CS 422: Data Mining

Department of Computer Science
Illinois Institute of Technology
Vijay K. Gurbani, Ph.D.

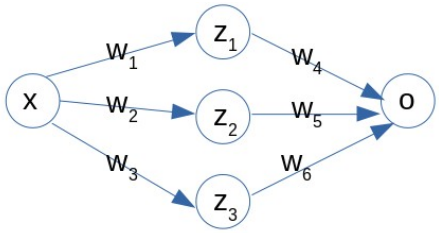
Fall 2024: Homework 5

Due date: Thursday Oct 31, 2024, 11:59:59 PM Chicago Time

Please read all of the parts of the homework carefully before attempting any question. If you detect any ambiguities in the instructions, please let me know right away instead of waiting until after the homework has been graded.

1. Exercises (Please submit a PDF file containing answers to these questions. Any other file format will lead to a loss of 0.5 point. Non-PDF files that cannot be opened by the TAs will lead to a loss of all points.)

1.1 (2 points) Consider the neural network shown in the figure below.



The weight matrix, W, is: $[1, 1, -1, 0.5, 1, 2]^T$. Assume that the hidden layer uses RelU and the output layer uses Sigmoid activation function. Assume squared error. The input x = 4, and the output y = 0.

Recall that RelU is defined as a function, f(x) = max(0, x). Its derivative is $f'(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$

Squared error is defined as $E(y, \hat{y}) = (y - \hat{y})^2$.

The partial derivative of E with respect to $\hat{y}$ is $-2(y - \hat{y})$.

Using this information, answer the following questions:
(Show all work, and all answers should be rounded to 3 decimal places OR POINTS WILL BE TAKEN OFF!)
(a) Use forward propagation to compute the predicted output.
(b) What is the loss or error value?
(c) Using backpropagation, compute the gradient of the weight vector, that is, compute the partial derivative of the error with respect to all of the weights.
(d) Using a learning rate of 1.0, compute new weights from the gradient. With the new weights, use forward propagation to compute the new predicted output, and the loss (error).
(e) Comment on the difference between the loss values you observe in (b) and (d).

1.2 [1 point] Tan Chapter 4, questions 14, 15 (b).

1.3 [1 point] Consider a dataset that has 8 predictors. You train a neural network with 3 hidden layers and an output layer that predicts a continuous value (a regression problem). The first hidden layer has 16 neurons, the second has 8 neurons, and the third has 4 neurons. In this network, how many total parameters will you have?

2 Practicum problems

Feed Forward Neural Networks

The dataset available for this problem on Blackboard (wifi_localization.csv) is described in detail at https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization. It contains signal for 7 WiFi access points, which are used to predict a user's location in one of four rooms. The response variable, therefore, is a room number ranging from 1-4. Using this dataset, you will train models to predict the user's location. 80% of the data will be used for training and 20% for testing.

Use the seed value of **1122** before you divide the data into a train and test set.

(a) [1 point] Train a **decision tree model** on the training set using all predictors, and fit the test set to the model. Create a confusion matrix from fitting the held out test set. Your function should return a confusion matrix object. (Examine the tree by plotting it; **DO NOT** submit the plot. Plotting it so you can see its decisions is instructive.)

(b) You will now train a neural network to predict the user's location in one of the four rooms. Loss is measured in terms of accuracy, using 'categorical_crossentropy' on the 'adam' optimizer (see the compile() function). For the fit() method, set the batch_size to 32 and the validation_split to 0.20. Train the network for 100 epochs.

For this question, use one hidden layer, and only one neuron in the hidden layer. Use the relu activation function for the neuron in the hidden layer.

(i) **[0.5 points]** Write a function that creates and trains a neural network as described above and returns the model object after training.

(ii) **[0.2 points]** Fit the held out test dataset to the model and create a confusion matrix. Your function should return a confusion matrix object.

(iii) **[0.2 points]** Examine the loss and accuracy of the model when it is fit to the test dataset. Write a function to return the loss and accuracy from fitting the test dataset. The function should return an unnamed vector, the first element of the vector is the loss (numeric, rounded to two decimal digits of accuracy) and the second element of the vector is the accuracy (numeric, rounded to two decimal digits of accuracy).

(iv) **[0.5 points]** Examine the values for loss and accuracy you got above. Given the shallow network you are using, there are four possible outcomes:

1. High value for accuracy, high value for loss.
2. High value for accuracy, low value for loss.
3. Low value for accuracy, high value for loss.
4. Low value for accuracy, low value for loss.

Where in the above spectrum does your loss and accuracy fall? Do the loss and accuracy meet your expectation? Please explain your answer in no more than 1-2 sentences. Put the answer to this question in the PDF you submit. Label the answer with "Q2(b)(ii) ..."

(v) **[0.3 points]** Examine the distribution of the predicted labels. What pattern do you see in the predicted labels? Please explain your answer in no more than 2-3 sentences. Put the answer to this question in the PDF you submit. Label the answer with "Q2(b)(iv) ..."

(vi) **[0.3 points]** Do you think we will get better results if we increase the training to 200 epochs? Why? Put the answer to this question in the PDF you submit. Label the answer with "Q2(b)(v) ..."

NOTE: Before starting the next question below, if your model was being stored in a global variable, make sure you null out the variable. The reason is that if you don't null it out, the global model object will have residual information left over from previous use and your results will not be quite accurate.

(c) We will now attempt to improve the neural network so it performs better (high accuracy during prediction and low loss).

(i) **[1 point]** Improve your neural network by adding one hidden layer (do not add more than one hidden layer). Experiment with the number of neurons you need in this hidden layer and play with the choice of the activation function for the neurons in the hidden layer. You aim is to get an accuracy ≥ 0.95 and loss $0.00 \leq \text{loss} \leq 0.13$.

Use the same training and test set used in 2(b). Train your model on the training set. As before, loss is measured in terms of accuracy, using 'categorical_crossentropy' on the 'adam' optimizer (see the compile() function). For the fit() method, set the batch_size to 32 and the validation_split to 0.20. Train the network for 100 epochs.

Your function should return the model object after training.

(ii) **[0.375 points]** Fit the held out test dataset to the model and create a confusion matrix. Your function should return a confusion matrix object.

(iii) **[0.375 points]** Examine the loss and accuracy of the model when it is fit to the test dataset. Write a function to return the loss and accuracy from fitting the test dataset. The function should return an unnamed vector, the first element of the vector is the loss (numeric, rounded to two decimal digits of accuracy) and the second element of the vector is the accuracy (numeric, rounded to two decimal digits of accuracy).

(iv) **[0.375 points]** Compare the loss and accuracy of your new model with the loss and accuracy obtained from the model in 2(b). Is it better or worse? Provide a rationale for your answer (1-2 sentences), and put the answer to this question in the PDF you submit. Label the answer with "Q2(c)(iv) …"

(v) **[0.375 points]** Based on the plots of accuracy and validation during testing, at what epoch do you think we should stop the training to minimize over-fitting? Your answer should be a number between 0 and the maximum epochs used during training. The function should return this answer as a numeric value.

(vi) **[0.5 points]** Study the confusion matrix you created in 2(c)(ii) and compare it with the confusion matrix you created for the decision tree model in in 2(a). Based on the confusion matrices and other knowledge you have gained in class on neural networks, if you had to deploy the model in 2(c) or the model in 2(a) for production, which one will you choose and why? Provide a rationale for your answer (1-2 sentences), and put the answer to this question in the PDF you submit. Label the answer with "Q2(c)(vi) …"