**CAP5415**

**Computer Vision**

**Fall 2022**

**Homework 5**

**Questions 1:**

If we use Stochastic Gradient Descent as an optimization function, we can always find the global optimum solution for our network.

True/False

**False**

**We don’t get global optimum solution always in SGD**

**Questions 2:**

Convolutional neural networks use correlation operation, although they can also use a convolution operation as well.

True/False

**True**

**Convolution and correlation operation both can be used for feature extraction.**

**Questions 3:**

To train a CNN, we have to compute gradients for all the learnable parameters. Therefore, all the layers in a CNN should be differentiable. However, this does not apply to activation functions as we don’t have to learn any parameters for activation functions, and therefore, we can use any activation function whether it is differentiable or not.

True/False

**False**

All layers have to be differentiable to compute gradients and train via back-propagation. This applies to activation layers as well. Some activations cannot be differentiable at very few points, such as ReLU at 0, but in implementation the actual gradient is not exactly 0 but a very small number close to zero. If the activation is non-differentiable, the network will not be able to do gradient based optimization.

**Question 4:**

Overfitting is good as the loss will be very small during the training. However, the only reason we don’t perform overfitting is because it requires a lot of parameters in the network.

True/False

**False**

Overfitting is not good since it memorizes the data distribution instead of learning their features. It does not generalize to new set of data, which is why we dont want to overfit training samples.

**Question 5:**

If the training loss starts increasing after some epochs, it could mean that the learning rate we are using is too high to train the network.

True/False

**True**

Of the many things that can cause increase in training loss, having high/very high learning rate is one of them. High learning rate means instead of moving towards the minima, it might overshoot or be too sensitive to noise, causing loss to increase instead.