Deep Learning

Final Exam
Deadline: May 9th – 11:59 PM

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", …

You can use any toolbox/package in any programming language to answer the questions.

The exam has 100 points.
Question 1: 60 points
Question 2: 40 points

Question 1 – Convolution Neural Networks:

The CIFAR-10 dataset is a famous computer vision dataset. It 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) Randomly choose 1000 images from the training set and put them in a 1000*3072-dimensional data matrix A . Note that 3072 is 32*32*3 where 32 is the height and weight of the images and 3 is the number of colors (channels). Run PCA dimensionality reduction method on data matrix A and obtain the top 120 principal components (directions/components which show highest variation in the data). Report those components and show how much of data variation is represented by those 120 directions.

c) 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?

Question 2 – 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. This problem is similar to the prediction problems we solved in the class.

a) Create a training set and a testing set for this problem. In your code, k should be an input 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.