**数据描述和说明**

**A\_share\_name.xlsx**

A股的股票代码和中文名称对应表，第一列是股票代码（整数型），例如，000001记为1；第二列是中文名。

**SZ399300.TXT**

沪深300指数数据，按列包含日期、开盘价、最高价、最低价、收盘价成交量、成交额数据，数据期间从2010年6月2日到2015年12月9日，共1340天。该数据的交易日期是完整的。

**hs300 文件夹**

沪深300成份股个股的原始数据。数据是从通达信软件下载的个股日行情，每个股票单独存在一个txt文件，按列包含日期、开盘价、最高价、最低价、收盘价成交量、成交额数据，数据期间从2010年6月2日到2015年12月9日，共1340天。300个股股票的数据由于可能停牌，退市或尚未上市，存在缺失数据需要补全。原始数据中仅包含有成交记录的日期，因此每个txt文件的长度可能是不同的。

**zip.train**

zip.train数据集在《ESL》书中使用。数据读取后得到一个7291×257 的数据矩阵，包含7291条数据，每条数据共257列。其中，第一列表示数据属于哪个数字。第i行数据的第一列取值为k，表示该行随后256个数据是对数字k的一个手写记录。数据记录是大小为 256× 1 的向量，由 16× 16 的像素图像所对应的 16 ×16 数据矩阵按行拼接得到。

**macro\_econ\_data.xls**

数据取自美国银行（Bank of America）内部整理的美国1990-2008年月度宏观经济数据共228个观测值，包括个人消费支出（Personal Consumption Expenditure， 记为Nominal PCE或 pce），未偿还个人消费贷款（Revolving Credit，记为debt），和违约率（Charge-off Rate）。

**t-bond rates.xlsx**

利率期限结构数据集记录了美国国债（T-Bond）自1990年1月到2009年8月每月初的利率期限结构。利率期限结构（又称为利率曲线）是指在一个时点不同期限的即期利率与到期期限的关系。例如，数据集的第一条数据记录了在1990年1月1日期限为3个月，6个月，1年，2年，3年，5年，7年和10年对应的国债利率。我们一共有236条记录，对应236个月份，每条记录是一个8维向量。