Advanced Machine Learning **Name:** _____
Spring 2022
Final Exam
4/30/2022
Time Limit: 1:00 pm - 3:00 pm
Instructor: Dr. Anahita Zarei

You need to do the following problems in R or Python and submit your rmd/jupyter notebook, AND html files on Canvas.

If your problem involves any randomization please set the seed to 123.

Please note that there's no restriction on the number of hidden layers, neurons, hyper-parameters, etc. in the following problems. But you need to briefly explain the rational behind the choices you have made in your solution. Your grade is based on different factors such as preprocessing, model implementation, appropriate plots, and model accuracy.

1. (20 points) This problem uses the file named "realestate.csv" posted on Canvas and contains information on real estate sales for a number of cities in the US.

   The features in the file are as follows:

   - inventory: Amount of time (number of months) it would take to sell all current listings at current pace of sales
   - year
   - listings: Total active listings
   - city: Name of MLS area
   - median: Median sale price

   Create a deep neural network to predict the median sale price based on the rest of the features and answer the following questions.

   (a) What is under-fitting? How did you make sure under-fitting didn't occur in your model implementation? Explain your answer with a relevant plot.

   (b) What is over-fitting? How did you make sure over-fitting didn't occur in your model implementation? Explain your answer with a relevant plot.

   (c) Report the unnormalized MAE for train, validation, and test set of your final model.

   (d) Create a single scatter plot that shows the prediction values (in blue), and the actual values (in red) for the test set.

2. (20 points) This problem uses the file named "salesData.csv" posted on Canvas and contains information on accumulative number of weekly sales for a company.

   Create a recurrent neural network model that uses the past 5 weeks historical data points to make predictions 4 weeks in the future. For example, you need to use the information

from week 1, 2, 3, 4, and 5 to predict the sales for week 9. You need to use 60% of the data for training, 20% for validation, and 20% for test.

(a) How did you make sure under-fitting didn't occur in your model implementation? Explain your answer with a relevant plot.

(b) How did you make sure over-fitting didn't occur in your model implementation? Explain your answer with a relevant plot.

(c) Report the unnormalized MAE for train, validation, and test set of your final model.

(d) Create a single scatter plot that shows the prediction values (in blue), and the actual values (in red) for the test set.