Data Preprocessing for Homework 2

In this document we provide a brief description for the preprocessed data for Homework 2. The steps performed are:

1. Download the GSPC (S&P 500) and FIGRX (Fidelity International Discovery) historical prices from finance.yahoo.com.
2. Download the EURUSD data from the University of British Columbia web site.
3. Align these three data sets; i.e., remove all the dates for which there are missing entries. We remove a date from the GSPC and FIGRX data if there is no entry for that date in the EURUSD data and vice versa.
4. Construct the daily percentage changes of the GSPC data by

$$\mathrm{GSPC}_{\mathrm{change}}(t) = \frac{\mathrm{GSPC}(t) - \mathrm{GSPC}(t-1)}{\mathrm{GSPC}(t-1)} \cdot 100\%$$

where GSPC(t) is today's market closing price, GSPC(t − 1) is yesterday's (previous business day) closing price, and $\mathrm{GSPC}_{\mathrm{change}}(t)$ is today's daily change (in percentage). Repeat this step for the FIGRX and EURUSD data and obtain $\mathrm{FIGRX}_{\mathrm{change}}(t)$ and $\mathrm{EURUSD}_{\mathrm{change}}(t)$, respectively.

The preprocessed data are provided in the structure data2004 (or data2005), which has the following fields,

• data2004.X is a 243×2 matrix with $\mathrm{GSPC}_{\mathrm{change}}(t)$ and $\mathrm{EURUSD}_{\mathrm{change}}(t)$ in the first and second column, respectively.

• data2004.Y is a 243 × 1 vector containing $\mathrm{FIGRX}_{\mathrm{change}}(t+1)$.

• data2004.y is a 243 × 1 vector containing the encoded class labels (1 or 0) for the FIGRX data. This can be obtained by thresholding the values of data2004.