<u>Computer Vision Discussion Quiz – 1</u>

1. Explain the sentence, "Convolutional Neural Networks (CNNs) are specifically designed to deal with the variability of 2D shapes".

Answer: Because pictures are 2D forms, this term suggests that Convolutional Neural Networks are especially built for them. As a result, CNNs are specifically built for image and handwriting recognition. CNN collects significant characteristics from pixel pictures quickly and effectively. Shift invariance is achieved automatically in convolutional networks by requiring the repetition of weight configurations across space. By limiting the receptive fields of hidden units to be local, these force the extraction of local characteristics.

2. What are the problems/limitations in traditional pattern recognition models and how they are overcome by the method presented in the paper?

Answer: One of the main problems in traditional pattern recognition models is feature extraction process. Extracting features manually would be tiring process so CNN is used to overcome this problem. Another problem in traditional pattern recognition models is training each module separately and training them again together in pairs, and this limitation was overcoming by global training, here we train all the modules globally, so we don't have train each module one after another.

3. All learning techniques need a minimal amount of prior knowledge about the task. How can we incorporate that prior information in the case of multi-layer neural networks?

Answer: We can incorporate the prior information in multi-layer neural networks with prior backpropagation knowledge. Multi-layer neural networks require little prior information, this prior information greats aids in lowering the training error. This prior backpropagation knowledge is used for weights altering. So this helps in lowering the training error and this improves the efficient outcome of the model.

4. What is the use of training data, validation data, and test data? Why the test data should be disjoint from training data?

Answer:

Training data: We consider a subset of data from a larger collection of data to be training material for machine learning algorithms. We supply input data to the algorithm. It evaluates the data on a regular basis to gain a better grasp of how it behaves, and then adapts itself based on the desired outcome.

Validation data: Validation data is the data we use to assess the model after we've trained it with training data in order to evaluate the data. Validation data is used as a first check to see how well the machine has been trained and how accurate the predictions are.

Test data: The remaining data from the given data is treated as test data, and it serves as a last check to ensure that the machine's predictions are correct and that it has been properly trained.

Test data should be disjoint from training data because this will help us to train the machine good and test the ability of the machine thoroughly because if test and training data is disjoint there's no data in common between them, so this helps to test the machine with unknown data so that we could it's true potential otherwise if we test with the same data we used for training, as the machine already is trained on it, it's familiar with it and it'll be very easy for the model to give desired output.

5. What is structural risk minimization? Explain how it controls the tradeoff between minimizing the training error and minimizing the expected gap between training and test error?

Answer: Structural risk reduction is based on establishing a series of increasing-capacity learning machines that correspond to subsets of the parameter space, each of which is a superset of the preceding subset. Minimizing Etrain + H(W) is used to perform structural risk minimization, where H(W) is a regularization function and is a constant. H(W) is designed so that it accepts big values for W parameters in high-capacity subsets of the parameter space. Minimizing H(W) controls the tradeoff between decreasing the training error and minimizing the predicted gap between the training and test errors by limiting the capacity of the accessible subset of the parameter space.

6. The presence of local minima in the loss function does not seem to be a major problem in practice in the multi-layer neural network, why?

Answer: The presence of local minima in the loss function is not a major problem in practice in multi-layer neural networks because all non-identifiable configurations operate in an indistinguishable manner. This leads to the production of same error in all training, validation and test data sets as the training dataset teaches all the models same. So, the presence of local minima is not intrinsically hazardous.

7. What are the advantages of a CNN over a fully-connected architecture for handwriting recognition task as discussed in paper 1?

Answer:

Advantages of CNN over a fully-connected architecture for handwriting recognition:

The adoption of CNN provides for substantial cost savings in terms of computation.
CNN gets key characteristics from pixel pictures directly.
Because CNN is extremely accurate, it doesn't require a lot of training data.
In comparison to fully-connected architecture, CNN is compatible with any hardware and device.
In fully-connected architecture, the topology of the input is completely disregarded, whereas CNNs require the extraction of local features by confining the receptive fields of hidden units to be local.

8. Define the terms: (a) receptive field, (b) feature map, (c) bi-pyramid structure of a network, (d) time-delay Neural Network

Answer:

a) Receptive field: These are the fields that assist neurons extract basic visual elements like oriented edges, endpoints, and corners from other information like speed spectrograms.

b) Feature map: A layer's units are grouped into planes, each of which has the same set of weights. A feature map is the collection of outputs from the units in such a plane. A feature map's units must all perform the same activity on various portions of the image. A complete convolutional layer is made up of numerous feature maps, allowing for the extraction of multiple features at each point.

c) Bi-pyramid structure of a network: The number of feature maps increases as the spatial resolution decreases at each layer of convolutions and sub-sampling, culminating in a "bi-pyramid."
d) Time-Delay Neural Network: Time-Delay Neural Networks are fixed-size convolutional networks that communicate weights along a single temporal dimension. Phoneme recognition, spoken word recognition, on-line recognition of isolated handwritten letters, and signature verification have all made use of them.

9. Explain the saturation of activation function. How can this be overcome? Can you name an activation function that does not have this problem?

Answer: The mean activation value of the layers is more than 0.5, and it decreases as we progress from the output layer to the input layer. In deeper networks with sigmoid activations, this type of saturation might endure for a long time. The saturation phase can be avoided with an intermediate number of layers. At the same time, the saturation of the top hidden layer decreases. The first concealed layer starts to saturate and, as a result, stabilizes. The combination of random initialization and the fact that a hidden unit output of 0 equates to a saturated sigmoid causes this behavior. Because of their symmetry around 0, deep networks with sigmoids but initialized from unsupervised pre-training (example: RBMs) and Hyperbolic tangent networks do not suffer from the saturation behavior of the top hidden layer seen with sigmoid networks.

10. Explain gradient vanishing and gradient explosion in the context of deep neural networks?

Answer: In the context of deep neural networks, gradient explosion is the explosion that occurs when the derivatives are large, and the gradient grows exponentially as we travel through the model. This kind of explosion is called as gradient explosion and gradient vanishing occurs when the derivatives are tiny, and the gradient falls exponentially as we travel through the model. This kind of vanishing process is called as gradient vanishing.

Note: For 10th question to get a clarity, I referred to this cite: https://towardsdatascience.com/the-vanishing-exploding-gradient-problem-in-deep-neural-networks-191358470c11

For 6th question to get an understanding, I referred to this cite: https://medium.com/@pranabbhadani/is-the-local-minima-a-real-issue-in-deep-neural-learning-6d812b28d684

And all other questions are answered as per my understanding and as per the given research papers.