Data Mining with Markowitz Portfolio Optimization in Higher Dimensions

Mark J. Bennett
Graduate Program in Analytics
University of Chicago

R in Finance, May, 2014
higher dimensional R-packages

Are they Data Mining cousins?
Historical Timeline

- Portfolio Theory (Markowitz)
- CAPM (Sharpe)
- Neural network (perceptron)
- Dual method (Goldfarb, Idnani)
- Decision Tree
- Support Vector Machine
- Neural Network
- Graphical Models
- *quadprog
- *huge Logistic Regression (large scale)
Why Are These Techniques Important?

- Portfolio Theory and Graphical Models are Natural Tools for Market Uncertainty, Portfolio Complexity
- Tie together Machine Learning and Financial Analytics: looking for Comovement in Multivariate Gaussian distributed returns
- Given dataset: \( N \) Training vectors \( \times p \) Dimensions
- Perform data mining where \( N \approx p \)
- Inspect and clean dataset: split-adjust
Historical $N + 1 = 1255$ prices for $p = 9$ securities
Historical $N = 1254$ log returns for $p = 9$ securities
Environment:

When we have $p$ time series of stock prices: $S_{ij}$ at time $i$ for security $j$, log returns are $R_{ij} = \ln(S_{ij}/S_{i-1j})$.

The full matrix of log returns is: $\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1p} \\ R_{21} & R_{22} & \cdots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{N1} & R_{N2} & \cdots & R_{Np} \end{bmatrix}$

We can find: $E\{\mathbf{R}\} = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$ and $\text{cov}(\mathbf{R}) = \Sigma_{p \times p}$
Portfolio Optimization:

Find \( w = \begin{bmatrix} w_1 \\ \vdots \\ w_p \end{bmatrix} \) the \textit{optimal weights} for the \( p \) securities given historical series of \( N \times p \) prices in order to forecast the best portfolio going forward:

\[
\arg\min_w w^T \Sigma w, \tag{1}
\]

for each target level of return \( \mu_P \) where \( w^T \Sigma w \) is the return variance of the portfolio, weighted by the vector \( w \) where below are the weight sum, portfolio mean, and no short sales constraints:

\[
\begin{pmatrix} 1^T w \\ \mu^T w \end{pmatrix} = \begin{pmatrix} 1 \\ \mu_P \end{pmatrix}; w \geq 0 \tag{2}
\]
Comovement Clusters for Weighted Stocks $p = 452$ securities

Huge $\lambda_{\text{min.ratio}}$: 0.4
NPN, Glasso

Cov Matrix Threshold: 1.7
Result Portfolio for $p = 452$ securities: Favoring CME, AAPL
Thank you!

References:
