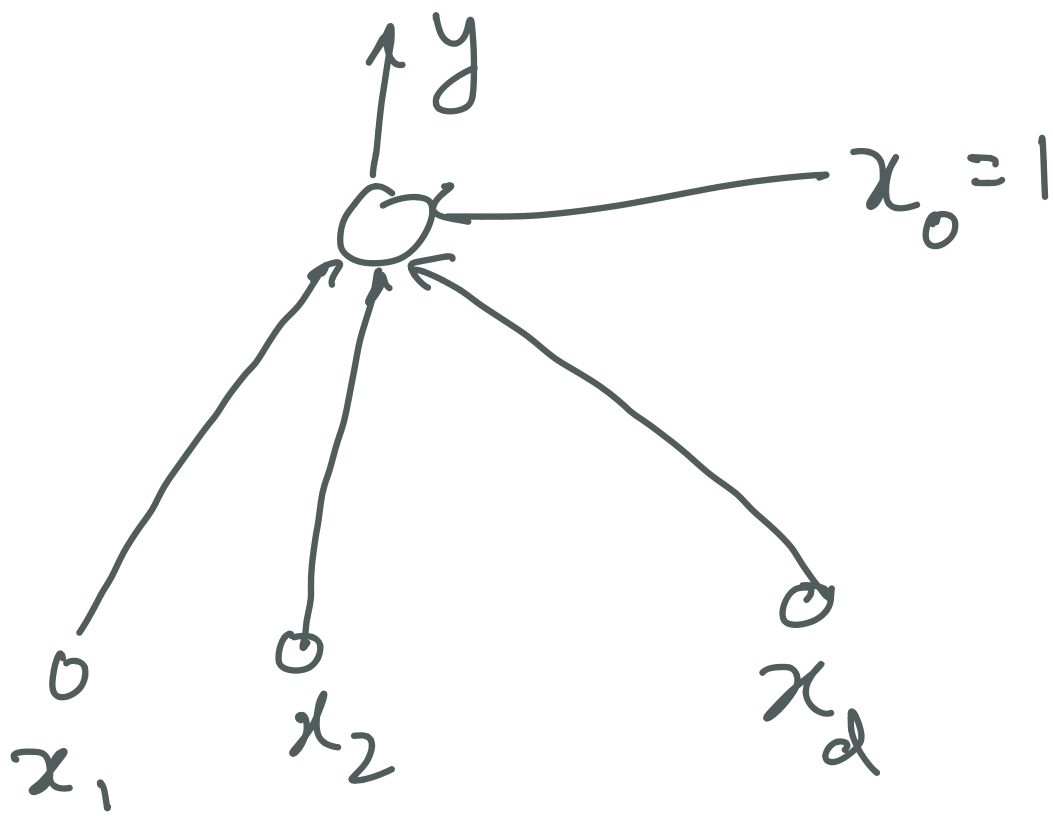
References

1. Pattern Classification

Duda, Hart, Stork

(Artificial Neural Networks

McCulloch-Pitts Neuron



$$y = \text{sign}\left(\sum_{j=0}^{d} w_j x_j\right)$$

$$\mathcal{D} = \left\{ (x^{(i)}, y^{(i)}) \;\middle|\; \begin{array}{c} x^{(i)} \in \mathcal{X},\ y^{(i)} \in \{-1, 1\} \\ i \in [N] \end{array} \right\}$$

Assume that there exists w^* such that

$$\text{sign}\left((w^*)^T x^{(i)}\right) = y^{(i)}.$$

$i \in [N]$.

$$\|x^{(i)}\| \leq R$$

$$w^{(1)} = 0$$

$$i = 1, \ldots, N$$

$$\Rightarrow y^{(i)} \left(w^{*T} x^{(i)}\right) > 0$$

$$\gamma = \min_i \frac{y^{(i)} (w^*)^T x^{(i)}}{\|w^*\|}.$$

$w^{(k)}$ be the current estimate.

Let $(x^{(l)}, y^{(l)}) \in \mathcal{D}$ such that

$$\text{sign}\left(\left(w^{(k)}\right)^T x^{(l)}\right) \neq y_l$$

$$\Rightarrow \; y^{(l)}\left(w^{(k)T} x^{(l)}\right) < 0$$

$$w^{(k+1)} = \begin{cases} w^{(k)} + y^{(l)} x^{(l)} & [\text{Update}] \\ w^{(k)} & \text{otherwise} \end{cases}$$

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

$$= \left(w^{(k+1)} - w^{(k)}\right)^T \left(w^{(k+1)} + w^{(k)}\right)$$

$$= \left(y^{(l)} x^{(l)}\right)^T \left(2w^{(k)} + y^{(l)} x^{(l)}\right)$$

$$= 2\, y^{(l)}\, w^{(k)T} x^{(l)} + \|x^{(l)}\|^2$$

$$\therefore \|w^{(k+1)}\|^2 - \|w^{(k)}\|^2 \leq \|x^{(l)}\|^2 \leq R^2$$

$$y^{(l)}\, w^{(k)T} x^{(l)} < 0$$

$$(w^*)^T \left(w^{(k+1)} - w^{(k)}\right)$$

$$= (w^*)^T \left(y^{(l)} x^{(l)}\right)$$

$$\geq \gamma \|w^*\|$$

Let M be the Number of updates till there are no mistakes

After M updates

$$\sum_{k=1}^{M} (w^*)^T \left(w^{(k+1)} - w^{(k)}\right)$$

$$\geq M\gamma \|w^*\|$$

1

$$w^{*T}\left(w^{(M+1)} - w^{(1)}\right) \geq M\gamma \|w^*\|$$

If we $w^{(1)} = 0$

$$(w^*)^T \left(w^{(M+1)}\right) \geq M\gamma \|w^*\|$$

$$M\gamma \leq \|w^{(M+1)}\|$$ (Because of Cauchy Schwartz)

$$\|w^{(M+1)}\|^2 \leq MR^2$$

$$M^2\gamma^2 \leq MR^2$$

$$M \leq \frac{R^2}{\gamma^2} = \frac{R^2 \|w^*\|^2}{\min_i y^{(i)}\left(w^{*T} x^{(i)}\right)}$$

Adapt the proof to $w^{(1)} \neq 0$

Perceptron $\rightarrow$ DHS.