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, *L*2 regularization with weight decay parameter 𝛽 = 10−6, and batch size = 32. Use appropriate scaling of input features.

a) Use the training dataset to train the model and plot both accuracies on training and testing data against epochs.

b) State the approximate number of epochs where the test error converges.

* 1. 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 to {4, 8, 16, 32, 64}. Plot the time taken to train the network for one epoch against different batch sizes.
  2. b) Select the optimal batch size and state reasons for your selection.
  3. 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.

* 1. 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}.
  2. b) Select the optimal number of neurons for the hidden layer. State the rationale for your selection.
  3. c) Plot the train and test accuracies against epochs with the optimal number of neurons.
  4. 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 to decay parameters to {0, 10−3, 10−6, 10−9, 10−12}.
  5. b) Select the optimal decay parameter. State the rationale for your selection.
  6. c) Plot the train and test accuracies against epochs for the optimal decay parameter.

* 1. *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.
  2. b) Compare and comment on the performances of the optimal 3-layer and 4-layer networks.