IQC delta scoring modeling

1. Defining our variables

Let $X_{n\times d}$ be a matrix that represents our set of alpha expression that passed all the IS conditions

The IS conditions are:

a) Fitness > 1
b) Sharpe > 1.25
c) 0.01 < Turnover < 0.7
d) Weight not too concentrated and adequate positions held

Let $Y_{n\times 1}$ represent the score delta for each of our alpha expression which is also our target expression

Let $Z_{n\times 1}$ represent the highest correlation value for each alpha expression

$n =$ number of alpha expression in our set

$d =$ number of features we have

In our case, $d = 42$

We have data on: *sharpe, turnover, fitness, returns, margin, long_count, short_count* for 2013 until 2018 which yields 42 different variables.

*Note: I intentionally excluded the highest correlation value since calculating it is infeasible because WebSim only allows us to run 50 correlation tests in a given time frame.

2. Goals

Objective function: $\max(\sum_{k\in\omega} y_k)$ s.t. $z_i < 0.7$

Find a subset $\omega \in X$

3. Metrics we would need to approximate

a. How does submitting one alpha affect the score delta of all the other alphas?

Let $M_{n\times n}^{(k)}$ represent the effect matrix where M_{ij} represents the effect of submitting alpha i on alpha j given that I submitted alpha k.

$$M_{n\times n}^{(k)} = \begin{bmatrix} 0 & & \\ & \ddots & \\ & & 0 \end{bmatrix}$$

4. Models to approximate this metric?

EM? Any way to decompose M into simpler parts?