**Final Set of Videos**

**Note:** Skip video 28 as it is same as video 27

**Video 27 (Convolutional Layers and Activation Maps)**

**Q1. Why we call convolutional act as feature extraction?**

Ans: In convolution, we apply filters on input layer that yields some parameter called features. Filters determine how many and in what fashion, parameters can affect the final output. That’s why it is called feature extraction.

**Q2. CNN is divided in two big steps. Feature Learning and Classification. What happens in each step?**

Ans: In feature learning

* Convolution is done by applying filter
* Activation functions are used to convert linear data to non-linear
* Pooling is performed but it is optional
* Flattening is performed on data got from pooling
* Fully connected layers perform above operations repeatedly

In classification

* Output layer gives scores for classes
* Softmax function converts scores into proper probabilities
* Loss function is used to measure the difference between actual and predicted value and hence to tune the weights.

**Q3. What is activation/feature map?**

Ans: It is the output of one filter when applied to the result of previous layer. Like image, it also has dimensions such as width x height x depth.

**Video 29 (Activation Functions in Deep Learning (1))**

**Q1. Why we take the depth of filter same as the depth of input image?**

Ans. In RGB image for example, if we apply only RG filter then we will miss the blue details. Filter should have same depth as input image otherwise we will miss major details of input image

**Q2. Why the width and Hight of activation map decreased gradually?**

Ans: Typically, when a filter is applied, it is not applied on border filters. In other words, border pixels are missed if we not use padding or replication techniques. So, the width and height of activation map is decreased gradually.

**Q3. What is the low-level feature, mid-level feature and high-level feature in term of face detection?**

Ans: Low-level features: Features extracted at first layer in CNN. Examples are Points and Edges

Mid-level features: Features extracted in intermediate layers. Examples are shapes made from low level features such as polygons and triangles.

High-level features: Features extracted in last layer. Examples are shapes made from mid-level features such as eyes and lips.

**Q4. Why initial layers detect low level features rather than high level features?**

Ans: High-level features are made up of many low-level features. So, in order to extract high-level features, we need to extract low-level features first. So, we extract low-level features at initial layers.

**Q5. If image size is 32\*28\*3 and we have filter size 5\*3\*3 and stride S=1 then what is the next output?**

Ans: Formula is (N-F) / (S + 1). So, x = ((32-5)/1)+1 = 28 and y = ((28-3)/1)+1 = 26. So, the final size will be 28x26.

**Video 30 (Activation Functions in Deep Learning(Linear and non linear))**

**Q1. What is the role of activation functions in a Deep learning model?**

Ans. Activation functions are used for conversion of linear data/operations to non-linear data/operations. Some of the activation functions are sigmoid, Tanh and Relu. Relu is most used activation function in deep learning applications.

**Q2. Discuss some types of Activation functions and their ranges.**

Ans: Sigmoid is S shaped function. It maps input within 0 and 1.

Tanh is S shaped function. It is zero-centric that is it is balanced on origin. It maps input within -1 and +1.

Relu is the most used function in deep learning applications. Its formula is max(0, x) where x is input.

**Video 31(Activation function in deep learning(sigmoid activation function))**

**Q1. Discuss some problems of sigmoid activation function.**

Ans: We don’t use Sigmoid function in DL because

* It converts all higher and lower values into nearly single constant.
* We cannot take derivative of constant values.
* Values cannot be optimized because derivative of constant is zero.
* It has an exponential component in the formula that is **expensive to compute.**
* It is not zero-centric, so it is not useful for data pre-processing, data cleaning and PCA

**Q2. What is meant by zero centric in term of any activation function?**

Ans: A curve is said to be zero centric if it is balanced on the origin. In other words, it should be symmetric about both x-axis and y-axis.

**Q3. What are advantages of zero centric function?**

Ans. For a line or curve, it is better for it to be zero-centric. So that,

* We don’t need to do an extra calculation i.e. finding the intercept.
* Used in data cleaning e.g. Principal component analysis (PCA).
* Models operating on normalized data enjoy faster convergence.

**Video 32(Activation function in deep learning(tanh activation function))**

**Q1. If your network is too deep and the computational load is a major problem then which of the activation function you would prefer to use and why?**

Ans. I would prefer Relu because it is least expensive in terms of computation because it doesn’t contain any exponential component in its formula.

**Q2. What is the difference in tanh and sigmoid function? In which case you prefer tanh over sigmoid?**

Ans. Tanh is zero-centric whereas sigmoid is not. Tanh maps output within -1 and +1 whereas Sigmoid maps output within 0 and 1.

**Video 33 (Activation function in deep learning(tanh activation function))**

**Q1. What is the main functionality of RelU? Why the computation power of RelU is faster than sigmoid and tanh?**

Ans. Relu has formula y = max(0, x). It maps all the negative values to zero and positive values unchanged. Relu does not contain any exponential term. So, it is 6 times more computationally efficient than Sigmoid and Tanh.

**Q2. Sigmoid is also not zero centric. What is the solution to overcome this negativity?**

Ans. Solution is to use any zero-centric activation function such as Leaky Relu that has formula max(0.01x, x). Also parametric Relu or pRelu that is max(alphaX, X)can be another option.

**Q3. Differentiate between Parametric Relu and Leaky Relu? How can you choose the value of ‘Alpha’?**

Ans. In Leaky Relu max(0.01X, X), the value 0.01 is fixed whereas pRelu or parametric Relu has alpha instead of 0.01. Alpha is determined while training the model.

**Video 34 ( CNN Architecture in deep leaerning (LeNet)**

**Q1. What is term architecture means in Deep Learning? Discuss important part of architecture(names only).**

Ans. The art or practice of designing and constructing the techniques and placing them in right order to achieve the desired output expectation. LeNet, AlexNet, VGG, GoogLeNet and ResNet are some architectures presented in ILSVRC.

**Q2. In 1\*1 filter why the width and Hight of the input and output image remain same?**

Ans. since 1\*1 filter causes operation on border pixels as well so, input and output size remains same

**Q3. When we applied 1\*1 filter why the computation power required for next layer is minimum as compare to the initial layer?**

Ans. Smaller filters such as 1x1 make the depth lesser as compared to the input image originally has.

**Q4. Inlist the sequence of LeNEt architecture layers and on which type of problems we can consider LeNet architecture?**

Ans. LeNet architecture has 2x convolution layer, 2x pooling layer, and at the end fully 2x fully conected layers. It can be used on MNIST dataset for Handwritten digit recognition.

**Video 35 (AlexNetCNN Architecture in deep learning (Convolutional Filters)**

**Q.1 What do we mean by deeper architecture? Compare AlexNet and LeNet.**

Ans. Deeper architecture means larger number of layers as compared to AlexNet and LeNet.

**Q2. Which Activation function is used by AlexNet?**

Ans. Alexnet used Relu as activation function.

**Video 36 (AlexNet CNN architecture in deep learning(1))**

**Q1. What the term wight initialize, weight optimize and weight update mean in deep learning?**

Ans. In deep learning the output of a neuron is calculated according to this formula xw + b = y Here x is input, w weights, b is base and y is output. Here the starting values of weights are called the weights initialization, in order to get the values, close to Y, the values of w are update and weight optimize mean the values of weight which make the output close to y.

**Q2. What do we mean by ground truth also describe what is loss?**

Ans. Ground truth means the label provided. Difference between predicted value and actual value is called loss and the function that measures it called loss function.

**Q3. What is difference between learning and weight optimization in deep learning?**

Ans. Learning is the initial process of assigning weights. Weight optimization is secondary and performed latter after the process of training.

**Video 37(ZF net Architecture in Deep Learning)**

**Q1. Discuss the ZF Net architecture in terms of convolutional layers and folters.**

Ans. ZF Net architecture is like AlexNet architecture. It consisted of 5 shareable convolutional layers, max-pooling layers, dropout layers, and 3 fully connected layers. Loss is 11.7%

**Q2. Why ZF net perform better than Alex Net?**

Ans. ZFnet performs better than Alex Net because ZF Net has 7x7 size filter in first convolutional layer while Alex Net has 11x11 size filter in first convolutional layer. Smaller filter need few parameter which are learned during training and with small filter there is less chance of losing pixel information.

**Q3. Why in ZF Net they reduced the size of filter? Why VGG-16 is Expensive to train?**

Ans. ZF Net they reduced the size of filter because smaller filter need few parameter which are learned during training and with small filter there is less chance of losing pixel information. VGG-16 is Expensive to train because it needs 138 million parameters for training.

**Video 38 (GoogLeNet Architecutre)**

**Q1. Discuss the architecture of googLeNet?**

Ans. In deep learning the term architecture consisted of a 22 layer deep CNN and no totally connected layers however reduced the amount of parameters from sixty million (AlexNet) to four million.

**Q2. What is inception module? What are its benefits?**

Ans. Inception Modules square measure utilized in Convolutional Neural Networks to permit for a lot of economical computation and deeper Networks through a dimensional reduction with stacked 1×1 convolution.

**Q3. What is meant by the term ‘Bottleneck’ layer?**

Ans. A bottleneck layer is a layer that contains few nodes compared to the previous layers.

**Video 39 (ResNet Architecture)**

**Q1. What are residual networks how they are different?**

Ans. Residual network consists of 152 layer model for ImageNet. It is different in a sense that it improves the error rate to 3.57% which is better than human performance. ResNet is able to train very deep networks without degrading.