CS 550 -- Machine Learning
Homework #2

Due: 23:55, December 7, 2020

In this homework, implement an artificial neural network (ANN) regressor and learn its weights implementing the backpropagation algorithm. Your implementation should support an ANN for linear regression (no hidden layer) and an ANN with a single hidden layer. You will use these ANNs in your experiments.

After implementing your ANN regressors, test them on the two datasets that are provided on the course web page. The instances in both of these datasets have 1-D inputs and 1-D outputs (so that you will easily plot and see what your ANNs learn). The datasets are provided as four text files (train1, test1, train2, test2). An individual line of each file corresponds to an instance where the first number corresponds to the input and the next one corresponds to the output.

While testing your implementation, follow the steps below:

(a) For each dataset, first find the configuration that learns a network on its training instances. When selecting your configuration, you may want to answer the following questions. Note that each question might affect whether or not your network learns.

- Is it sufficient to use a linear regressor or is it necessary to use an ANN with a single hidden layer? If it is latter, what will be the minimum number of hidden units?
- What is your activation function to define the hidden units?
- What is your loss function?
- What is a good value for the learning rate?
- How to initialize the weights?
- How many epochs should you use? How to decide when to stop?
- Is it necessary to use momentum?
- Will you use a stochastic or a batch learning algorithm?
- Does normalization affect the learning process for this application?

After finding such a configuration, plot the actual outputs for the given input points. Then draw a curve for the outputs estimated by your selected model (on the same plot). While drawing this estimated output curve, do not just use the given input points, but draw a curve for data points uniformly selected in the range of your input points (so that you can see a smooth curve). Provide these plots for training and test sets. Here you will have two plots for each of the datasets.

Additionally, for the data points uniformly selected in the range of your input points, draw a curve for each of the hidden units (put all these curves on the same plot). It is adequate to provide these curves on the training sets. Provide these curves in a separate plot for each dataset.

Then, for each dataset, write the answers to the aforementioned questions in the following format.

ANN used (specify the number of hidden units):
Selected activation function:
Selected loss function:
Learning rate:
Range of initial weights:
Number of epochs:
When to stop:
Is momentum used (if so, value of the momentum factor):
Is normalization used:
Stochastic or batch learning:
Training loss (averaged over training instances):
Test loss (averaged over test instances):

(b) The following experiments will help you understand how learning occurs and what factors affect the learning process.

For each of the experiments required in (c)–(f),

- If you are asked to provide plots, provide them on the training instances. You need to provide two sets of plots, one set for the first dataset and the other for the second dataset. For each plot, draw the actual outputs for the given input points. Then draw a curve for the outputs estimated by a model on the same plot. As explained before, while drawing this estimated output curve, do not just use the given input points, but draw a curve for data points uniformly selected in the range of your input points (so that you can see a smooth curve). Note that here you will end up with many plots. You need to put them "nicely" to your report, sticking to the page limit. For example, you may have a row of five small plots to observe the effect of a factor. For this observation, you just need to see what the estimated curves look like (you do not need to see the estimated values in detail, which means that you do not need to put a plot into your report as a large-size figure).
- If you are asked to report the training (or the test) set loss, report the loss averaged over the training (or the test) instances and its standard deviation. Put these values in a table. Format the numbers "nicely" (for example, do not use 6-10 digits after the decimal point).

(c) For each dataset, use the linear regressor and ANNs with a single hidden layer. Use 2, 4, 8, and 16 hidden units. For each ANN, provide the plots as explained in (b). Report the training and test set losses.

Note that for each ANN, you may need to use a different configuration. That is, try different configurations (different learning rates, different ranges of initial weights, etc.) when learning these ANNs.

Observe how the complexity (number of hidden units) of an ANN determines its ability to learn. Summarize your observations in 6-10 sentences.

(d) *Only for the second dataset*, use an ANN with a single hidden layer containing 8 hidden units. Then, run your ANN by selecting your learning rates as 1, 0.1, 0.01, 0.001, and 0.0001. For each run, calculate the average loss on training instances at the end of each epoch and stop if this average loss is below a threshold. Select the threshold based on your findings in (c) and report the selected threshold value. Then, for each run, provide a plot as explained in (b) and report the number of epochs that you reach the threshold. Compare your results in 3-5 sentences.

(e) *Only for the second dataset*, use an ANN with a single hidden layer containing 8 hidden units. Similar to (d), calculate the average loss on training instances at the end of each epoch and stop

if this average loss is below the threshold that you select for (d). Then, run your ANN with and without momentum. For each, provide a plot as explained in (b) and report the training set loss and the number of epochs that you reach the selected threshold. Compare your results in 3-5 sentences.

(f) *Only for the second dataset*, use an ANN with a single hidden layer containing 8 hidden units. First use a stochastic learning algorithm, then use a batch learning algorithm. For each, find a configuration that leads to good results. Are there any differences between these configurations? If so, explain all of the different choices that you need to make (e.g., the learning rates might be different for batch and stochastic learning).

This homework asks you to implement a neural network regressor by **writing your own codes**. Thus, you are not allowed using any machine learning package. In your implementation, you may use any programming language you would like.

You are expected to write your report neatly and properly. The format, structure, and writing style of your report as well as the quality of the tables and figures will be a part of your grade. Use reasonable font sizes, spacing, margin sizes, etc. You may submit either a one-column or a double-column document. In your report, do not give any screen shots (except the plots). Do not forget to address the questions specifically asked to you. Your report should be a maximum of 5 pages.

Please email the pdf of your report and the source code of your implementation before the deadline. The subject line of your email should CS 550: HW2.