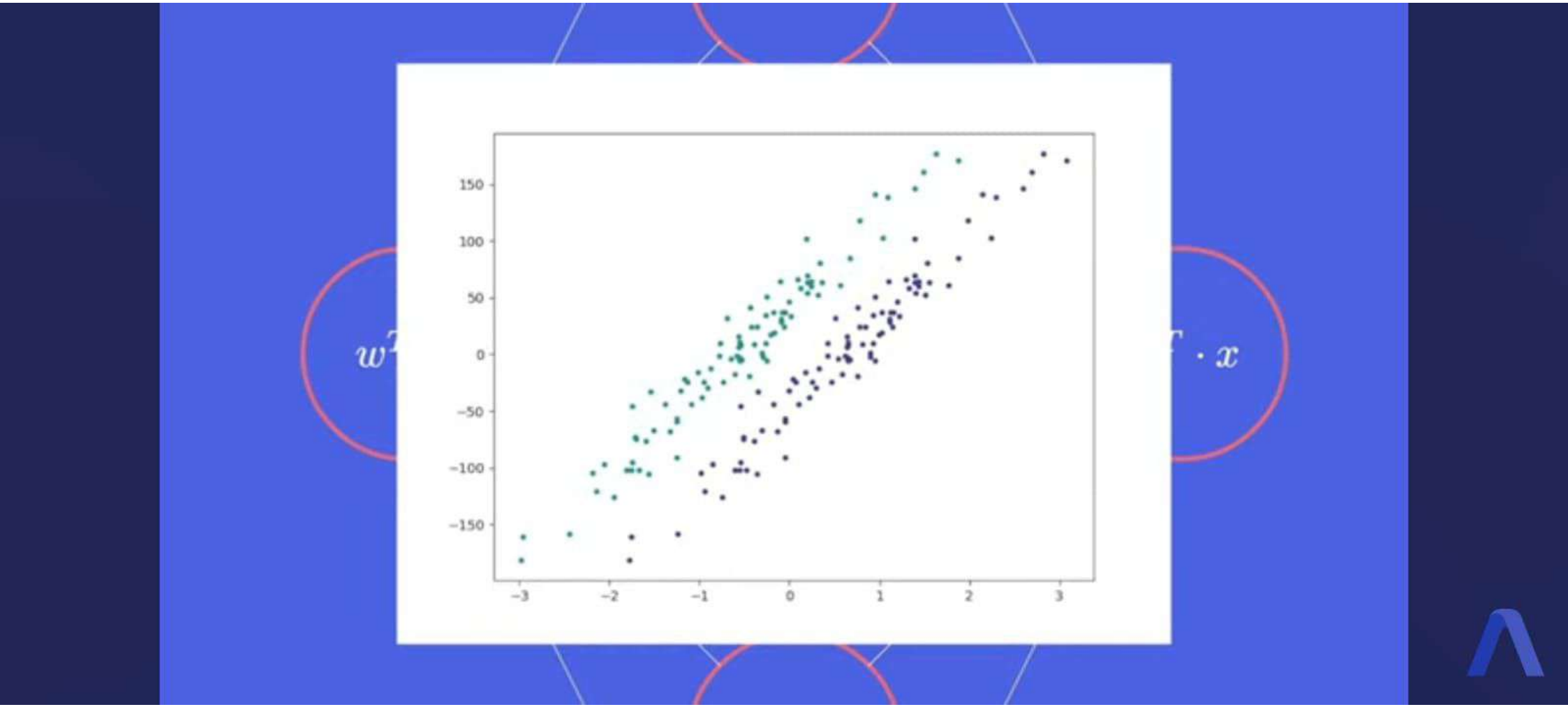
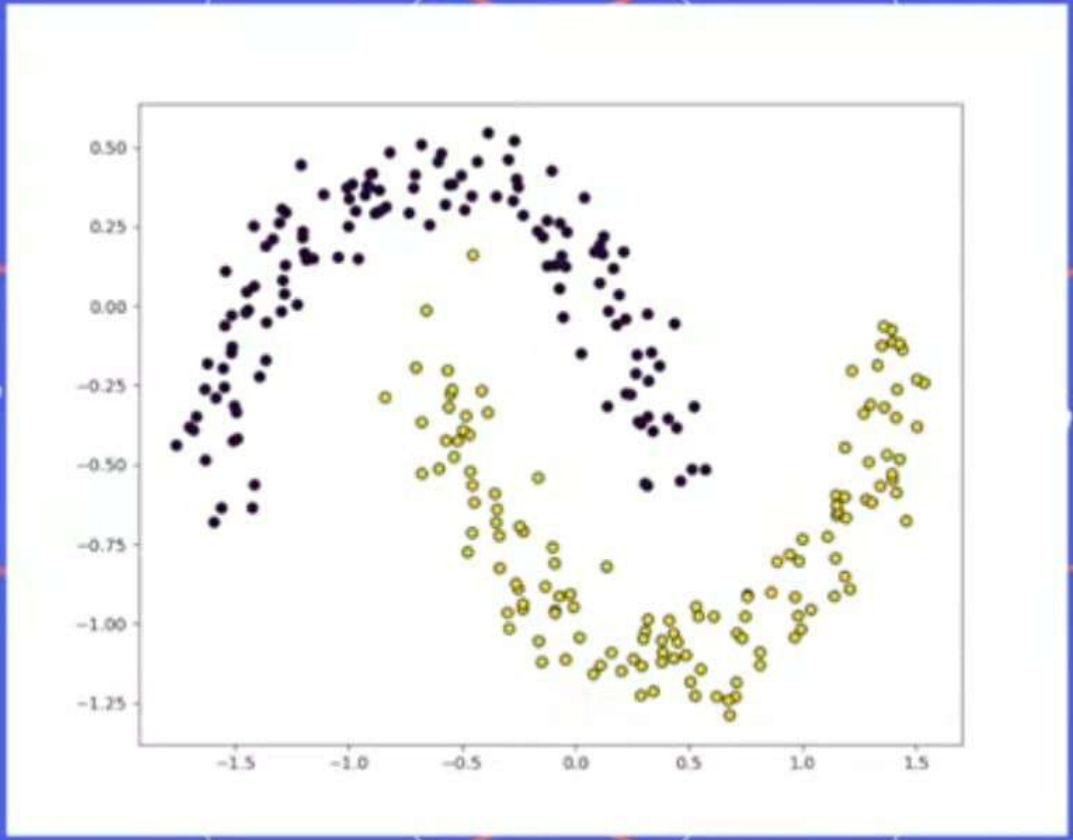

150
100
50
0
−50
−100
−150
−3
−2
−1
0
1
2
3

w^T

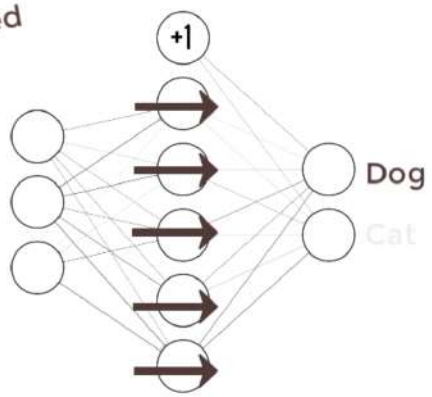


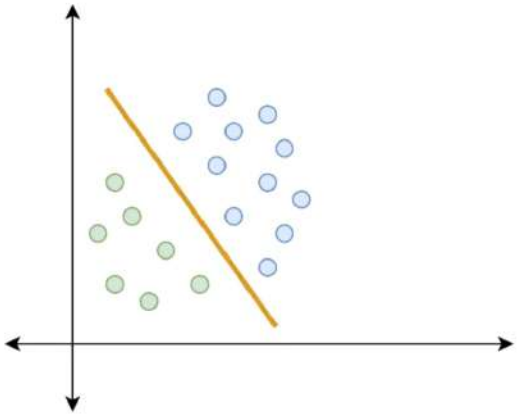
$T \cdot x$

Activation Functions

Case I

Just pass through unfiltered

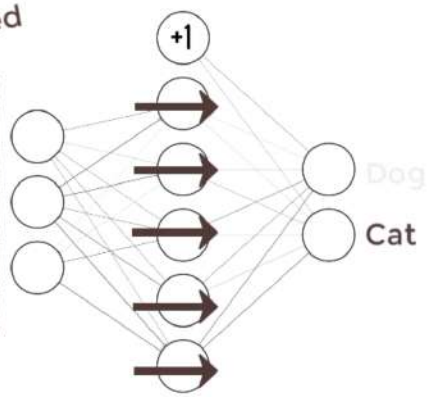




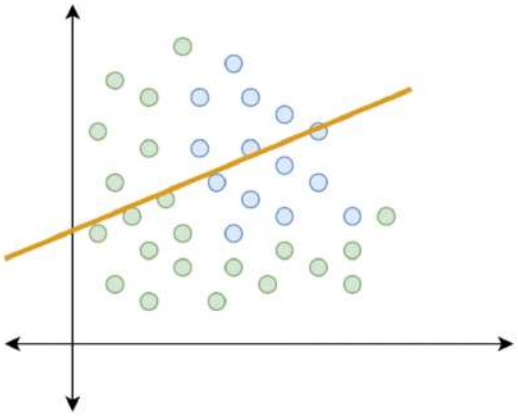
Case II

Activation Functions

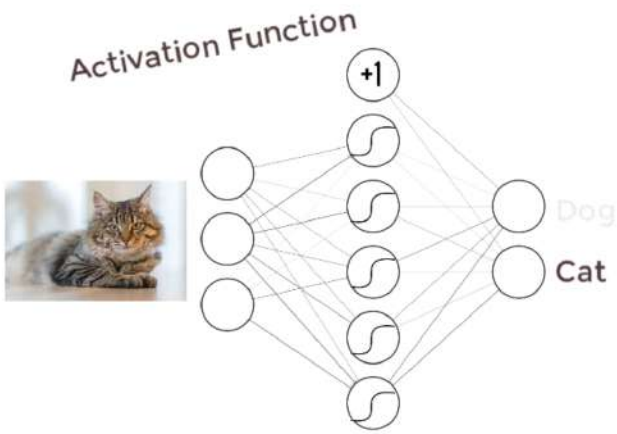
Just pass through unfiltered

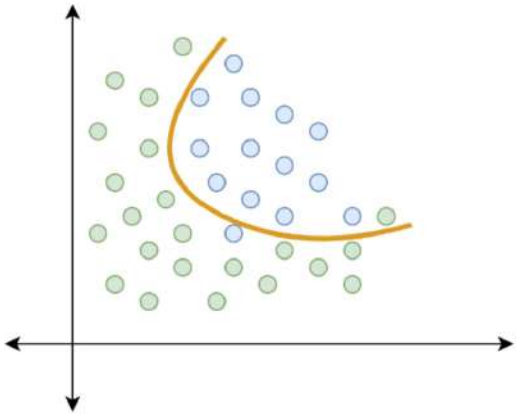


Notice Boundary does not curve



Case II Activation Functions



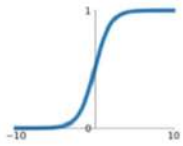


towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044

Pawan Jain Follow
Jun 12, 2019 · 6 min read

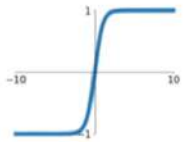
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



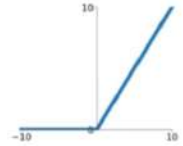
tanh

$$\tanh(x)$$



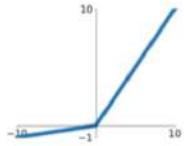
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

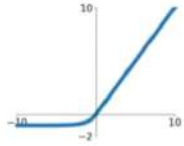


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

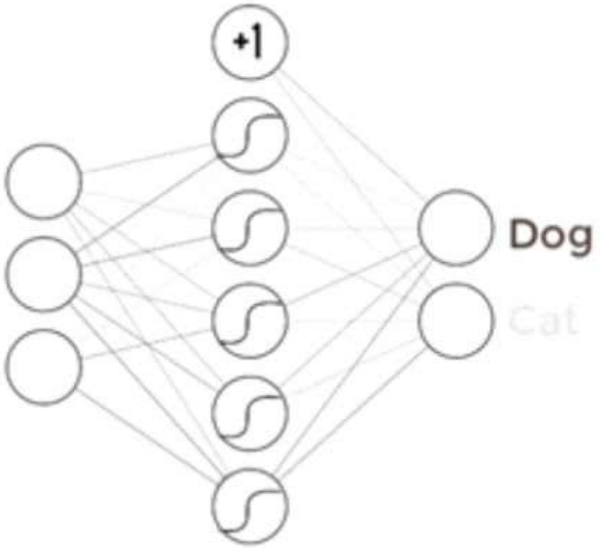
ELU

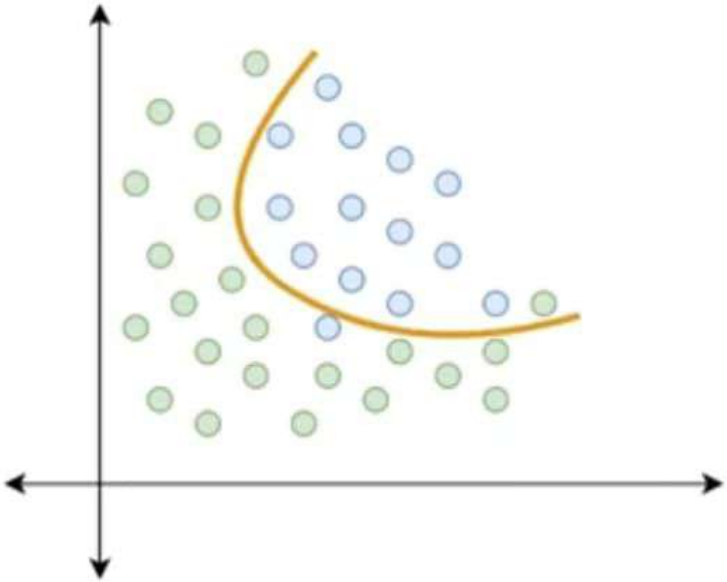
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



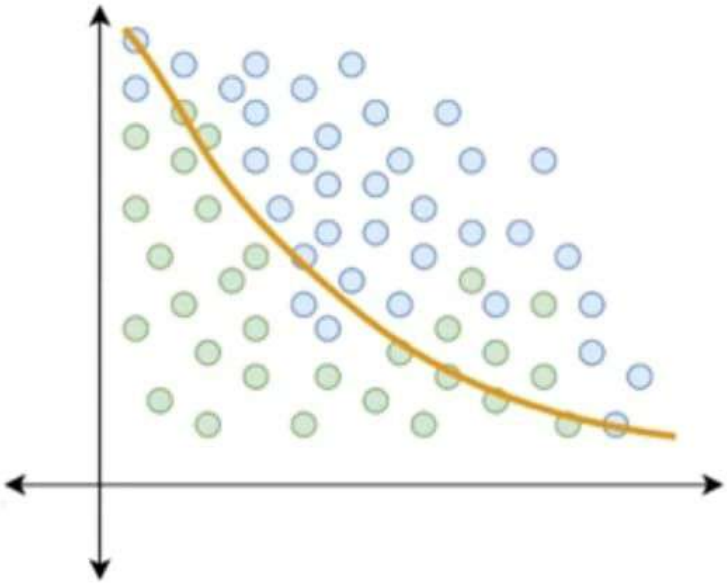
Case II

Activation Functions





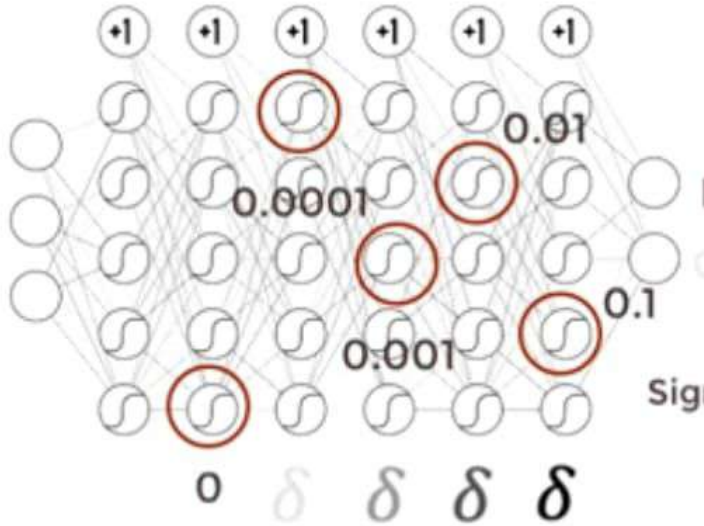
Case III Activation Functions



Vanishing Gradient Problem

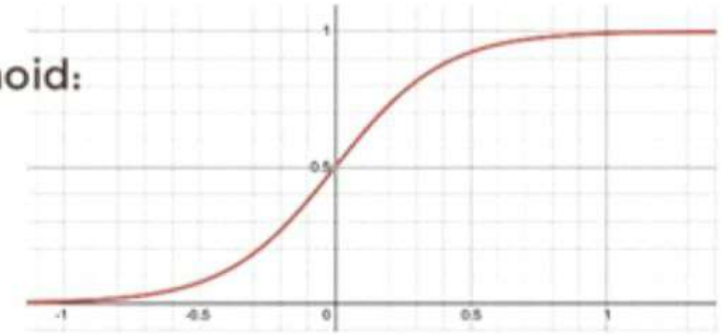
Case III Activation Functions

Why do gradient vanish with sigmoid?



Vanishing Gradient Problem

Sigmoid:



Sigmoid Gradient :