ML-Final Exam

Friday, May 14, 2021 1:56 PM

Q1. A

Q2. B

Q3. B, C

Q4. A, B

Q5. A, D

Q6. Training error is high, this indicates a bias problem. Thus increasing the amount of data will be reducing the variance, but is not likely going to solve our problem. A good method would be to decrease the bias of the model by maybe adding more layers / learnable parameters. It is a possibility that training converges to a local optimum. Training longer / using a better optimizer / restarting from a different initialization could work.

Q7. Benefits of using a Convolutional Neural Network layers insteads of fully connected are as follows:

- Filters in CNN work as feature extractors for visual tasks and are better than Deep Networks as filters localize over a patch of the image while Deep networks work at a pixel level.
- The complexity of CNN in terms of computation and memory is better compared to Deep networks.
- CNN has less parameters.
- CNN filters processes image at multiple levels thus becoming a better feature extractor as size of objects in image doesn't matter now.

Q8.

Layer	Activation map dimensions	No. of parameters
Input	128 x 128 x 3	0
Conv-9-32	120 x 120 x 32	32 x (9x9x3 + 1)
Pool-2	60 x 60 x 32	0
Conv-5-64	56 x 56 x 64	64 x (5x5x32 + 1)
Pool-2	28 x 28 x 64	0
Conv-5-64	24 x 24 x 64	64 x (5x5x64 + 1)
Pool-2	12 x 12 x 64	0
FC-3	3	3 x (12x12x64 + 1)

Q9.

(i) The amount of training data is insufficient.

(ii) DNN's accuracy can be improved by:

- Regularization
- Drop-out
- Reduce model complexity /order
- Early stopping
- Taking a larger training data.

(iii) The data size is sufficient for the capacity of the model.

Q10.

False. Adding more hidden layers will make the network deeper increasing the chance of vanishing gradients.

False. This will increase the magnitude of loss thus pulling gradients close to 0. This will make vanishing GD worse.

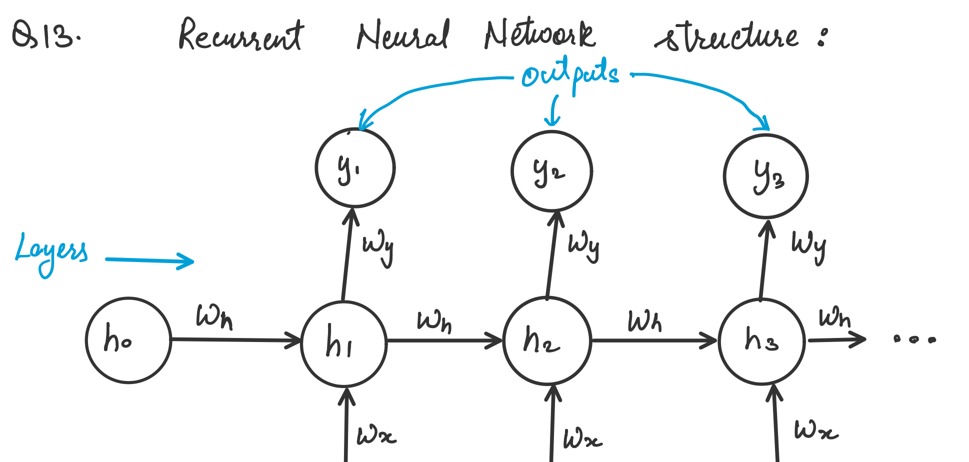
True. By setting a cut-off value to our gradient descent, we make sure that our GD value doesn't surpass the threshold value. Thus this solves the exploding gradients problem.

Q11. (i) This means that the model is underfitting.

(ii) Increasing the model complexity can help with underfitting problem. Additionally if a regularization term is used, then decreasing the regularization term can help tackle underfitting problem.

Q12. Drop-out happens to be a regularization technique used in artificial neural networks for reducing overfitting. In this technique, the nodes at each layer are randomly dropped out of the network at each iteration thus preventing complex co-adaptations on the training data. It happens to be an efficient method to tackle overfitting.

Q13. Recurrent Neural Network structure:



Input data types that require memory eg: Natural Language Data are suitable for recurrent neural networks.

Q14. False. The goal of using PCA is dimentionality reduction, which is the process of reducing the number of variables in a dataset. This is an unsupervised method.

Q15. Autoencoders: are an unsupervised learning technique in which we leverage neural networks for the task of representing learning. We can use them by designing a NN such that we impose a bottleneck in the network which forces a compressed knowledge representation of the original input. This can be used in image processing.

Semi-Supervised learning: In this type learning technique, the machine is provided with a small portion of data which is labeled where majorly other data is unlabeled. The labeled data is then used to label the unlabeled data by building a model on labeled data. Then the whole data is available to be used. It can be used in natural language processing to analyse speech.

Transfer Learning: is a technique that focuses on storing knowledge gained while solving one problem and applying it to a different, but similar problem. This can be applied to text classification.

Q16. Principal Component Analysis finds only orthogonal basis vectors, where non-linearity can also be implied. Linear PCA is equivalent to linear auto-encoder whereas non-linear PCA is equivalent to auto-encoder with non-linear activation. Although the feature space learned by NN or CNN based auto-encoder may or may not be orthogonal or orthonormal like PCA. So they are equivalent, but not exactly same.