AlexNet and VGGNet

Paper 1: Gradient-Based Learning Applied to Document Recognition.

Paper 2: Understanding the Difficulty of training Deep feedforward Neural Networks.

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

2. What are the problems/limitations in traditional pattern recognition models shown in Fig. 1 (paper1) and how they are overcome by the method presented in this paper?

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?

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

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?

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?

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

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

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

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