1. What is the concept of cyclical momentum?

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

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

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

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

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

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

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

9. What is the purpose of the POOLING LAYER?

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

11. What do you mean by PARAMETERS?

12. What formulas are used to measure these PARAMETERS?

Answers:

1. Cyclical momentum is a technique in which the momentum value used during gradient descent is varied periodically between two bounds. This can help accelerate convergence and prevent getting stuck in local minima.
2. The Recorder callback keeps track of hyperparameter values and other data during training, such as losses and metrics.
3. In the color dim plot, one column of pixels represents the values of the color channels (red, green, and blue) for a single pixel in the image.
4. Poor teaching in the color dim plot refers to having color channels that are highly correlated, meaning they provide redundant information. This can make it harder for the model to distinguish between different objects in the image.
5. Yes, a batch normalization layer has trainable parameters, specifically scale and shift parameters that are learned during training.
6. During training, the batch mean and variance are used to normalize the data, while during validation the running mean and variance (calculated during training) are used.
7. Batch normalization layers help models generalize better by reducing internal covariate shift, which is the phenomenon of the distribution of activations changing throughout the network during training. By normalizing the activations, batch normalization helps stabilize the learning process and allows the model to learn more robust features.
8. Max pooling and average pooling are both types of pooling layers used in convolutional neural networks. Max pooling returns the maximum value within each pooling window, while average pooling returns the average value within each window.
9. The purpose of the pooling layer is to reduce the spatial dimensionality of the output from convolutional layers, while retaining the most important features. This reduces the number of parameters in the network and helps prevent overfitting.
10. Fully connected layers, also known as dense layers, are used at the end of the network to perform the final classification or regression. They take the output from the convolutional and pooling layers and produce a prediction for each class.
11. Parameters refer to the weights and biases of the neural network. They are the learnable components of the model that are adjusted during training to minimize the loss function.
12. The formulas used to measure the parameters of a neural network depend on the specific type of layer and activation function being used. For example, the parameters of a fully connected layer include the weight matrix and bias vector, while the parameters of a convolutional layer include the kernel weights and biases. The number of parameters in a layer is calculated as the product of the number of inputs and outputs, plus the number of biases.