Assignment 04 Solutions

1. What is the concept of cyclical momentum ?

Ans. Momentum in neural networks is a variant of the stochastic gradient descent. It replaces the gradient with a momentum which is an aggregate of gradients as very well explained here. It is also the common name given to the momentum factor

2. What callback keeps track of hyperparameter values (along with other data) during training ?

Ans. Common standards are json, yaml or cfg files. Below is an example yaml file, that presents multiple hyperparameters for random forest along with more general info like project and experiment name. Similarly to the dictionary-based style, you just need to version control this file to keep track of hyperparameters.

3. In the color dim plot, what does one column of pixels represent ?

Ans. For additive colour on computers, the colours red, green and blue (RGB) are used. Each pixel on a screen is typically made up of three tiny "lights"; one red, one green, and one blue. By increasing and decreasing the amount of light coming out of each of these three, all the different colours can be made.

4. In color dim, what does "poor teaching" look like? What is the reason for this ?

Ans. The challenge with images having multiple color channels is that we have huge volumes of data to work with which makes the process computationally intensive. In other worlds think of it like a complicated process where the Neural Network or any machine learning algorithm has to work with three different data (R-G-B values in this case) to extract features of the images and classify them into their appropriate categories.

The role of CNN is to reduce the images into a form that is easier to process, without losing features critical towards a good prediction. This is important when we need to make the algorithm scalable to massive datasets.

5. Does a batch normalization layer have any trainable parameters ?

Ans. a Batch Norm layer also has parameters of its own: Two learnable parameters called beta and gamma. Two non-learnable parameters (Mean Moving Average and Variance Moving Average) are saved as part of the 'state' of the Batch Norm layer.

6. In batch normalization during preparation, what statistics are used to normalize? What about during the validation process ?

Ans. Batch normalization solves a major problem called internal covariate shift. It helps by making the data flowing between intermediate layers of the neural network look, this means you can use a higher learning rate. It has a regularizing effect which means you can often remove dropout.

7. Why do batch normalization layers help models generalize better ?

Ans. Batch normalization is a technique to standardize the inputs to a network, applied to ether the activations of a prior layer or inputs directly. Batch normalization accelerates training, in some cases by halving the epochs or better, and provides some regularization, reducing generalization error.

8.Explain between MAX POOLING and AVERAGE POOLING is number eight ?

Ans. Average pooling method smooths out the image and hence the sharp features may not be identified when this pooling method is used. Max pooling selects the brighter pixels from the image. It is useful when the background of the image is dark and we are interested in only the lighter pixels of the image.

9. What is the purpose of the POOLING LAYER ?

Ans: Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network.

The pooling layer summarises the features present in a region of the feature map generated by a convolution layer.

10. Why do we end up with Completely CONNECTED LAYERS ?

Ans. Fully connected layers are global (they can introduce any kind of dependence). This is also why convolutions work so well in domains like image analysis - due to their local nature they are much easier to train, even though mathematically they are just a subset of what fully connected layers can represent.

11. What do you mean by PARAMETERS ?

Ans. In a CNN, each layer has two kinds of parameters : weights and biases. The total number of parameters is just the sum of all weights and biases. Let's define, = Number of weights of the Conv Layer.

12. What formulas are used to measure these PARAMETERS?

Ans. Thus number of parameters = 0. CONV layer: This is where CNN learns, so certainly we'll have weight matrices. To calculate the learnable parameters here, all we have to do is just multiply the by the shape of width m, height n, previous layer's filters d and account for all such filters k in the current layer.