

26-10-2021

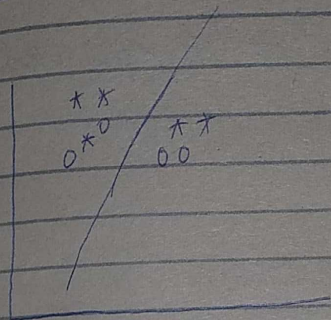
week #9 lecture #12

Tuesday

L_2 regularization = prefers ^{small} weights & distributed weights

If size of dataset is increased, we can be saved from overfitting.

High Variance & High Bias:-



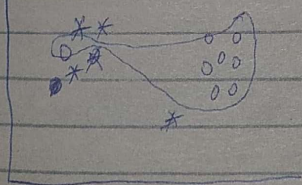
This classifier is really bad bcz it doesn't fit on the data.

This is called High Bias bcz we assumed a ^{simple} ~~good~~ classifier could fit the data set.

Train set performance = bad

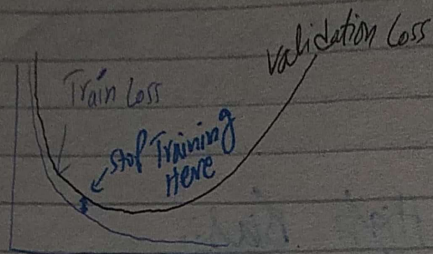
Test set performance = bad

High Variance:-



Training Set Performance = very Good

Test Set Performance = Bad



If we have more difference in these loss, then we should stop our Training.

Solution:

① Regularization

② Increase Dataset

To avoid High Variance

Exploding Gradient Problem.

If we don't use Sigmoid function, then weights Gradients are larger values,

2000
 w_1

50000
 w_2

20000
 w_3

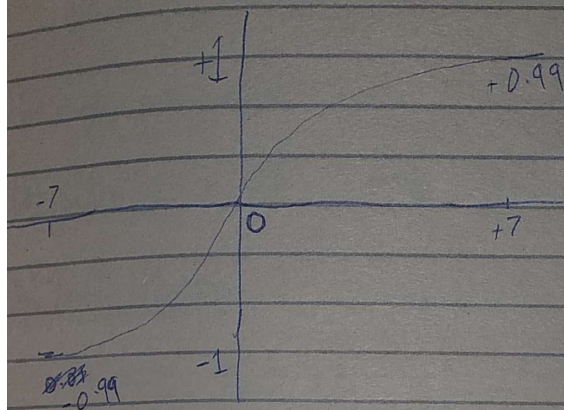
250000
 w_4

all multiplied together, will get

larger gradient values.

tanh Activation function:

$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



We prefer this Activation function over sigmoid. In this weights, gradients, are not zero centered, it can give negative values in its output.

for example:

$$\tanh = \frac{e^2 - e^{-2}}{e^2 + e^{-2}}$$

It doesn't saturate our output that much than

Sigmoid, still it saturates our output. Range has been increased.

$$-7 \leftrightarrow +7$$

Its output can be ~~to~~ $-ve$ or $+ve$, then the gradients doesn't
move only in $-ve$ or $+ve$ directions so weights will move
in both directions independently.