ISE 5123 **Software Tools-Dec Support**
Spring 2020

Final Exam
May 8, 2020

NAME _____

Instructions:

1. You may use books, notes, python codes, but all work is to be done on your own. The following statement must be typed in the beginning of your notebook file in order for your exam to be graded:

   "I understand that it is considered cheating to give or receive any unauthorized assistance on this test."

2. You need to answer to ONLY three problems. The choice is between question 2 and quetion 4. Questions 1 and 3 need to be answered.

3. You may ask Dr. Mohebbi any questions and post them on the discussion board, but you may not discuss the test or compare answers with anyone else.

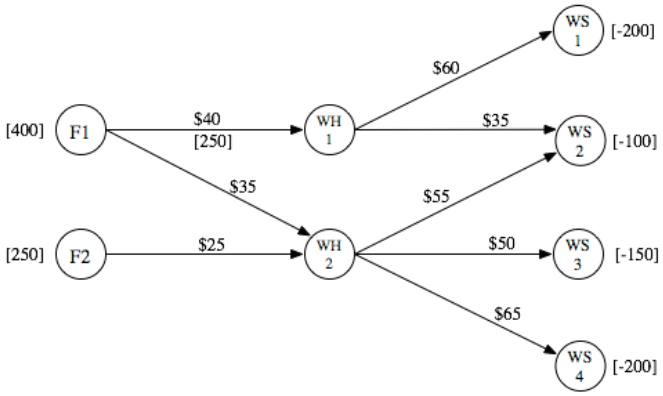
4. The test is untimed, but due before 8:00 p.m. on Friday, May 8, 2020.

5. Submit the codes for all problems in a notebook file on Canvas.

6. Good Luck!

1. **(30 points)** The list method **reverse** reverses the elements in the list. Define a function named **reverse** that reverses the element in its list argument (without using the method **reverse!**). Try to make this function as efficient as possible, and state its computational complexity using big-O notation.

2. **(40 points)** Heart Beats is a manufacturer of medical equipment. The company's primary product is a device used to monitor the heart during medical procedures. This device is produced in two factories and shipped to two warehouses. The product is then shipped on demand to four third-party wholesalers. All shipping is done by trucks. The product distribution network is shown below. The annual production capacity at Factories 1 and 2 is 400 and 250, respectively. The annual demand at Wholesalers 1, 2, 3, and 4 is 200, 100, 150, and 200, respectively. The cost of shipping one unit in each shipping lane is shown on the arcs. Due to limited truck capacity, at most 250 units can be shipped from Factory 1 to Warehouse 1 each year. Formulate and solve a network optimization model using Python and Gurobi to determine how to distribute the product at the lowest possible annual cost.



3. **Titanic Survivors (30 points)** A few minutes before midnight on April 14, 1912, the luxury ocean liner Titanic struck an iceberg in the North Atlantic Ocean. By 2:15 a.m., the Titanic, thought by many to be unsinkable, had sunk, leaving more than 2000 passengers and crew to meet their icy fate. Data on survival status of passengers and the type of passage they booked are given below.

Survival/Class	**First**	**Second**	**Third**	**Crew**	**Total**
Survived	203	118	178	212	711
Died	122	167	528	673	1490
Total	325	285	706	885	2201

(a) Show the percentage distribution of survival status by type of passage or the percentage distribution of type of passage by survival status, whichever you think is more appropriate.

(b) Construct a grouped bar graph that corresponds to the percentage distribution you chose.

(c) Interpret the results.

4. **(40 points)** The features of the 506 examples in the Housing dataset have been taken from the original source that was previously shared on `https://archive.ics.uci.edu/ml/datasets/Housing` and summarized here:

   - CRIM: Per capita crime rate by town
   - ZN: Proportion of residential land zoned for lots over 25,000 sq. ft.
   - INDUS: Proportion of non-retail business acres per town
   - CHAS: Charles River dummy variable (= 1 if tract bounds river and 0 otherwise)
   - NOX: Nitric oxide concentration (parts per 10 million)
   - RM: Average number of rooms per dwelling
   - AGE: Proportion of owner-occupied units built prior to 1940
   - DIS: Weighted distances to five Boston employment centers
   - RAD: Index of accessibility to radial highways
   - TAX: Full-value property tax rate per $10,000
   - PTRATIO: Pupil-teacher ratio by town
   - B: 1000(Bk – 0.63)2, where Bk is the proportion of [people of African American descent] by town
   - LSTAT: Percentage of lower status of the population
   - MEDV: Median value of owner-occupied homes in $1000s

   The house prices (MEDV) is our target variable - the variable that we want to predict using the 13 independent variables. You can load it into a DataFrame as below:

   $>>> import\ \ pandas\ \ as\ \ pd$
   $>>> df = pd.read\_csv('https\ :\ //raw.githubusercontent.com/rasbt/'python - machine-learning-book-3rd-edition/'master/ch10/housing.data.txt', header = None, sep ='\s+')$
   $>>> df.columns = ['CRIM','ZN','INDUS','CHAS','NOX','RM','AGE','DIS','RAD','TAX', 'PTRATIO','B','LSTAT','MEDV']$

   (a) Print descriptive statistics for all variables.

   (b) Explore the data structure using pairwise correlation and heatmap. Interpret the graphs and results.

   (c) Do you see any outliers, i.e. unusual observations, in the dataset (13 independent variables)? You may use z-scores to detect outliers.

   (d) Fit a linear regression model to predict MEDV using the 13 independent variables. Print the model summary and interpret the results.