1. What is the concept of cyclical momentum?

Ans:

Cyclical momentum is a technique used in optimization algorithms for deep learning that aims to improve the efficiency of stochastic gradient descent (SGD) by varying the momentum coefficient during training.

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

Ans: TensorBoard callback is used to keep track of hyperparameters.

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

Ans: each column represents the intensity of R, G and B.

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

Ans:

In the context of color vision, "poor teaching" refers to a situation where an individual's visual system does not receive adequate or accurate information during development about how to distinguish between colors. This can result in difficulties distinguishing between certain colors, such as red and green, which are commonly affected in people with color vision deficiencies.

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

Ans: Yes, a batch normalization layer has two trainable parameters per feature/channel - scale parameter (gamma) and shift parameter (beta).

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

Ans: minmax scaler can be used for normalization which is applied on every feature. We calculate the mean and std deviation of the feature by using the fit() method and the transform() method is used to transform the values into normalised values.

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

Ans: normalization is used to scale down the values. The higher magnitude of values slow down the process of reaching the global minima in the optimization process. So scaling down the values through normalisation helps reach the global minima faster in an attempt to minimise the cost function.

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

Ans: Max pooling layer extracts the maximum number from the localised area whereas, average pooling is used to extract the average value from the localised area.

9. What is the purpose of the POOLING LAYER?

Ans: Pooling layer is used to extract the most important information from the localised area after applying filtering. That is also used to reduce the spatial dimension of the image as we only choose the maximum number in a particular localised area. We keep moving the pooling matrix with a particular stride.

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

Ans: Connected layers or ANN are used to calculate the cost function and minimise the cost function by updating weights with the help of optimiser such as Adam optimiser. The Optimiser algorithm find the global minima where the cost function is minimum wrt weights.

The reason of using a fully connected layer towards the end of the network to transform the previous layer output into final output. It is used to learn non linear relationship between features with the help of activation function and the level of influence of inputs on outputs with the help of weights.

11. What do you mean by PARAMETERS?

Ans: parameters are learned during the training process. For example the weights and biases are parameters which are learned during the model training and after training these parameters values are applied to the unseen data. The parameters values are adjusted during the training process with the help of optimisers.

12. What formulas are used to measure these PARAMETERS?

Ans: we use wight updation formula in optimisers to update the weights in which we find the differentiation of the activation function and loss function to be able to calculate weights after every epoch.