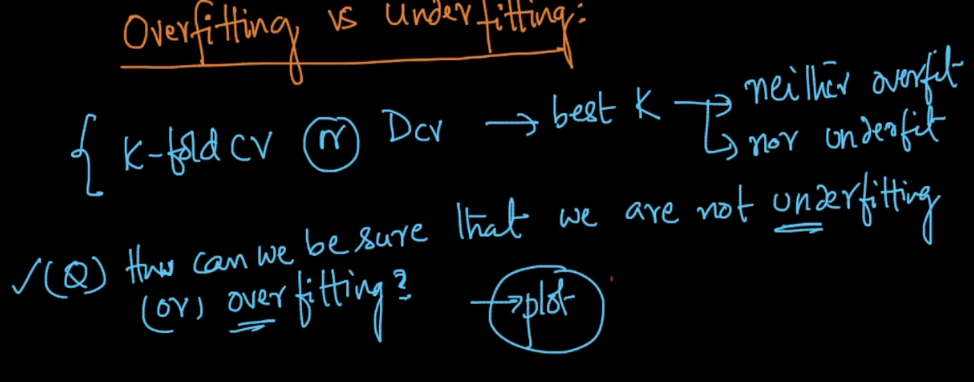
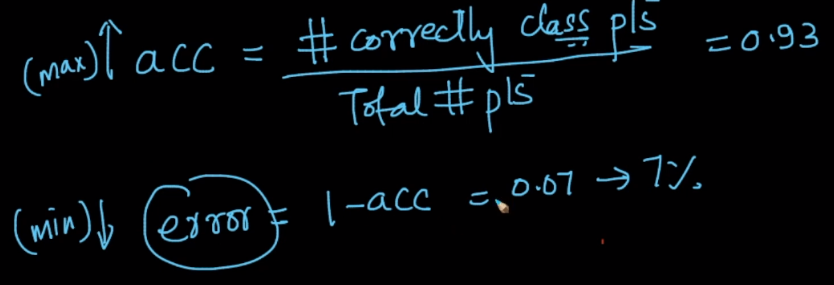
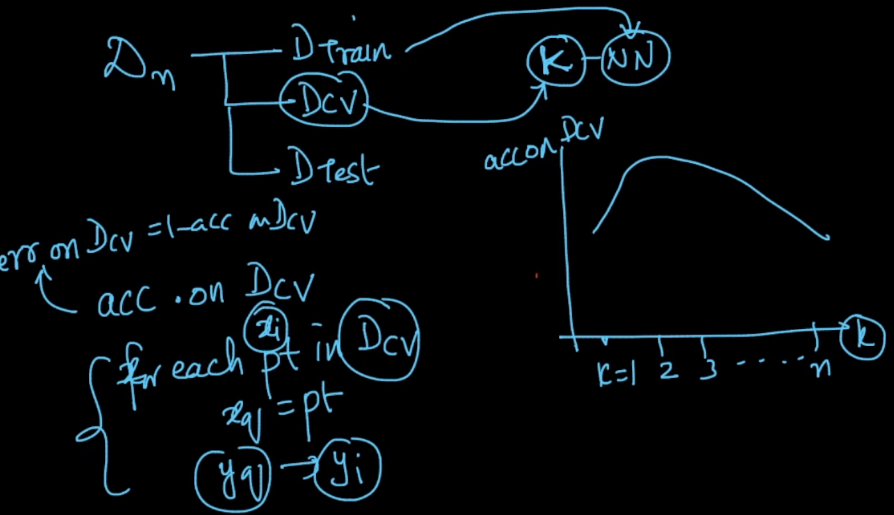
This will tell us how to determine that we are overfitting or underfitting.



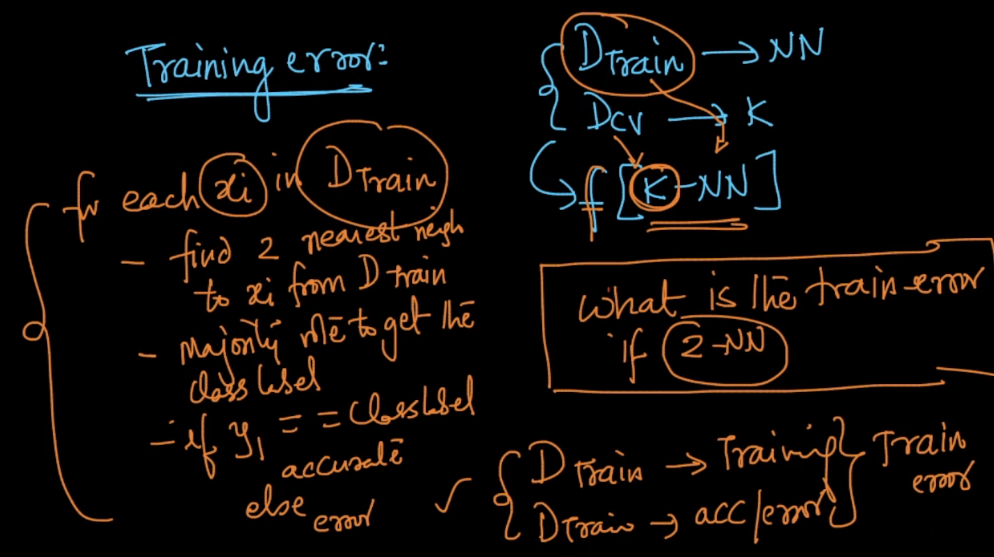
Accuracy and error are calculated as:

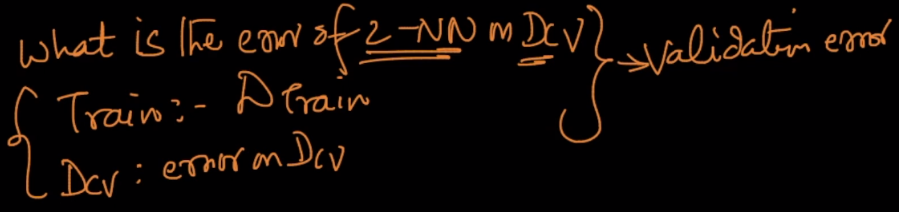


First we find CV accuracy and CV error using CV dataset.



Now we find training accuracy and training error using training dataset applied on k-NN trained using training dataset.





Now if plot CV error and train-error for different k values, our plot look like as given in below image.

At k = 1, train error will be 0, because k-NN model trained using training data, and we are again applying train data to get train error for k=1, so in this case each point exactly lies on corresponding real position of train data, therefore error will be 0.

Now the region where CV-error is high and train-error is low, is called overfit.

And the region where CV-error is high and train-error is also high is called underfit.

And the region where CV-error is minimum, is a good fit.

