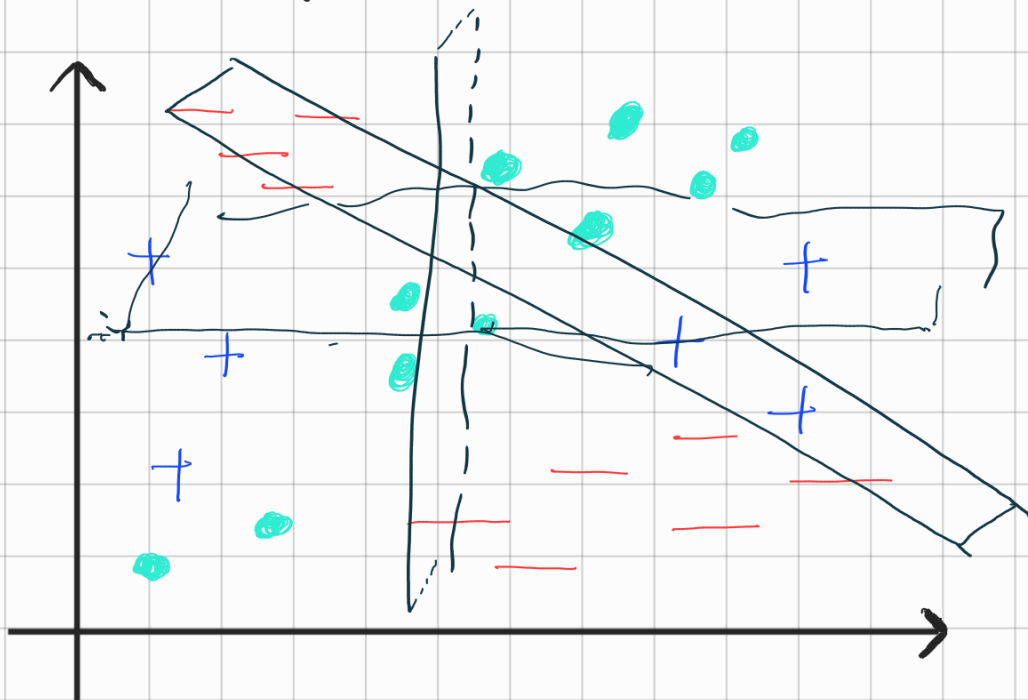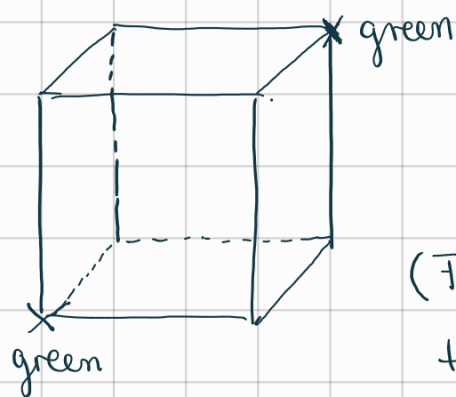# 1. Hand-Crafted Network

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

2. Linear Activation Function

$$Z_0 = X$$
$$\overline{Z}_\ell = Z_{\ell-1} \cdot B_\ell + b_\ell$$
$$Z_\ell = \phi_\ell(\overline{Z}_\ell)$$

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
$$Z_\ell = \phi_\ell(\overline{Z}_\ell)$$
$$= id(\overline{Z}_\ell)$$
$$= id(Z_{\ell-1} \cdot B_\ell + b_\ell)$$
$$= Z_{\ell-1} \cdot B_\ell + b_\ell$$

$$Z_3 = Z_2 \cdot B_3 + b_3$$
$$= (Z_1 \cdot B_2 + b_2) \cdot B_3 + b_3$$
$$= ((Z_0 \cdot B_1 + b_1) \cdot B_2 + b_2) \cdot B_3 + b_3$$
$$= (X B_1 B_2 + b_1 B_2 + b_2) \cdot B_3 + b_3$$
$$= X B_1 B_2 B_3'$$

Let ℓ be greater than 1

$$b_1 B_2^3 + b_2 B_3 + b_2$$

$$Z_\ell = Z_{\ell-1} \cdot B_\ell + b_\ell$$
$$= X \cdot \prod_{i=1}^{\ell} B_i + \sum_{i=1}^{\ell} b_i \prod_{j \geq 1}^{\ell-1} B_{j+1}$$

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 $Z_1 = Z_0 \cdot B_1 + b_1$
$$= X \cdot B_1 + b_1$$

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

$\square$