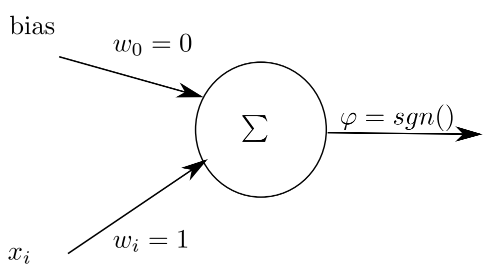
ML II: Exercise 2

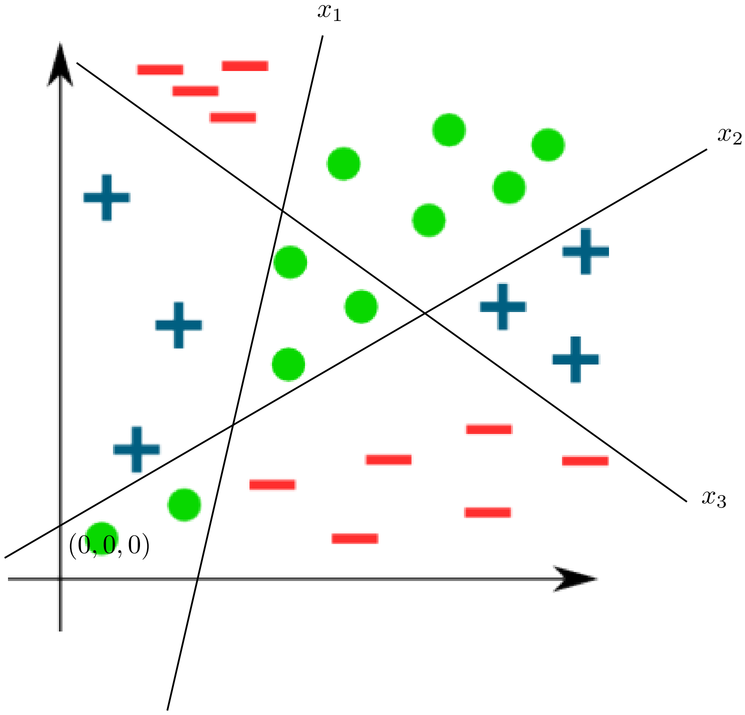
Tobias ..., Philipp...., Lina Gundelwein

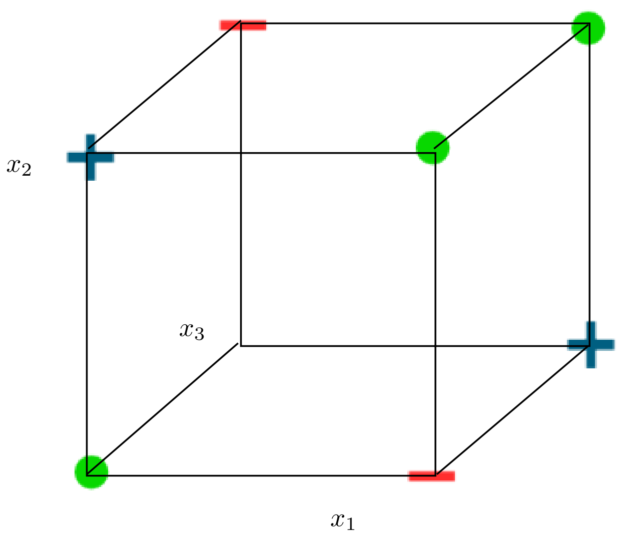
May 15, 2016

Simple Networks



Generalization to higher dimensions: Draw hyperdimensional decision boundaries each dividing the input space into two subspaces. The number of boundaries relates to the dimension of the hypercube.





Linear Activation Function

Using any linear activation function $\varphi(x) = x$ leads to the following outputs:

$$z_1 = \varphi(B_1 z_0) = B_1 z_0$$
$$z_2 = B_2 z_1 = B_2 B_1 z_0$$

This could be replaced by a single layer with the parameters $B = B_1 B_2$.

Weight Decay

1.

$$\begin{aligned} Loss(w) &= Loss_0(w) + \frac{\lambda}{2N} w^\tau w \\ \frac{\partial Loss}{\partial w} &= \frac{\partial}{\partial w} Loss_0 + \frac{\lambda}{N} w \\ w &= w - \tau \frac{\partial}{\partial w} Loss_0 - \tau \frac{\lambda}{N} w \\ &= (1 - \frac{\tau\lambda}{N}) w - \tau \frac{\partial}{\partial w} Loss_0 \\ \rightarrow \epsilon &= \frac{\tau\lambda}{N} \end{aligned}$$

2. The weight decays in proportion to its size. Thus, larger weights are penalized and weights with a small magnitude are preferred which avoids overfitting.

3.

$$\begin{aligned} Loss(w) &= Loss_0(w) + \frac{\lambda}{2N} |w| \\ \frac{\partial Loss}{\partial w} &= \frac{\partial}{\partial w} Loss_0 + \frac{\lambda}{N} sgn(w) \\ w &= w - \tau \frac{\partial}{\partial w} Loss_0 - \tau \frac{\lambda}{N} sgn(w) \end{aligned}$$

4. Since the biases are fixed and representing the offset, not the curvature of the model, the regularization has little effect on them.