EE 626 : End Semester Examination

Duration: 3 hours **Marks: 40** **Date: May 4' 2024**

1. A set of 32 input feature maps of size 64×64 are passed through a convolution layer **C1** comprising 16 filters of size 1×1 . Thereafter, the resulting feature maps are processed via another convolution layer **C2** comprising 256 5×5 filters. Subsequent to this, a max pooling layer **M1** is introduced. Adequate padding is provided together with a stride of 1 during the process of convolution and pooling operations. Based on this information, answer the following questions.

   (a) What is the size of each feature map in the layers **C1** and **C2**.

   (b) Compute the number of filter parameters (excluding bias) to be learnt in
      (i) **C1**
      (ii) **C2**
      (iii) **M1**.

   (c) Had the 32 input feature maps of size 64×64 were to be passed through **C2** and **M1**, what will be the number of filter parameters to be learnt (excluding bias).

   [2 + 3 + 2 = 7 marks]

2. A multi layer perceptron is trained to recognize the digits in $S = \{1, 2, 3\}$ by using a batch of samples. The targets are encoded using the one-hot scheme, whereby the i^{th} neuron in the output layer can provide the probability for the i^{th} digit in S. A simple stochastic gradient descent scheme is adopted for back-propagation

   (a) If the feature vector representing the digit 2 is sent as input, what will the target vector **t**

   (b) If the predicted values at the output neurons are $(1, 1.5, 2)$ respectively, use softmax function to map them to probability values.

   (c) By considering the targets and mapped probability values in (a) and (b), compute the following loss.
      (i) multi-category cross entropy
      (ii) sum of squared error

   64 − 5 + 1

   60

   [1 + 1 + 3 = 5 marks]

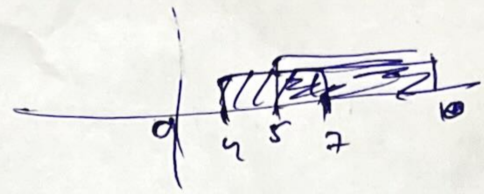
3. (a) Find w that minimizes the objective function: $E(w) = (w-3)^2 + 1$ using the Gradient Descent algorithm with momentum. Consider using 90% of the previous update at each iteration (starting from the second). Initialize w with -1 and use learning rate = 0.3. Run for 3 iterations

   (b) If the Ada grad version of gradient descent is used to minimize the function $E(w) = (3w-2)^4 + 1$, find the resulting value of w after 3 iterations. Initialize w with -2 and use learning rate = 0.3.

   [3+3 = 6 marks]

1.392 −1.728

12(3w−2) −96



4. In a signature verification system, there are two categories : forgery ω_1 and genuine ω_2. The idea is to select a threshold, so that samples with scores less than this value can be regarded as being verified as genuine. It is known that the scores of the genuine and forgery follow a uniform distribution $U(4,7)$ and $U(5,10)$ respectively. Assuming a decision threshold of 6 and prior $P(\omega_1) = 0.6$, with zero one loss function, compute the

(a) average probability of a signature to be accepted as a genuine. →0.66

(b) average probability of a signature to be correctly accepted as a genuine. → 0.46 0.66

(c) average probability of a signature to be rejected as a forgery. → 0.8

(d) average probability of a signature to be correctly rejected as a forgery. → 0.6

0.66

[8 marks]

5. Consider a two-dimensional data set $\mathbf{D}$, obtained by pooling the training samples from 2 classes ω_1 and ω_2. It is desired to learn three centroids by means of a k-means algorithm.

$$\mathbf{D} = \left\{ \begin{bmatrix} -5 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 6 \end{bmatrix}, \begin{bmatrix} -1 \\ -7 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \end{bmatrix}, \begin{bmatrix} 6 \\ -1 \end{bmatrix} \right\}$$

The algorithm is initialized with cluster means: $\mu_1 = \begin{bmatrix} -7 \\ 4 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$ and $\mu_3 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$.

(a) What are the values of the centroids at convergence?

(b) What is the minimum number of iterations required for convergence

[3 + 1 = 4 marks]

6. Consider a multi-layer perceptron with the following specifications:

- d nodes in the input layer with sigmoidal activation functions at the M nodes in the hidden layer.
- linear activation functions at the c nodes in the output layer.

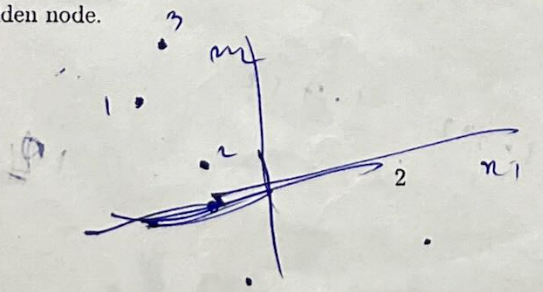
The weights of the network are to be learnt with N training samples $\{(\mathbf{x}_i, t_i)\}_{i=1}^{N}$ using the error function

$$E = \sum_{n=1}^{N} E^n = \frac{1}{6} \sum_{n=1}^{N} \sum_{k=1}^{c} (t_k^n - y_k^n)^6$$

(a) Derive an expression for the predicted output $y_k^n = y_k(\mathbf{x}_n, \mathbf{w})$ in terms of the input $\mathbf{x}_n$, activation functions and network weights.

(b) Derive an expression for the updating of weight w_{kj} , connecting the j^{th} hidden node to the k^{th} output node.

(c) Derive an expression for the updating of weight w_{ji} , connecting the i^{th} input node to the j^{th} hidden node.



[3 + 3 + 4 = 10 marks]