Department of Computer Science and Engineering
Indian Institute of Technology Madras

January_May 2024 Semester

CS6910: Fundamentals of Deep Learning
UG Section
Midsemester Examination

Date: 20th March, 2024 Time: 7.30AM to 9.00AM

1. (a) Give the parameter update expressions (no need to derive them) for each of the following methods: (i) Generalized delta rule, (ii) AdaGrad, (iii) RMSProp, (iv) AdaM. Define the terms used in the update expressions. Explain why the AdaM method is expected to give the best convergence.

   (b) Explain the phenomenon of internal covariant shift. Explain the post-activation function batch normalization that can be used to mitigate the effect of the internal covariant shift. (No need to derive the parameter update expressions).

   (3 + 2 = 5 Marks)

2. (a) Give the structure of an autoassociative neural network (AANN) that can be used to obtain a nonlinearly compressed representation of the input, without any restriction on the range of values that the compressed representation can take. The structure needs to be specified in terms of the layers, the relation between the number of nodes in the layers and the activation function used in each layer.

   (b) Explain how a 4-level stacked autoencoder built with AANNs as in Part (a) can be used to pre-train a deep feedforward neural network (DFNN).

   (c) Give the structure of the DFNN after merging of layers, for which the fine-tuning is to be done.

   (1 + 1 + 1 = 3 Marks)

3. Consider a shallow convolutional neural network built for classification of grey level images with the image size of WxH pixels. The convolution layer has D feature maps and uses kernels of size FxF. The stride parameter is 1 and there is no padding. The pooling layer does max pooling on 2x2 receptive fields with no overlap. The network has a full connected layer with L nodes, between the pooling layer and the output layer. The hyperbolic tangent function is used as the activation function of neurons in the convolution layer and the fully connected layer. The output layer has K neurons that use the logistic sigmoid activation function. The model is trained in the pattern mode, with the sum-of-squared errors as the loss function. The delta rule is used for the weight update.

(a) Derive an expression for the number of connections that have weight (other than bias) parameters associated with them, and an expression for the number of weight parameters (excluding bias parameters) to be learnt.

(b) Derive the weight update expression for the weight w_{lk} of a connection from lth node in the fully connected layer to kth node in the output layer.

(c) Derive the weight update expression for the weight w_{mijl} of a connection from the neuron at position (i, j) in the mth feature map in the convolutional layer to the lth node in the fully connected layer.

(d) Derive the weight update expression for the weight w_{pq}^{m} of a connection from the position (p, q) in the receptive field in the input image to the mth feature map in the convolutional layer.

(2 + 2 + 2 +2 = 8 Marks)

4. Give a diagram showing the typical structure of an inception module used in GoogLe Net. State the main purpose of the inception module and its components. State how the kernel factorization method can be used to reduce the number of parameters in the inception module.

(3 Marks)

5. (a) Explain the soft-assignment based VLAD aggregation method for an image.

   (b) Explain the NetVLAD aggregation method for an image.

(1 + 2 = 3 Marks)

6. Explain each of the following methods to obtain word embeddings:

   (a) Singular value decomposition (SVD) of PPMI matrix

   (b) Skip-gram model with negative sampling (Word2Vec)

   (c) GloVe method

(1 + 1 + 1 = 3 Marks)