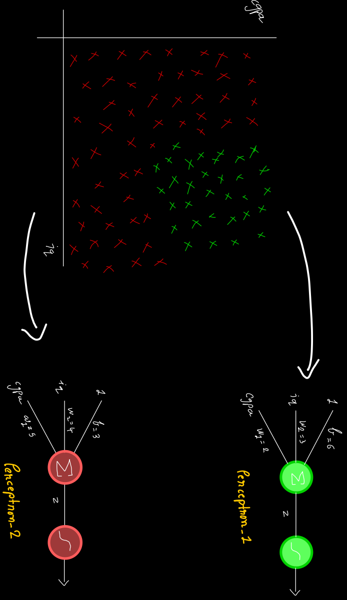
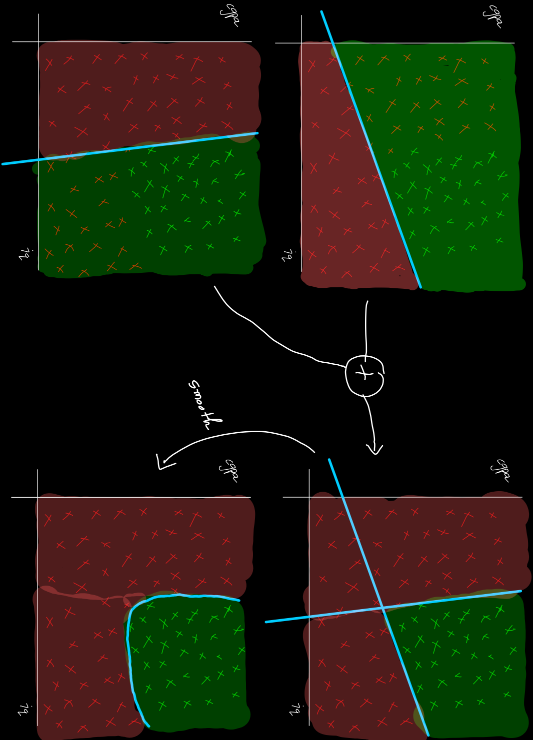
MLP (Multi-layer perceptron)

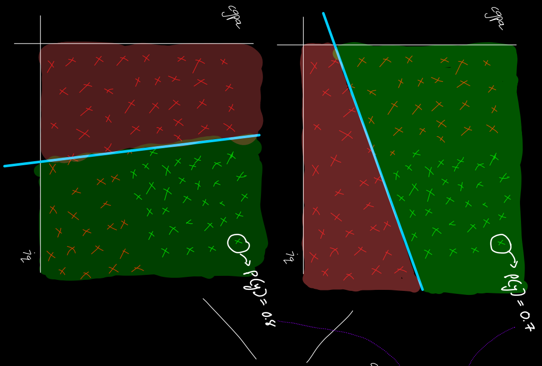
-> A single perceptron can't handle non-linear data hence we use a structure with multipal neurons to handle non-linearity of data.

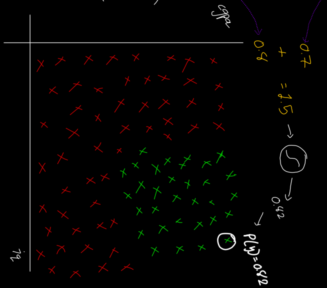



cgpa
iq
+
smooth

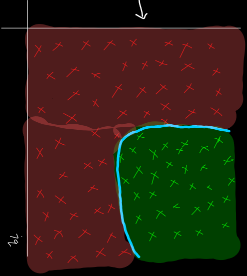
Mathematical intuition

↳ How this process works?

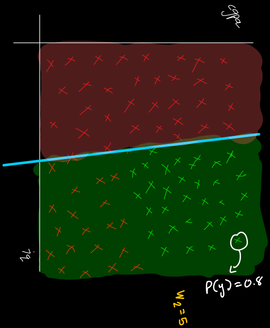


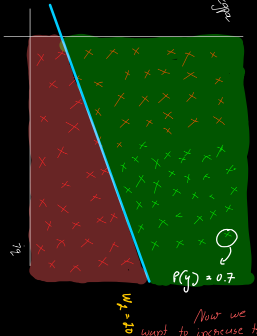


Now repeat this process for all data points.



Dominant Perceptron



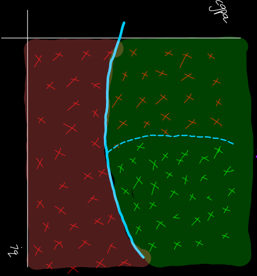


$0.7 \times W_1 + 0.8 \times W_2$

Now we want to increase the importance of perceptron-1 in final output

→ Hence we will multiply output of both perceptrons with a we

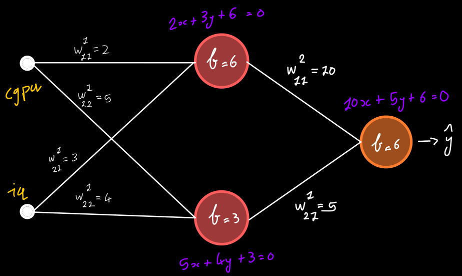
Perceptron-1 is dominating.



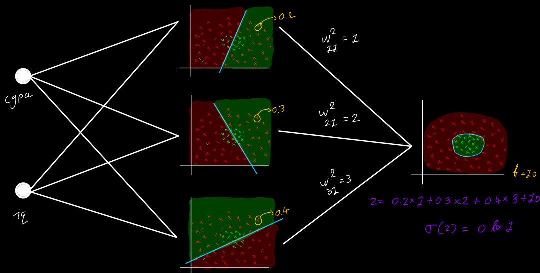
$\leftarrow b = 3$

$z = 0.7 \times 10 + 0.8 \times 5 + 3$

$\sigma(z) = 0.9999$



Adding more nodes



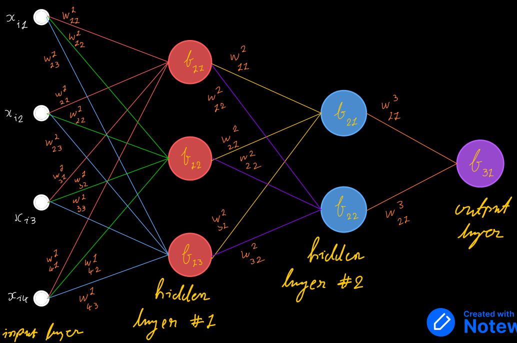
$z = 0.2 \times 1 + 0.3 \times 2 + 0.4 \times 3 + 10$

$\sigma(z) = 0$ to 1

→ More nodes = More complex non-linear boundaries

Forward Propagation

$a^{[0]}$



$$\begin{bmatrix} w^1_{11} & w^1_{12} & w^1_{13} \\ w^1_{21} & w^1_{22} & w^1_{23} \\ w^1_{31} & w^1_{32} & w^1_{33} \\ w^1_{41} & w^1_{42} & w^1_{43} \end{bmatrix}^T \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \\ b_{14} \end{bmatrix}$$

4×3 $\qquad$ 4×1 $\qquad$ 4×1

for hidden layer #1

$$\begin{bmatrix} w^1_{11} x_{i1} + w^1_{21} x_{i2} + w^1_{31} x_{i3} + w^1_{41} x_{i4} + b_{11} \\ w^1_{12} x_{i1} + w^1_{22} x_{i2} + w^1_{32} x_{i3} + w^1_{42} x_{i4} + b_{12} \\ w^1_{23} x_{i1} + w^1_{23} x_{i3} + w^1_{33} x_{i3} + w^1_{43} x_{i4} + b_{13} \end{bmatrix}$$

for hidden layer #2

$$u^{[1]} \begin{bmatrix} O_{11} \\ O_{12} \\ O_{13} \end{bmatrix} \begin{bmatrix} w^2_{11} & w^2_{12} \\ w^2_{21} & w^2_{22} \\ w^2_{31} & w^2_{32} \end{bmatrix}^T + \begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix}$$

$$\begin{bmatrix} w^2_{11} O_{11} + w^2_{21} O_{12} + w^2_{31} O_{13} + b_{21} \\ w^2_{12} O_{11} + w^2_{22} O_{12} + w^2_{32} O_{13} + b_{22} \end{bmatrix}$$

<u>for output layer</u>

$$a^{[2]} \quad \begin{bmatrix} O_{21} \\ O_{22} \end{bmatrix} \begin{bmatrix} w^3_{11} \\ w^3_{21} \end{bmatrix}^T + \begin{bmatrix} b_{31} \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} O_{21} w^3_{11} + O_{22} w^3_{21} + b_{31} \end{bmatrix} \right) = \hat{y}_i$$

$$\boxed{\sigma(W^T X + b)}$$

$$a^{[1]} = \sigma(a^{[0]} w^{[1]} + b^{[1]})$$

$$a^{[2]} = \sigma(a^{[1]} w^{[2]} + b^{[2]})$$

$$a^{[3]} = \sigma(a^{[2]} w^{[3]} + b^{[3]})$$

$$\underbrace{\sigma\Big(\underbrace{\sigma\big(\sigma(\underbrace{a^{[0]}w^{[1]}+b^{[1]}}_{a^{[1]}})\,w^{[2]}+b^{[2]}\big)}_{a^{[2]}}\,w^{[3]}+b^{[3]}\Big)}_{a^{[3]}}$$