Homework 4

Stochastic Control and Optimization: MIS 381N

This assignment is graded on Credit/No-Credit.

That is, if you complete the homework and it is acceptable, you will get credit. If you don’ t submit or if the submitted work is not acceptable, you will not get credit. Getting a credit is required to obtain a grade for the group project that follows.

Please write a report that solves the following problems. Make sure that your report includes the mathematical formation of the optimization problem.

Problem 1:

It costs a company $12 to purchase an hour of labor and $15 to purchase an hour of capital. If L hours of labor and K units of capital are available, then $0.05L^{2/3}K^{1/3}$ machines can be produced. Suppose the company has $100,000 to purchase labor and capital.

What is the maximum number of machines it can produce?

Problem 2:

The file homework4stocks.csv contains historical monthly returns for 27 companies. The first row contains stock names and the first column contains the dates. For each company, calculate the estimated mean return and the estimated variance of return. Then calculate the estimated correlations between the companies' returns.

Find a portfolio that achieves an expected monthly return of at least 1% and minimizes portfolio variance. What are the fractions invested in each stock? What are the portfolio's estimated mean, variance, and standard deviation? Assume no short selling is allowed.

Problem 3:

The file 'variable_selection.csv' contains observations of variables y, x_1, x_2, and x_3. Here, y is the dependent variable. We want to choose a linear model that uses at most 2 independent variables such that the sum of squared residuals is minimized. This can be formulated as a constrained quadratic programming problem.

$$\min \sum_{i=1}^{n} (y_i - x_1\beta_1 - x_2\beta_2 - x_3\beta_3)^2$$

$$\textit{Subject to: at most two of } \beta_1, \beta_2, \textit{and } \beta_3 \textit{ are nonzero}$$

This is called best subset problem which is usually very hard to solve.

More concretely, you can solve it using the following binary quadratic program.

$$
\begin{aligned}
&\min_{\forall\beta_i,\forall b_i} \frac{1}{2}\beta^\top Q\beta \\
&st: \begin{aligned}[t] & |\beta_i| \le b_i M, \forall i \\ & \sum_{\forall i} b_i \le 2 \\ & b_i \in \{0,1\}, \beta_i \text{ are unconstrained} \end{aligned}
\end{aligned}
$$

However, binary quadratic programs are difficult to solve. One way to solve the problem is to use the branch and bound method that you learnt in the integer programming assignment. First fix a binary variable to 0/1 and solve the QP. Then build the binary tree in pre-order (i.e. run a calculation at a node, then recursively traverse each of its children) branching while fixing binary variables till you get an optimal solution. Here, M is some large constant. If you have n binary variables, then you will need to solve at most $2^{n+1}-1$ QPs. However, we will take a simpler approach to solving this problem.

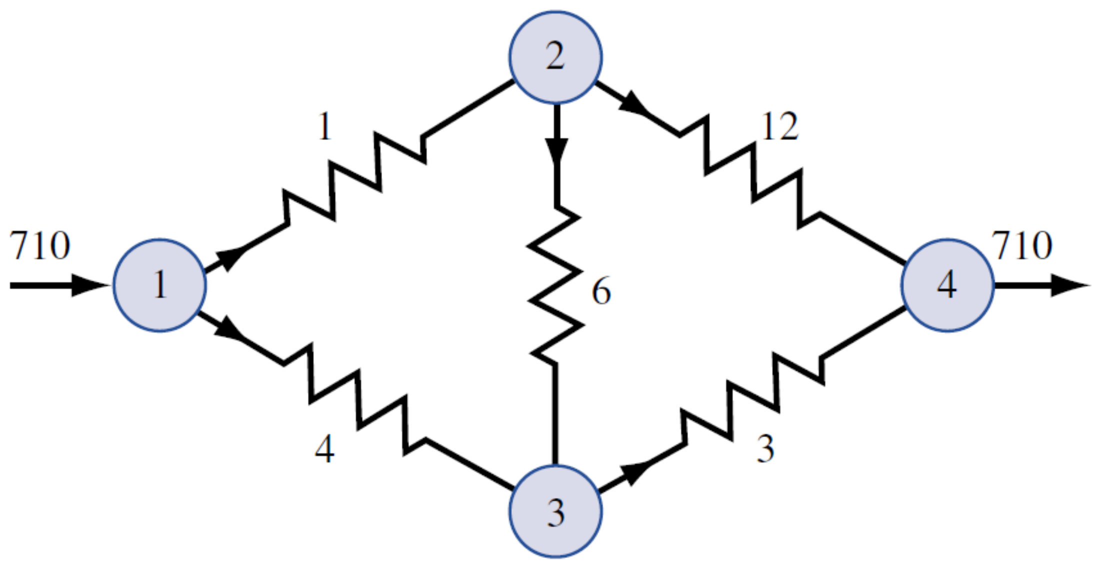
The number of variables in this problem is small and one can enumerate all possibilities. Run six OLS regressions (3 with one independent variable and 3 more with two variables each) and choose the regression that best fits the data.

It's just 6 linear regressions. This is the toy problem to let you warm up for your project. In your project, you will build on this problem.

Problem 4:

In an electrical network, the power loss incurred when a current of I amperes flows through a resistance of R ohms is I^2R watts. In the figure below, 710 amperes of current must be sent from node 1 to node 4. The current flowing through each node must satisfy conservation of flow. For example, for node 1, 710 = flow through 1-ohm resistor + flow through 4-ohm resistor. Remarkably, nature determines the current flow through each resistor by minimizing the total power loss in the network. Assume the current is non-negative.

1. Formulate a quadratic programming problem whose solution will yield the current flowing through each resistor.
2. Use R to determine the current flowing through each resistor.



Problem 5:

The file nflratings.csv contains the results of 256 regular-season NFL games from the 2009 season. The teams are indexed 1 to 32 as shown below:

Index	Team Name	Index	Team Name	Index	Team Name	Index	Team Name
1	Arizona Cardinals	9	Dallas Cowboys	17	Miami Dolphins	25	Pittsburgh Steelers
2	Atlanta Falcons	10	Denver Broncos	18	Minnesota Vikings	26	St. Louis Rams
3	Baltimore Ravens	11	Detroit Lions	19	New England Patriots	27	San Diego Chargers
4	Buffalo Bills	12	Green Bay Packers	20	New Orleans	28	San Francisco 49ers
5	Carolina Panthers	13	Houston Texans	21	New York Giants	29	Seattle Seahawks
6	Chicago Bears	14	Indianapolis Colts	22	New York Jets	30	Tampa Bay Buccaneers
7	Cincinnati	15	Jacksonville	23	Oakland Raiders	31	Tennessee Titans

	Bengals		Jaguars				
8	Cleveland Browns	16	Kansas City Chiefs	24	Philadelphia Eagles	32	Washington Redskins

The csv data file contains a matrix with the following columns:

- Week (1-17)
- Home Team Index (1-32 from the table above)
- Visiting Team Index (1-32 from the table above)
- Home Team Score
- Visiting Team Score

For example, the first game in the matrix is team 25 Pittsburgh versus team 31 Tennessee, played at Pittsburgh. Pittsburgh won the game by a score of 13 to 10, and the point spread (home team score minus visitor team score) is 3. A positive point spread means that the home team won; a negative point spread indicates that the visiting team won. The goal is to determine a set of ratings for the 32 NFL teams that most accurately predicts the actual outcomes of the games played. Use NLP to find the ratings that best predict the actual point spreads observed. The model will estimate the home team advantage and the ratings. The objective is to minimize the sum of squared prediction errors.

You will need to calculate the following:

- Actual Point Spread = Home Team Score – Visiting Team Score
- Predicted Spread = Home Team Rating – Visitor Team Rating + Home Team Advantage
- Prediction error = Actual Point Spread – Predicted Point Spread

You will also need to normalize the ratings. To do this, you set the actual average of the ratings to be 85 (this is somewhat arbitrary but based on the well-known Sagarin rating system).

What do these ratings mean? If two teams had ratings of 82 and 91, then the second team would be predicted to win by 9 points if the game was played on a neutral field.

This problem is similar to the problem you solved in the first homework. You have a quadratic objective and minimizing the prediction error is similar to solving the least squares problem, $y = X\beta + \epsilon$. If you solve the problem using the "lm" function in R then you will notice that your solution has an NA. This is because your matrix, $X^\top X$ is singular or does not have full rank. However, once you normalize your ratings or add a constraint to your data matrix, X , all variables are real.

Deliverables

You can either hand write or type your report, but make sure that you submit a PDF file. Please name your report as hw4_x.pdf(where x is your eid).

You can either integrate R code into your report or submit a separate R/Rmd file.