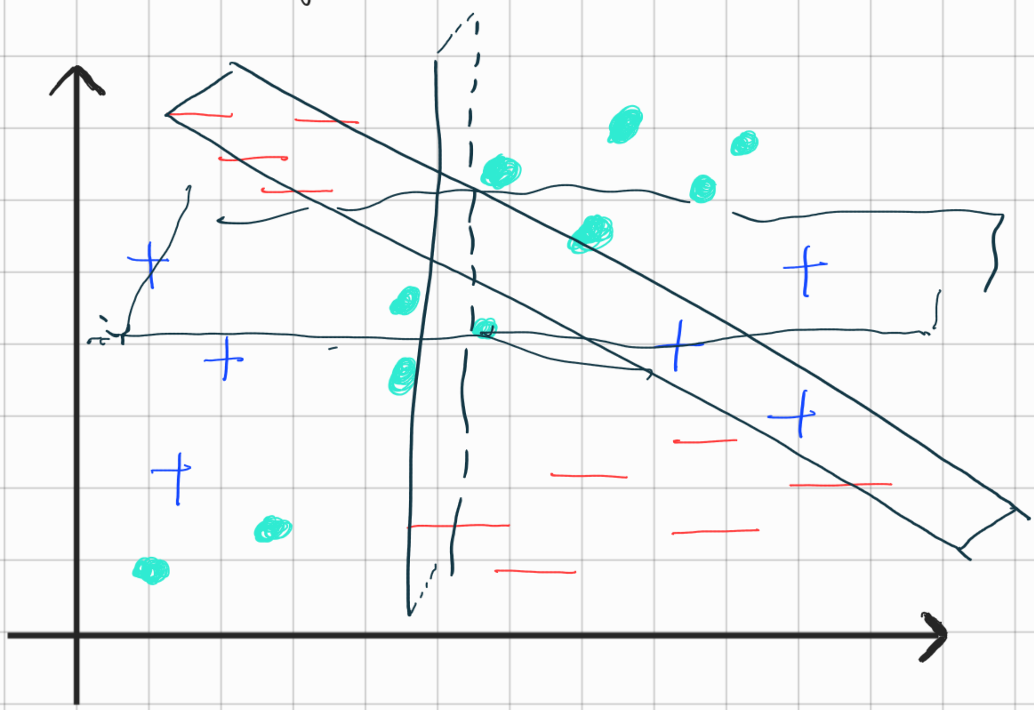
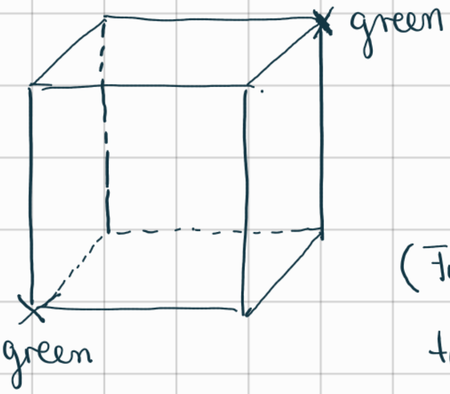
1. Hand-Crafted Network



The general idea in this case is to use two corners of the $\{0,1\}^3$ hypercube for one of the classes present here. In this case:



(Fig 1: For instance green takes these two corners)

The first layer of the neural network classifies the data as the depicted Y_i in the specific region of the corner.

Class 1 → $\{(1,0,0), (0,1,1)\}$
Class 2 → $\{(0,1,0), (1,0,1)\}$
Class 3 → $\{(0,0,1), (1,1,0)\}$

2. Linear Activation Function

$$z_0 = X$$
$$\bar{z}_\ell = z_{\ell-1} \cdot B_\ell + b_\ell$$
$$z_\ell = \phi_\ell(\bar{z}_\ell)$$

If ϕ_ℓ is the identity function then any network with depth $L > 1$ is equivalent to a 1-layer neural network.

Proof:

If ϕ_ℓ is the identity then

$$\begin{aligned} z_\ell &= \phi_\ell(\bar{z}_\ell) \\ &= id(\bar{z}_\ell) \\ &= id(z_{\ell-1} \cdot B_\ell + b_\ell) \\ &= z_{\ell-1} \cdot B_\ell + b_\ell \end{aligned}$$

$$\begin{aligned} z_3 &= z_2 \cdot B_3 + b_3 \\ &= (z_1 \cdot B_2 + b_2) \cdot B_3 + b_3 \\ &= ((z_0 \cdot B_1 + b_1) \cdot B_2 + b_2) \cdot B_3 + b_3 \\ &= (X B_1 B_2 + b_1 B_2 + b_2) \cdot B_3 + b_3 \\ &= X B_1 B_2 B_3 + \\ & \quad b_1 B_2 B_3 + b_2 B_3 + b_3 \end{aligned}$$

Let ℓ be greater than 1

$$\begin{aligned} z_\ell &= z_{\ell-1} \cdot B_\ell + b_\ell \\ &= X \cdot \prod_{i=1}^{\ell} B_i + \sum_{i=1}^{\ell} b_i \prod_{j=1}^{\ell-1} B_{j+1} \end{aligned}$$

Now we define $B := \prod_{i=1}^{\ell} B_\ell$ and $b := \sum_{i=1}^{\ell} b_j \prod_{j=1}^{\ell-1} B_{j+1}$

$$z_\ell = X \cdot B + b$$

This is the same expression as
$$\begin{aligned} z_1 &= z_0 \cdot B_1 + b_1 \\ &= X \cdot B_1 + b_1 \end{aligned}$$

So in this case any network with depth $L > 1$ is equivalent to a 1-layer neural network

$\square$