Artificial Neural Networks

Final Exam

Deadline: December 16th – 9:00 AM

You have 1 week to submit the solutions on Harvey.

For each question, **type a report in word** that shows your **answer** and your **results**. Please save your word file as **PDF** and ***submit your PDF + all your source codes*** in separate folders named for example "Q1-a" , "Q1-b", …

The exam has 100 points.

Question 1: 40 points

Question 2: 40 points

Question 3: 20 points

Question 1 – Restricted Boltzmann Machines:

In our class, we saw how to implement a deep belief network using 3 RBMs for classification tasks. We also worked with the MNIST dataset for the classification of handwritten digits 0,1,…,9 (with 60000 images for training and 10000 images for testing).

a) Using the packages discussed in the class, implement a deep belief network with L number of RBMs. Change L, try different values 2,3,4,…,8, and show the training and testing accuracy plot and loss plot. Find the best L and justify your answer.

b) implement one Gaussian RBM from scratch without using any RBM packages (you can use standard packages such as Numpy, pandas, Matplotlib, …). Use your RBM to learn unsupervised features of the training set images in the MNIST dataset (60000 images). Draw the features learned by your Gaussian RBM (visualize the hidden layer h) for 30 random images in the dataset. Did your RBM learn anything interesting about the dataset?

c) implement the classic RBM (not Gaussian RBM) from scratch using no RBM packages (you can use standard packages such as Numpy, pandas, Matplotlib, …). Build a 3-layer, a 4-layer, and a 5-layer deep belief network and classify the MNIST dataset. Report the training loss and accuracy as well as the testing loss and accuracy. Compare the results with a 3-layer, a 4-layer, and a 5-layer stacked autoencoder.

Question 2 – Autoencoder and Convolution Neural Networks:

The CIFAR-10 dataset consists of 60000 32x32 color images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. Download the dataset using this link:

https://www.cs.toronto.edu/~kriz/cifar.html

(a) Plot 5 random images inside the training set and 5 random images inside the testing set.

(b) Train a convolution neural network in Keras with L number of layers. Each “layer” is a convolution+ReLU+Pooling. Set L to be 2,3, and 4. The Pooling can be Max Pooling or Average Pooling. Plot the training and testing loss and accuracy as a function of number of training epochs. Report the results of all 6 settings: L=2 with Max pooling, L=2 with Avg pooling, L=3 with Max pooling, L=3 with Avg pooling, L=4 with Max pooling, L=4 with Avg pooling.

What setting is the best setting? Why?

(c) How can you improve your optimal neural network in (b)? Please improve it and report the new results (compare with the optimal result you got in (b)). Explain how that change could lead to a better neural network.

(d) In our class, we discussed Autoencoders and Convolutional Neural Networks separately. Here, we want to implement a “Convolution Autoencoder (CAE)” which is an autoencoder that uses convolution and deconvolution layers to reconstruct the images. This way, the CAE will learn meaningful features from the image. These features can then be used by any classification method to classify the images. Please do your research on CAEs and implement a CAE using the convolution and deconvolution layers in Keras. Then, train the CAE in an unsupervised fashion (the CAE observes the images of CIFAR-10 and reconstructs them). When you are done, use the encoder part of the CAE and train a simple classifier (any classifier) to classify the images of CIFAR-10. Report the classification results (loss and accuracy) as a function of training epochs used for classification.

(e) Use L1 and L2 regularization in item d. These techniques are available in Keras. Do these techniques help you to get better results? Why? Plot the confusion matrix for L1 and L2 regularization results.

Question 3 – Recurrent Neural Networks:

The excel file “daily-min-temperatures.csv” describes the minimum daily temperatures over 10 years (1981-1990) in the city of Melbourne, Australia. The units are in degrees Celsius and there are 3650 observations. The source of the data is credited as the Australian Bureau of Meteorology. At each time t, we want to observe the measurements of:

t-k , t-k+1 , t-k+2 , …, t-1 , t

To predict the measurement at time t+1.

(a) Create a training set and a testing set for this problem. In your code, k should be a variable. You can assume 80% of the data is used for training and 20% for testing.

(b) Train an LSTM to predict the measurement at t+1, and report the mean squared error and mean absolute percentage error of the training and testing as a function of the number of epochs used. Do this for four different values of k.

(c) do your research and find what a Gated Recurrent Unit (GRU) is. What is the difference between LSTM and GRU? Implement a GRU for this question and compare its results with your LSTM in both the training and testing datasets.

(d) In our class, we saw how to implement a simple feed-forward neural network from scratch (without any neural network packages). Please use that code and develop a simple recurrent neural network (not an LSTM). Run item (b) using that simple recurrent neural network. Compare the results with your LSTM implemented in (b) for four different values of k.