ISE 789/OR 791 Homework #4

Issued: April 2, 2024 Due: April 15, 2024

Problem 1: (30 points) Consider using a 3-layer input-hidden-output shallow network as we studied in the class (Lecture 8, Slide #29) for on-line (example by example) learning.
Derive the "gradient information" of the error function with respect to the weights (Slide #33) for adopting (i)Sigmoid function, (ii) Hyperbolic Tangent function, (iii) Rectified Liner Unit function as the activation function of your neural network.

Problem 2: (70 points) We have learned the basics of artificial neural networks in our class. In this problem, you are expected to learn more about designing a "good" neural network.

(1) Given that $y = f(x) = \frac{\sin x+1}{2}$ and $z = f(2x) = \frac{\sin 2x+1}{2}$, please plot the graphs of $y = f(x)$ and $z = f(2x)$ for $x \in [0, 2\pi]$.

(2) Randomly generate a set of 150 data points $\mathbf{Y} = \{(x_i, y_i) \mid i = 1, \ldots, 150\}$ and another set of 150 data points $\mathbf{Z} = \{(x_i, z_i) \mid i = 1, \ldots, 150\}$ where x_i's are uniformly distributed over $[0, 2\pi]$.

(3) Evenly divide $\mathbf{Y}$ into 5 portions such that $\mathbf{Y}_1 = \{(x_i, y_i) \mid i = 1, \ldots, 30\}$, $\mathbf{Y}_2 = \{(x_i, y_i) \mid i = 31, \ldots, 60\}$, $\ldots$, $\mathbf{Y}_5 = \{(x_i, y_i) \mid i = 121, \ldots, 150\}$. Follow the same way to divide $\mathbf{Z}$ into 5 portions of $\mathbf{Z}_1, \ldots, \mathbf{Z}_5$.

(4) Consider 4 possible artificial neural networks:

(N1) is a "shallow" neural net with 1 input node,
2 nodes in one hidden layer,
and 1 output node.

(N2) is also a "shallow" neural net with 1 input node,
4 nodes in one hidden layer,
and 1 output node.

(N3) is a "deep" neural net with 1 input node,
2 nodes in the first hidden layer,
2 nodes in the second hidden layer,
and 1 output node.

(N4) is a "deeper" neural net with 1 input node,
1 node in the first hidden layer,
2 nodes in the second hidden layer,
1 node in the third hidden layer,
and 1 output node.

Please draw the diagram of each of these four neural networks.

(5) You may use any neural network software package (or write your own programs) to conduct the following experiments on (N1), (N2), (N3) and (N4), respectively:

(E1): (a) Use $\mathbf{Y}_1$ as the training data set and $\mathbf{Y}_5$ as the validation data set.
(b) Use $\mathbf{Z}_1$ as the training data set and $\mathbf{Z}_5$ as the validation data set.

(E2): (a) Use $\mathbf{Y}_1 \cup \mathbf{Y}_2$ as the training data set and $\mathbf{Y}_5$ as the validation data set.
(b) Use $\mathbf{Z}_1 \cup \mathbf{Z}_2$ as the training data set and $\mathbf{Z}_5$ as the validation data set.

(E3): (a) Use $\mathbf{Y}_1 \cup \mathbf{Y}_2 \cup \mathbf{Y}_3$ as the training data set and $\mathbf{Y}_5$ as the validation data set.
(b) Use $\mathbf{Z}_1 \cup \mathbf{Z}_2 \cup \mathbf{Z}_3$ as the training data set and $\mathbf{Z}_5$ as the validation data set.

(E4): (a) Use $\mathbf{Y}_1 \cup \mathbf{Y}_2 \cup \mathbf{Y}_3 \cup \mathbf{Y}_4$ vs. $\mathbf{Y}_5$.
(b) Use $\mathbf{Z}_1 \cup \mathbf{Z}_2 \cup \mathbf{Z}_3 \cup \mathbf{Z}_4$ vs. $\mathbf{Z}_5$.

For each experiment, record the mean squared error (MSE) over each validation data set.

(6) Which activation function will you choose for your experiment? Why?

(7) In this assignment, we only care about ANN's best power of approximation. Generalization is not a concern. Which stopping rule will you use in your experiment? Why?

(8) Carefully analyze the results you obtained in (5), and report your findings as insightful as possible. (*Hint: # of nodes, shallow vs. deep, training data sets, design rules, ...*)