

Assignment 1

CE/CZ4042: Neural Networks and Deep Learning

Deadline: 16th October 2020

- ✓ The project is to be done individually.
- ✓ Need to complete both parts A and B of the assignment and submit a report and source codes online via NTULearn before the deadline.
- ✓ Data files for both parts are found in Assignment 1 folder under Assignments on NTULearn. You can use starter codes start_1a.py and start_1b.py to begin the assignment.
- ✓ The report would contain sections such as (i) introduction, (ii) methods, (iii) experiments and results, and (iv) conclusions. The conclusion is to include the answers to all the queries and a discussion on the results.
- ✓ The assessment will be based on both the project report and the correctness of the codes submitted.
- ✓ 10 marks are reserved for correctness, extra analysis, and presentation providing intuitive figures.
- ✓ Late submissions will be penalized: 5% for each day up to three days.
- ✓ The report has to be submitted in the following format:
 - lastname_firstname_A1_report.pdf (report in pdf format); and
 - lastname_firstname_A1_codes.zip (all the source codes)
- ✓ TA Dr. Charmara Kasun (chamarakasun@ntu.edu.sg) is in charge of this assignment. Please email or see him at the SCALE Lab (N4-B3b-13) during his office hours: Friday 3:30 P.M. – 5:30 P.M., in case you face issues.

Part A: Classification Problem

This project aims at building neural networks to classify the Cardiotocography dataset containing measurements of fetal heart rate (FHR) and uterine contraction (UC) features on 2126 cardiotocograms classified by expert obstetricians [1]. The dataset can be obtained from: <https://archive.ics.uci.edu/ml/datasets/Cardiotocography>.

The cardiotocograms were classified by three expert obstetricians and a consensus classification label with respect to a morphologic pattern and to a fetal state (N: Normal; S: Suspect; P: Pathologic) was assigned to each of them. The aim is to predict the N, S and P class labels in the test dataset after training the neural network on the training dataset.

Read the data from the file: ctg_data_cleaned.csv. Each data sample is a row of 23 values: 21 input attributes and 2 class labels (use the NSP label with values 1, 2 and 3 and ignore the other). First, divide the dataset in 70:30 ratio for training and testing. Use 5-fold cross-validation on the training dataset for selecting the optimal model, and test it on the testing data.

1. Design a feedforward neural network which consists of an input layer, one hidden layer of 10 neurons with ReLU activation function, and an output softmax layer. Assume a learning rate $\alpha = 0.01$, L_2 regularization with weight decay parameter $\beta = 10^{-6}$, and batch size = 32. Use appropriate scaling of input features.
 - a) Use the training dataset to train the model and plot accuracies on training and testing data against training epochs.
 - b) State the approximate number of epochs where the test error begin to converge.

(10 marks)
2. Find the optimal batch size by training the neural network and evaluating the performances for different batch sizes.
 - a) Plot cross-validation accuracies against the number of epochs for different batch sizes. Limit search space to batch sizes $\{4, 8, 16, 32, 64\}$. Plot the time taken to train the network for one epoch against different batch sizes.
 - b) Select the optimal batch size and state reasons for your selection.
 - c) Plot the train and test accuracies against epochs for the optimal batch size.

Note: use this optimal batch size for the rest of the experiments.

(9 marks)
3. Find the optimal number of hidden neurons for the 3-layer network designed in part (2).
 - a) Plot the cross-validation accuracies against the number of epochs for different number of hidden-layer neurons. Limit the search space of number of neurons to $\{5, 10, 15, 20, 25\}$.
 - b) Select the optimal number of neurons for the hidden layer. State the rationale for your selection.
 - c) Plot the train and test accuracies against epochs with the optimal number of neurons.

(9 marks)

4. Find the optimal decay parameter for the 3-layer network designed with optimal hidden neurons in part (3).
 - a) Plot cross-validation accuracies against the number of epochs for the 3-layer network for different values of decay parameters. Limit the search space of decay parameters to $\{0, 10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}\}$.
 - b) Select the optimal decay parameter. State the rationale for your selection.
 - c) Plot the train and test accuracies against epochs for the optimal decay parameter. (9 marks)
5. After you are done with the 3-layer network, design a 4-layer network with two hidden-layers, each consisting 10 neurons, and train it with a batch size of 32 and decay parameter 10^{-6} .
 - a) Plot the train and test accuracy of the 4-layer network.
 - b) Compare and comment on the performances of the optimal 3-layer and 4-layer networks. (10 marks)

Hint: Sample code is given in file 'start_1a.py' to help you get started with this problem.

Part B: Regression Problem

This assignment uses the data from the Graduate Admissions Predication [2]. The dataset contains several parameters like GRE score (out of 340), TOEFL score (out of 120), university Rating (out of 5), strengths of Statement of Purpose and Letter of Recommendation (out of 5), undergraduate GPA (out of 10), research experience (either 0 or 1), that are considered important during the application for Master Programs. The predicted parameter is the chance of getting an admit (ranging from 0 to 1). You can obtain the data from:

<https://www.kaggle.com/mohansacharya/graduate-admissions>

or from file 'admission_predict.csv'.

Each data sample is a row of 9 values: 1 serial number (ignore), 7 input attributes and the probability of getting an admit as targets. Divide the dataset at 70:30 ratio for training and testing.

1. Design a 3-layer feedforward neural network consists of an input layer, a hidden-layer of 10 neurons having ReLU activation functions, and a linear output layer. Use mini-batch gradient descent with a batch size = 8, L_2 regularization at weight decay parameter $\beta = 10^{-3}$ and a learning rate $\alpha = 10^{-3}$ to train the network.
 - a) Use the train dataset to train the model and plot both the train and test errors against epochs.
 - b) State the approximate number of epochs where the test error is minimum and use it to stop training.
 - c) Plot the predicted values and target values for any 50 test samples.

(12 marks)

2. Recursive feature elimination (RFE) is a feature selection method that removes unnecessary features from the inputs. Start by removing one input feature that causes the minimum drop (or maximum improvement) in performance. Repeat the procedure recursively on the reduced input set until the optimal number of input features is reached. Remove the features one at a time. Compare the accuracy of the model with all input features, with models using 6 input features and 5 input features selected using RFE. Comment on the observations.

(15 marks)

3. Design a four-layer neural network and a five-layer neural network, with the hidden layers having 50 neurons each. Use a learning rate of 10^{-3} for all layers and optimal feature set selected in part (3).

Introduce dropouts (with a keep probability of 0.8) to the layers and report the accuracies. Compare the performances of all the networks (with and without dropouts) with each other and with the 3-layer network.

(16 marks)

Hint: Sample code is given in file 'start_1b.py' to help you get started with this problem.

References:

- [1] Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J Matern Fetal Med 5:311-318.
- [2] Mohan S Acharya, Asfia Armaan, Aneeta S Antony: A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019.