

**CZ 4042**

**Neural Networks and Deep Learning**

**Assignment 1 Report**

**Tan Chuan Xin**

**U1821755B**

Table of Contents

[Part A: Classification Problem 3](#_Toc53840412)

[Introduction 3](#_Toc53840413)

[Methods 3](#_Toc53840414)

[Experiments 3](#_Toc53840415)

[Results 3](#_Toc53840416)

[Conclusions 3](#_Toc53840417)

[Question 1 (10 marks) 3](#_Toc53840418)

[Question 2 (9 marks) 3](#_Toc53840419)

[Question 3 (9 marks) 4](#_Toc53840420)

[Question 4 (9 marks) 4](#_Toc53840421)

[Question 5 (10 marks) 4](#_Toc53840422)

[Part B: Regression Problem 6](#_Toc53840423)

[Introduction 6](#_Toc53840424)

[Methods 6](#_Toc53840425)

[Experiments 6](#_Toc53840426)

[Results 6](#_Toc53840427)

[Conclusions 6](#_Toc53840428)

[Question 1 (12 marks) 6](#_Toc53840429)

[Question 2 (15 marks) 6](#_Toc53840430)

[Question 3 (16 marks) 6](#_Toc53840431)

Prologue

Part A: Classification Problem

# Introduction

# Methods

# Experiments

# Results

# Conclusions

## Question 1 (10 marks)

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 𝛼 = 0.01, L2 regularization with weight decay parameter 𝛽 = 10−6 , and batch size = 32. Use appropriate scaling of input features.
   1. Use the training dataset to train the model and plot accuracies on training and testing data against training epochs.
   2. State the approximate number of epochs where the test error begin to converge.

## Question 2 (9 marks)

1. Find the optimal batch size by training the neural network and evaluating the performances for different batch sizes.
   1. 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.
   2. Select the optimal batch size and state reasons for your selection.
   3. 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.

## Question 3 (9 marks)

1. Find the optimal number of hidden neurons for the 3-layer network designed in part (2).
   1. 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}.
   2. Select the optimal number of neurons for the hidden layer. State the rationale for your selection.
   3. Plot the train and test accuracies against epochs with the optimal number of neurons.

## Question 4 (9 marks)

1. Find the optimal decay parameter for the 3-layer network designed with optimal hidden neurons in part (3).
   1. 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}.
   2. Select the optimal decay parameter. State the rationale for your selection.
   3. Plot the train and test accuracies against epochs for the optimal decay parameter.

## Question 5 (10 marks)

1. 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 .
   1. Plot the train and test accuracy of the 4-layer network.
   2. Compare and comment on the performances of the optimal 3-layer and 4-layer networks.

Part B: Regression Problem

# Introduction

# Methods

# Experiments

# Results

# Conclusions

## Question 1 (12 marks)

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, 𝐿2 regularization at weight decay parameter 𝛽 = 10−3 and a learning rate 𝛼 = 10−3 to train the network.
   1. Use the train dataset to train the model and plot both the train and test errors against epochs.
   2. State the approximate number of epochs where the test error is minimum and use it to stop training.
   3. Plot the predicted values and target values for any 50 test samples.

## Question 2 (15 marks)

1. 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.

## Question 3 (16 marks)

1. 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.