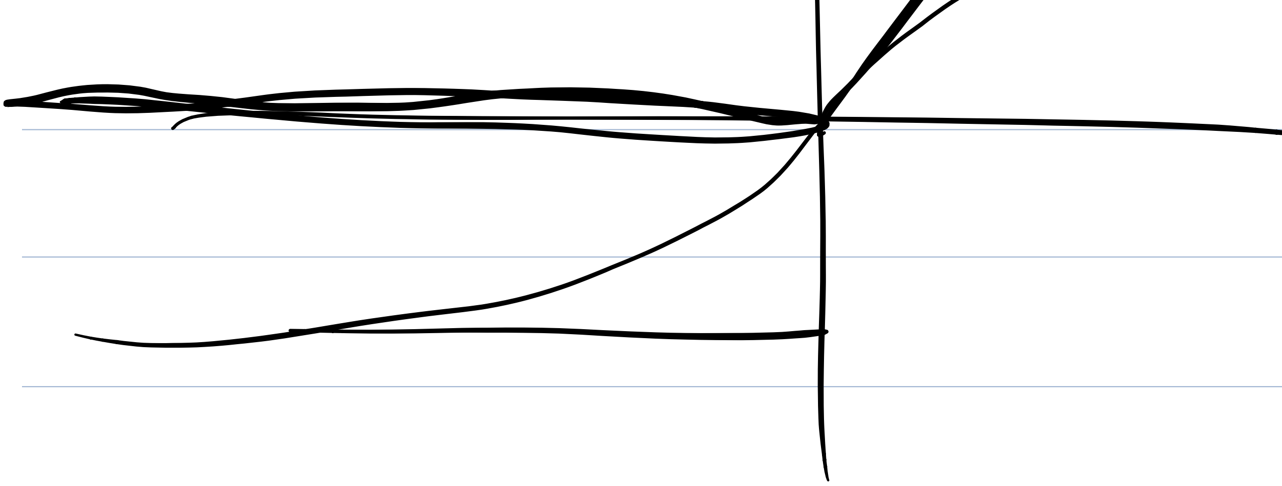
$$1\cdot 3\cdot 2\cdot 4 + 2\cdot 4\cdot 2\cdot 4 + 1\cdot 2\cdot 4$$

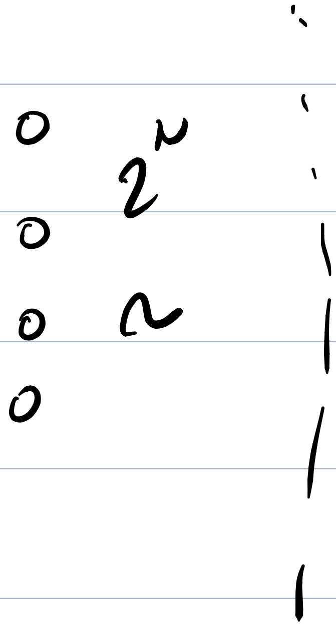
$$1\cdot 3\cdot 1\cdot 3 + 1\cdot 3\cdot 2\cdot 4 + 1\cdot 3\cdot 1$$

$$\left(\langle \tilde{x}, \sum w_i x_i \rangle + 1 \right)^2 = \text{"real"} \langle \cdot , \cdot \rangle$$

$$\left(\left(\sum_i \langle \tilde{x}, x_i \rangle w_i \right) + 1 \right)^2 =$$







Perceptron

XOR