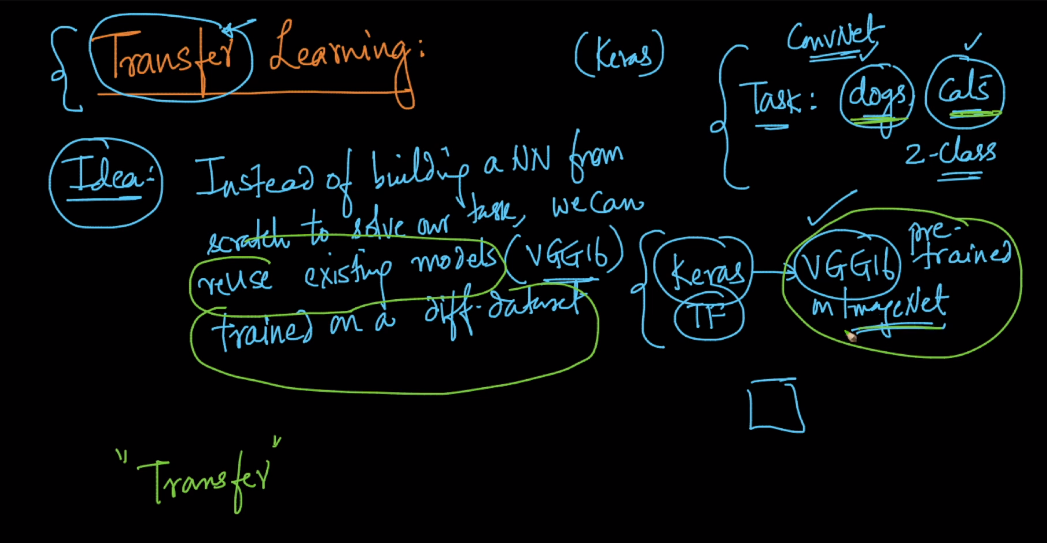
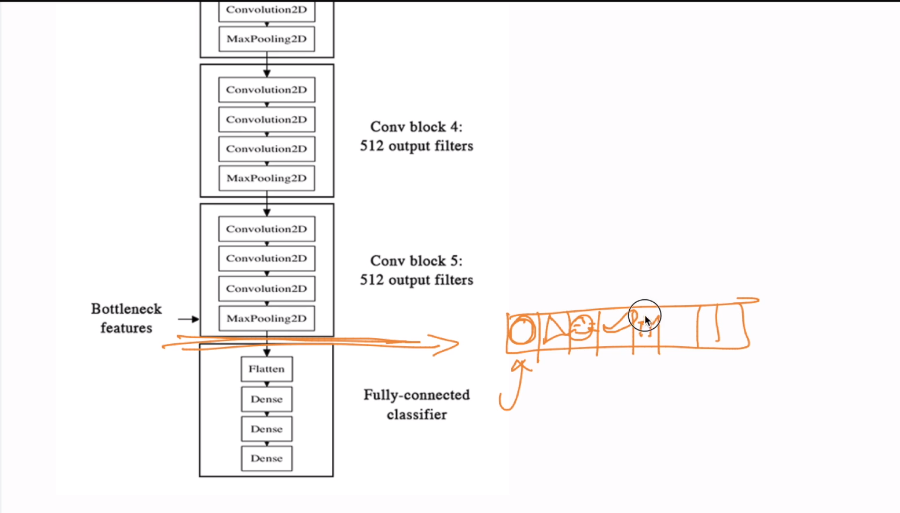
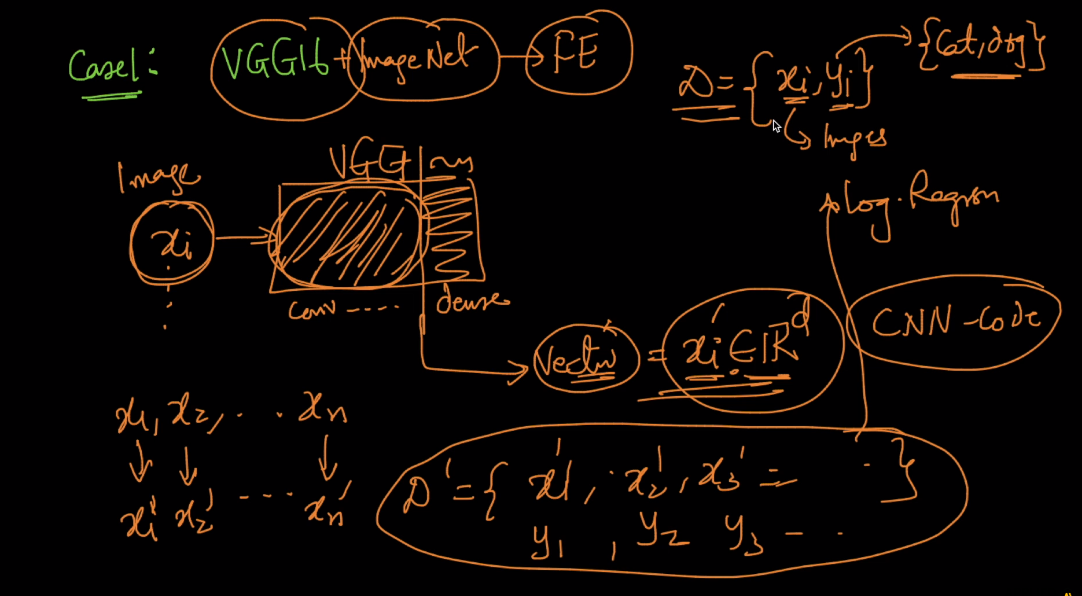
**Transfer learning :**



Case1 :  first we are removing the last dense layers from VGG 16 network(**we dont train this**). Suppose we have data point (xi,yi).Now we are passing our input data xi to this network. The output obtained from this VGG 16 is considered as xi'. Now we train a new model like logistic regression with (xi',yi) to predict whether it is dog or cat.

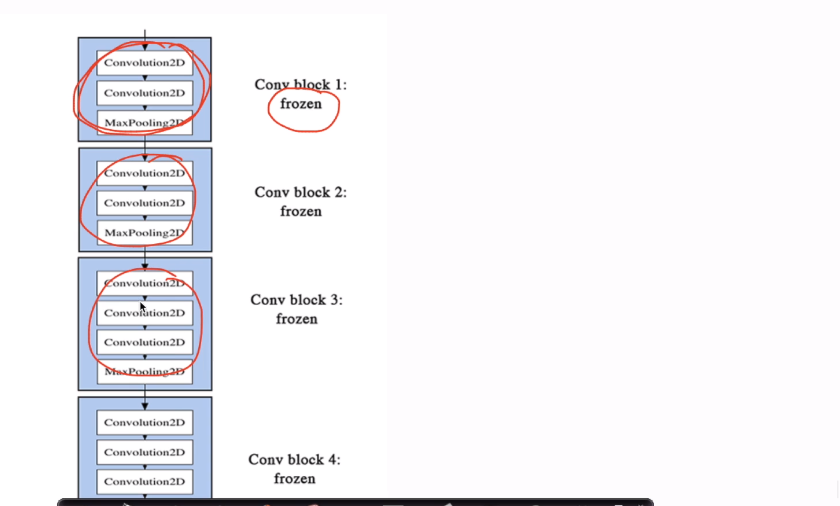
Here we use vgg16 as a feature engineering.

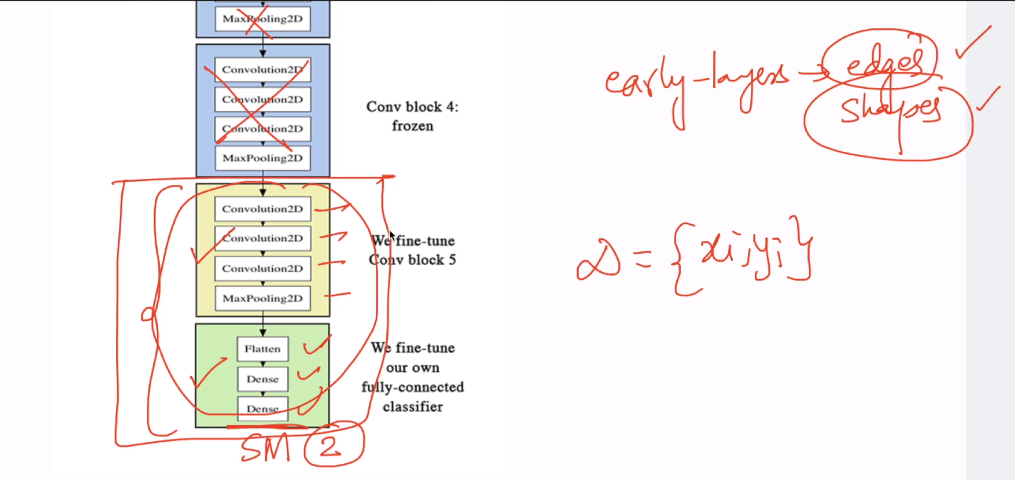


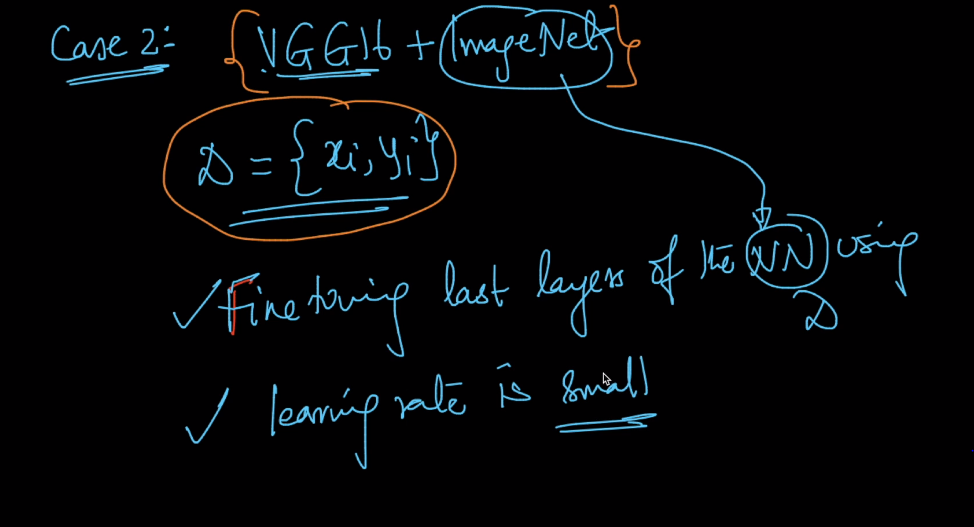


Case 2: we are making the weights W of starting layers constant (not updating them during training) and only train the last few layers of VGG 16 on our (xi,yi).

In this we frozen all layers and then we just train few low layers.

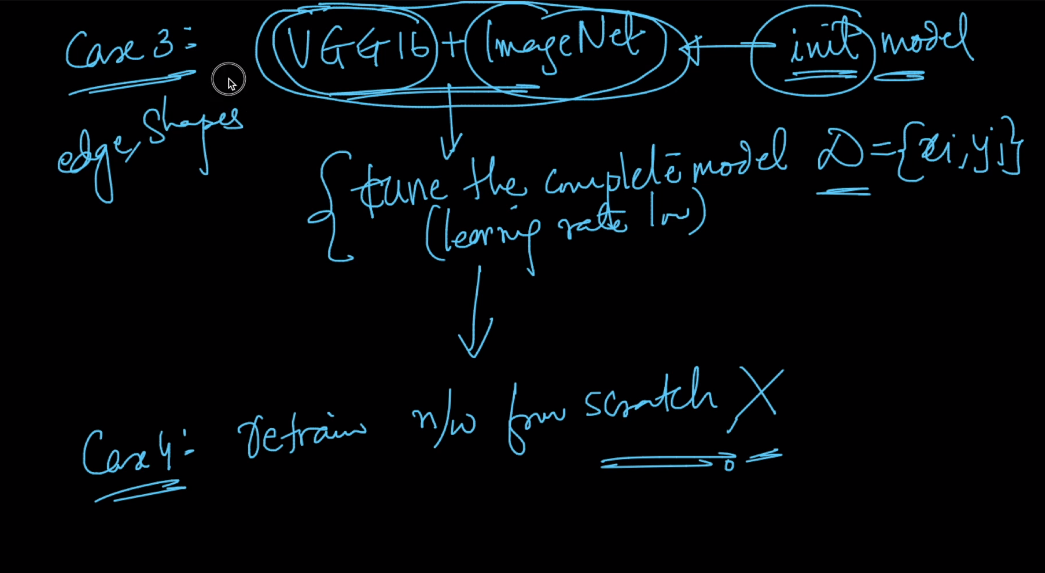




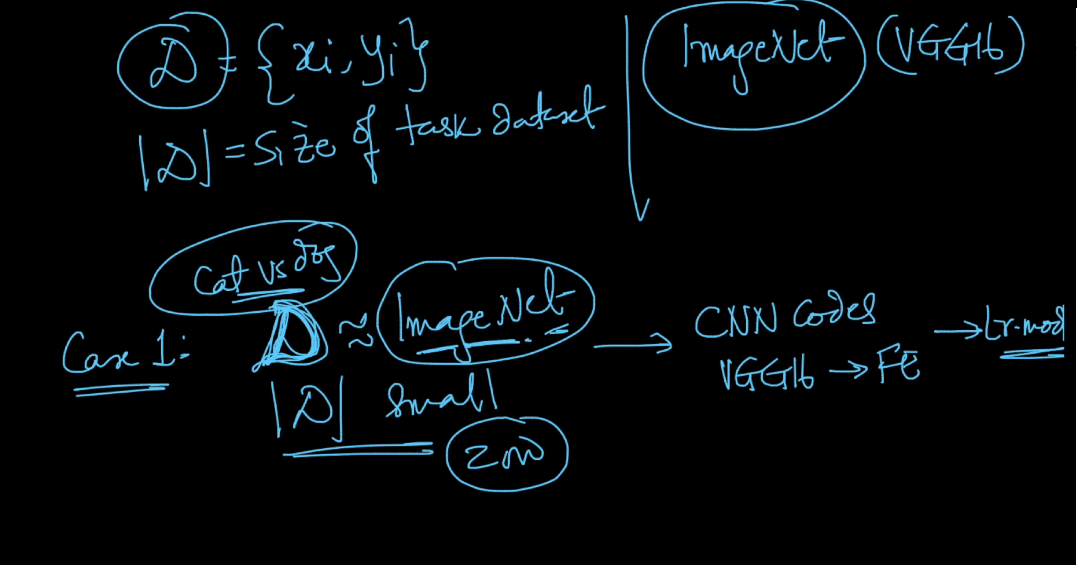


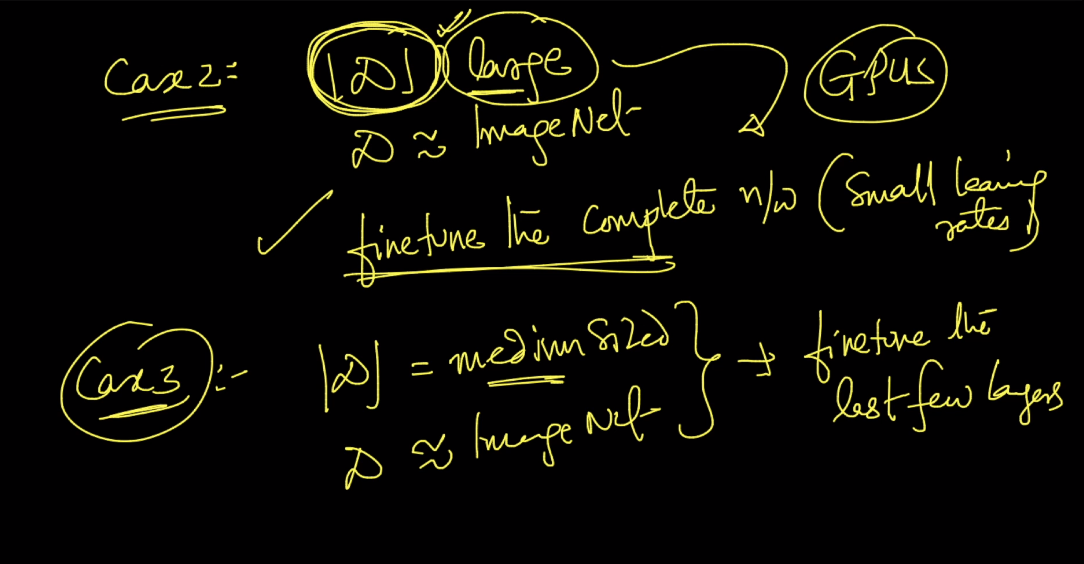
Case3 : In this case we use vgg16 + imagenet to initialize model and train/fine tune the whole model with low learning rate.

We use vgg16 + imagenet bcz it is well trained model and help to train our model better.



When to use which ?





Case 4 : when d is small and it is not like imagenet then we can use few initial layers like edge and shape and flatten this layer thus we got feature map/cnn codes.

