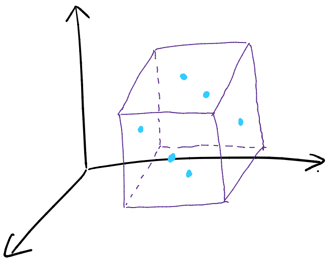
Problem Set3 Problem1

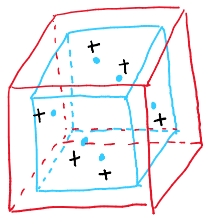
VC dimension $= 2\times 3 = 6$

I. Construct a dataset contains 6 points lying in the center of each side of a cube as follows:

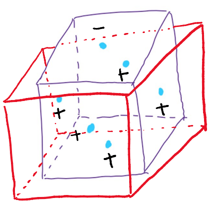


Then prove that the dataset is separable using the hypothesis space with any labeling.

① 6+, 0−



② 5+, 1−

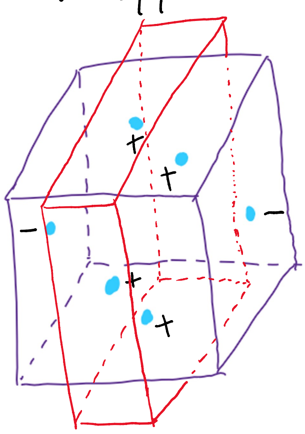


Because of the symmetry of cube, it's the same case when any of the other 5 points is labeled as −.

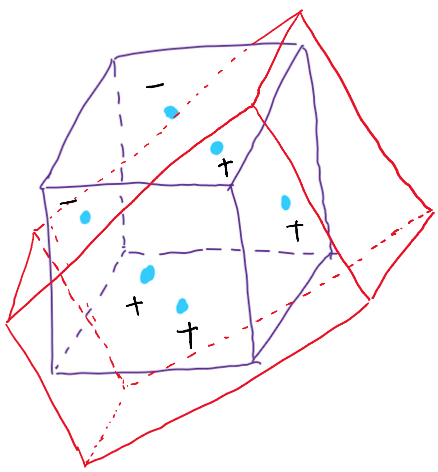
③. 4+, 2−

The two points of − will be on opposite side or adjacent side.

Case 1: opposite side.

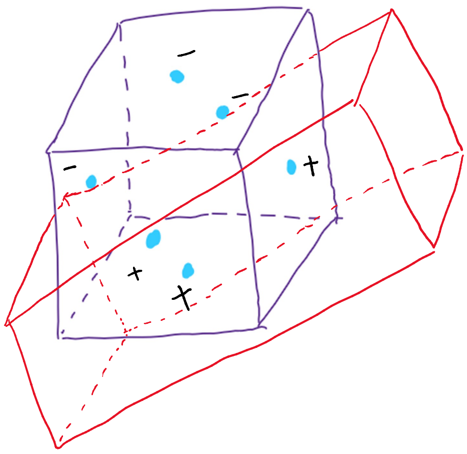


Case 2: adjacent side

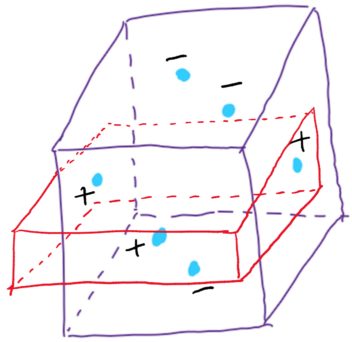


④ 3+ , 3−

case 1: among the 3 points of −, any two of them are on adjacent sides.



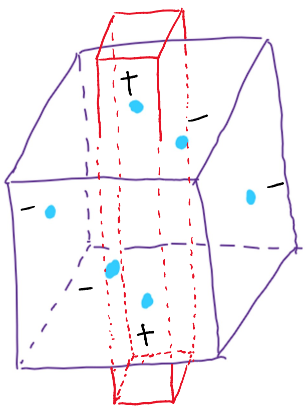
case 2: among the 3 points of −, two pairs are on adjacent side, one pair are on opposite side.



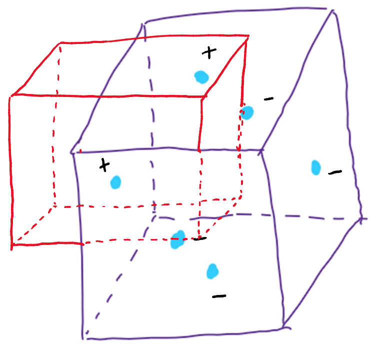
⑤ 2+, 4−

The two points of + can be on the adjaent side or opposite side.

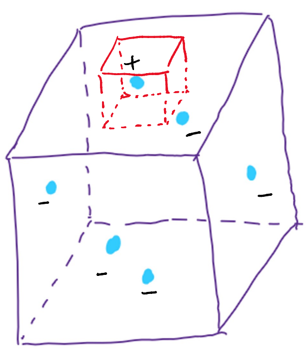
case 1: opposit side



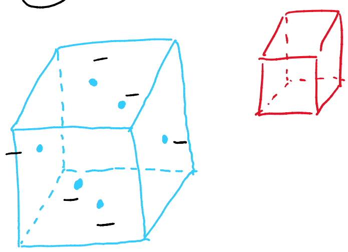
case 2: adjacent side



⑥ 1+, 5−



⑦ 0+, 6−



II. Show that for any 7 sample points, $\exists$ a labeling such that no hypothesis can separate + and −.

Denote the coordinate of point as $(X_1^{(m)}, X_2^{(m)}, X_3^{(m)})$, $m=1,\cdots,M$.

Labeling the points with $\max_m X_1$, $\min_m X_1^{(m)}$, $\max_m X_2^{(m)}$, $\min_m X_2^{(m)}$, $\max_m X_3^{(m)}$, $\min_m X_3^{(m)}$ to be +, the other point is −.

Then we cannot separate this dataset by any box in the hypothesis space.

Following the same idea, the VC dimension of the hypothesis space consisting of axis aligned boxes in $\mathbb{R}^d$ is $\boxed{2d}$.