Northeastern University
CS 5100 Foundations of Artificial Intelligence
Homework#6   [300 points]

PART A: Theory and Algorithms   *[100 points]*   * *See PART B Prog Assignment on Page 3*

Please - clearly write your **full name** on the first page. Submit a single PDF file.

Please provide brief but complete explanations, using diagrams where necessary, and suitably using your own words. While presenting calculations, explain the variables and.

Study Chapter 16 and 18 of Russel AI textbook – selected sections only, plus ML and DL notes provided. Answer the below:

1. **From textbook p.642 - 644**   [45 points]

   A. Problem 16.5 a only
   B. Problem 16.16 a and b only

2. **From textbook p.763**   **[40 points]**
      18.1
      18.3
      18.17
      18.19

   3. Study the Home Credit Default Risk Kaggle competition and data sets.   **[15 points]**
https://www.kaggle.com/c/home-credit-default-risk

Specify 4 machine learning questions relevant to this use case.

Against each ML question also list one or more algorithms which would help answer it.

PART B: Implementation – Programming Assignment (PA)   *(200 points)*

In this assignment, we'll focus on Ch. 16 Decision Networks and ML algorithms.

Section 1: **DECISION NETWORKS**   [100 points]

A decision network (Chapter 16, AIMA section 16.5) uses a Directed Acyclic Graph (DAG) to represent a set of random variables and their conditional dependencies within a probabilistic model, while a decision network extends the Bayesian network to include decision nodes and utility nodes. As in the example given in Figure 1, there are three types of nodes: Rectangles represent decision nodes, Ovals represent chance nodes, and Diamonds represent utility nodes. In this assignment, you will write code to perform inference in Decision Networks of discrete variables.
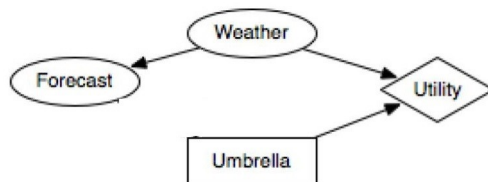
You will be given a decision network, which may have several decision nodes, several chance nodes, and at most one utility node. You will be asked to answer queries using the given network: Calculate the expected utility of a particular decision, or determine the decision with the maximum expected utility, given the joint probabilities.

Files provided:
- Assignment Notes for Decision Networks – which gives a deep insight of this particular problem
- Question.py
- TestCases folder which contains the input files.

Submission Checklist:
- Output files for all the test cases
- Decision_network_soln.py

Section 2:

Ridge and Lasso Regression for Regularization   [100 points]

In this assignment, you might need Jupyter notebook, as you will play around with the models. For this, you would need to install it. **Study and refer** to this implementation: http://www.science.smith.edu/~jcrouser/SDS293/labs/lab10-py.html

Install Jupyter notebook and ipynb

with conda:
    conda install ipython jupyter

with pip:
    # first, always upgrade pip!

    pip install --upgrade pip

    pip install --upgrade ipython jupyter

To open the ipynb files, you can open it using the command ipython notebook filename.ipynb from the directory it is downloaded on to. After running this command, you will be navigated to the Juypter notebook running on localhost:8888.

Zip file contents:
        Train.csv
        House-Price-Prediction-Ridge-Lasso.ipynb
The Jupyter notebook is trying to solve the below problem statement

Problem Statement for the example provided is below:

A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. The company wants to know
- Which variables are significant in predicting the price of a house, and
- How well those variables describe the price of a house.

Assignment Question:

Based on the above example, you pick a data set and a problem of your choice and come up with some prediction scenario and use ridge and lasso regression to come up with the factors that influence your dataset the most.

A good place for Data sets (not needed for this PA; just a reference)
>   https://www.kaggle.com/datasets