

NN: Practical aspects of designing MLPs

What can be done to make the model perform better?

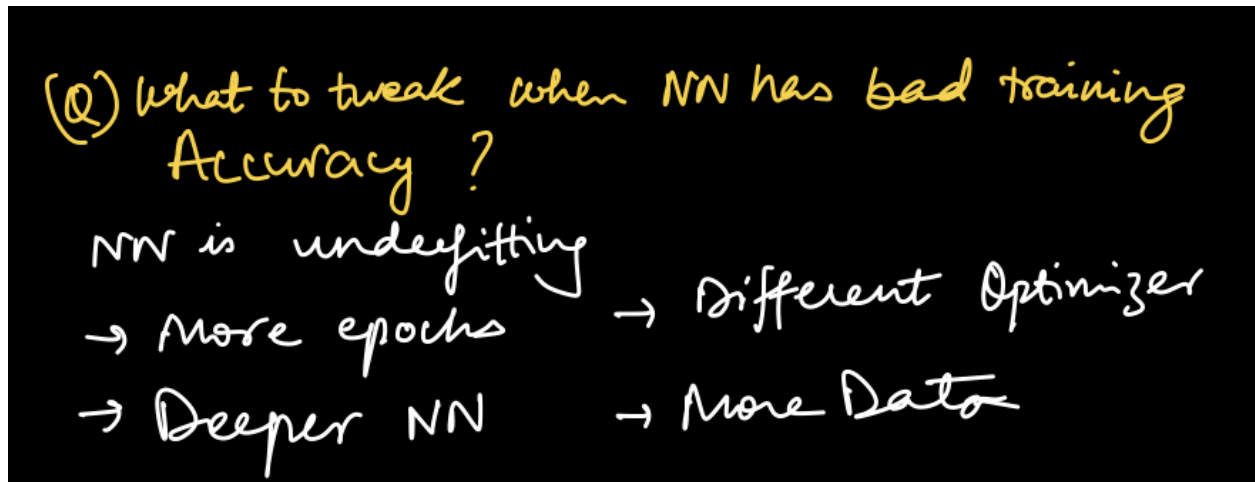
Ans: many things can be done to the baseline model like:

- Training for more epochs
- Trying a Deeper Neural Network
- Adding L2 Regularization
- Using Batch Normalization
- Trying Dropout
- Simply Changing the NN Architecture
- Or finally Collecting more data

Note: It is really important to know what to try from all of this:

- to achieve the desired outcome
- Or else, you will spend months experimenting, and never getting any results

This process of knowing what to tune to achieve one effect is called **Orthogonalization**



Now if the model, does not fit well on the training set, what to do ?

Ans: Bigger Network or use a different Optimizer (Preferably Adam)

What if the model fits well on the training data but performs poorly in Validation set?

Ans: clearly its the case of **Overfitting**, hence:

- Regularization should be used
- Also, getting a bigger training data with diverse sample works

What is diverse sample in training data ?

Ans: The new samples should be completely different from the original training set

What if the model fits well on the training and validation data but performs poorly in test set?

Ans: Though the chances of this happening is very low, but if it does happen:

Try getting a bigger validation data. Or change the Loss function of the model

(Q) what are the main hyperparameter to look for when using NN?

- Learning Rate (α)
- β value in GD with momentum
- $\beta_1, \beta_2 \in$ in Adam
- Number of layers of NN
- Number of neurons/hidden unit
- Learning Rate decay
- Batch size

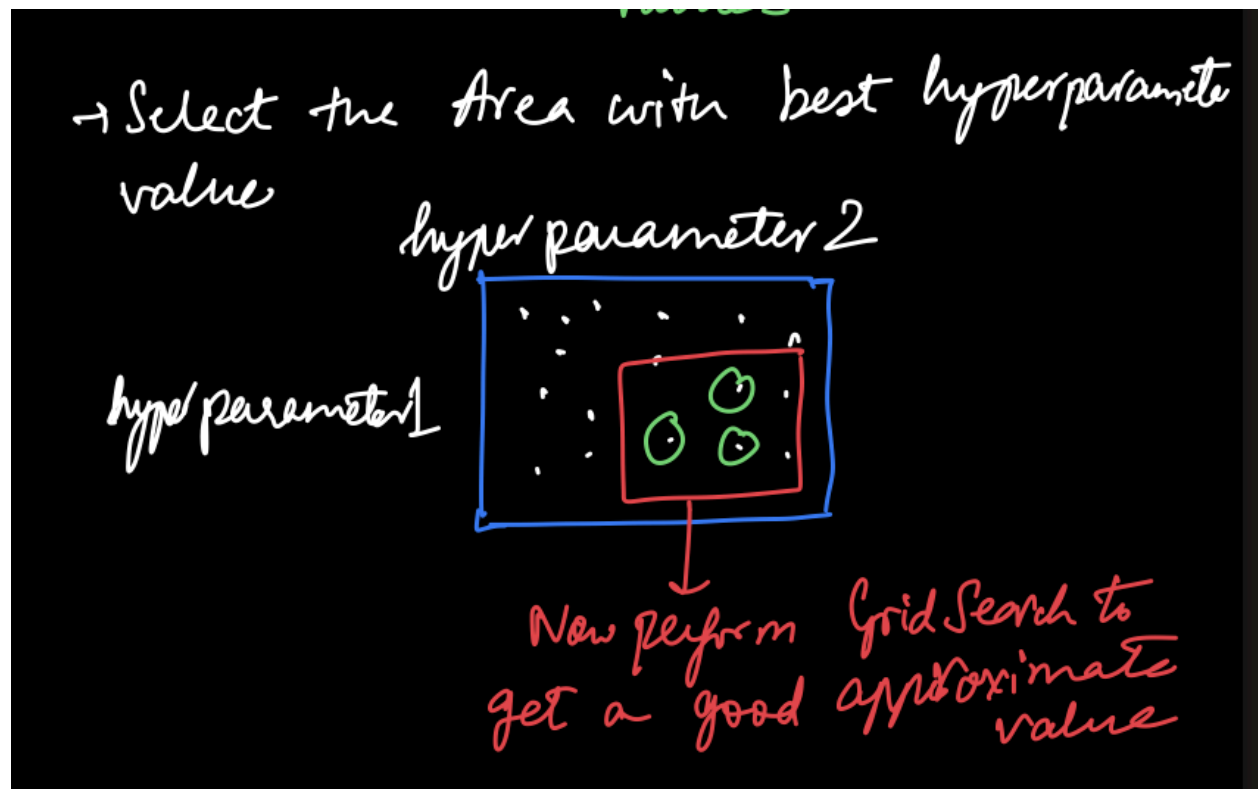
(Q) What should be the order in which these hyperparameters be tuned?

- 1) Learning Rate (α)
- 2) β value of GD with momentum
- 3) Number of Neuron/hidden unit
- 4) Batch-size
- 5) Number of layers of NN
- 6) Learning Rate decay
- 7) β_1 , β_2 , ϵ of Adam (Not Recommended for tuning)

What if there are N hyperparameters, will grid search be the right approach then ?

Ans: No, as the hyperparameters increases so does the time of Grid Search

- Hence to avoid it there is a hack to perform
- Try RandomSearch first to get the best hyperparameter values



After Random Search,

- Then select the region/area of the best hyperparameter
- And perform Grid search on the selected region
- to have a good approximate value of the hyperparameters to be used