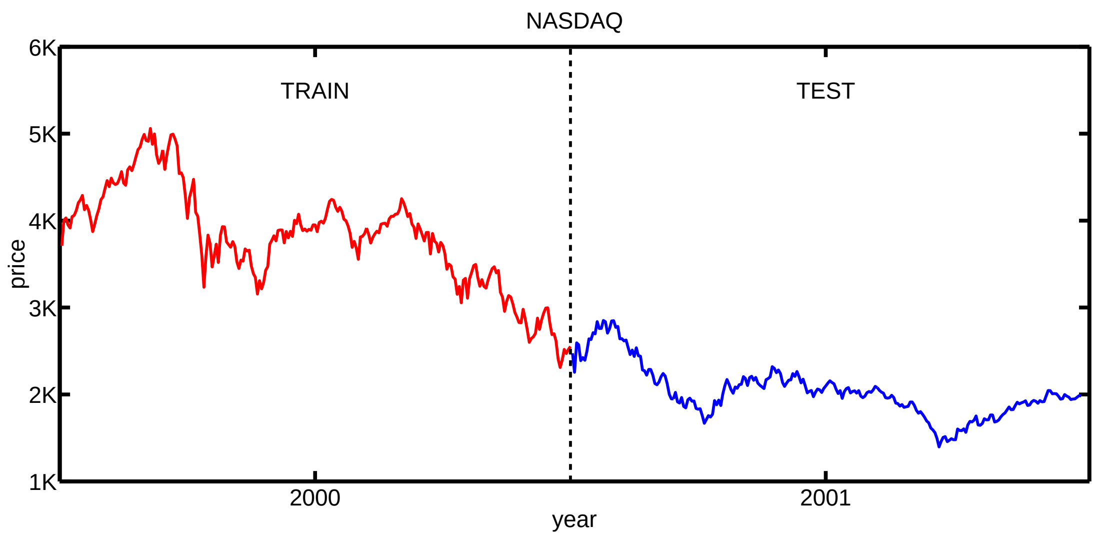
**CSE 291d. Assignment 4**

**Out:** *Thu Feb 01*
**Due:** *Fri Feb 09*

## 4.1 Stock market prediction

In this problem, you will apply a simple linear model to predicting the stock market. From the course web site, download the files `nasdaq00.txt` and `nasdaq01.txt`, which contain the NASDAQ indices at the close of business days in 2000 and 2001.




(a) How accurately can the index on one day be predicted by a linear combination of the three preceding indices? Using only data from the year 2000, compute the linear coefficients $(w_1, w_2, w_3)$ that minimize the mean squared prediction error:

$$\varepsilon(\vec{w}) = \frac{1}{T}\sum_t \left(x_t - w_1 x_{t-1} - w_2 x_{t-2} - w_3 x_{t-3}\right)^2,$$

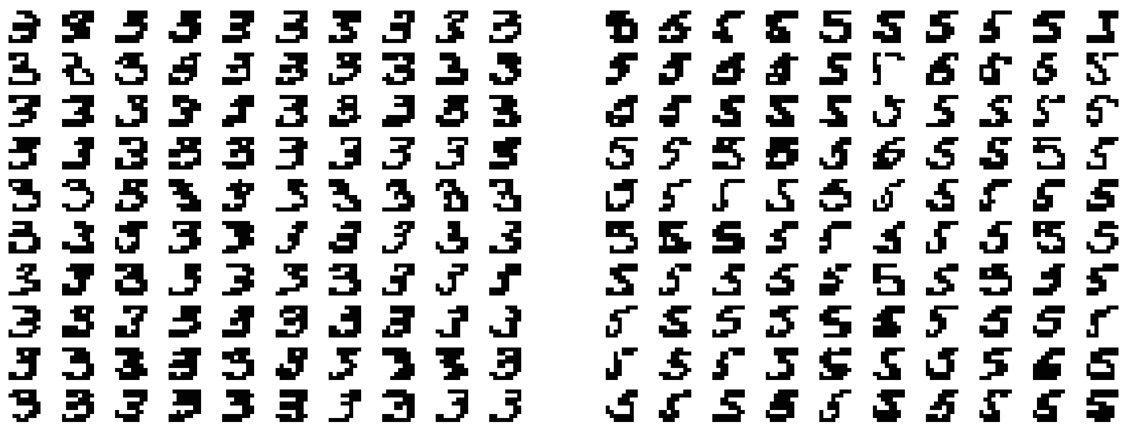
where the sum is over business days in the year 2000 (starting from the fourth day).

(b) For the coefficients estimated in part (a), compare the mean squared prediction errors on the data from the years 2000 and 2001. Would you recommend this linear model for stock market prediction?

Turn in your source code, your solution for the linear coefficients, and your results for the mean squared prediction errors. You may program in the language of your choice, and you may solve the required system of linear equations either by hand or by using built-in routines (e.g., in Matlab, Maple, Mathematica, etc.).

## 4.2 Handwritten digit classification

In this problem, you will use logistic regression to classify images of handwritten digits. From the course web site, download the files `digits3a.txt`, `digits3b.txt`, `digits5a.txt`, and `digits5b.txt`. These files contain data for binary images of handwritten digits. Each image is an 8x8 bitmap represented in the files by one line of text. Some of the examples are shown in the following figure.



(a) Perform a logistic regression (using gradient ascent or Newton's method) on the images in files `digits3a.txt` and `digits5a.txt`. Indicate clearly the algorithm used, and provide evidence that it has converged (or nearly converged) by printing out the log-likelihood on several iterations of the algorithm, as well as the percent error rate on the images in these files. Also, print out the 64 elements of your solution for the weight vector as an 8x8 matrix.

(b) Use the model in part (a) to label the images in the files `digits3b.txt` and `digits5b.txt`. Report your percent error rate on these images.

Again, turn in your source code. You may program in the language of your choice.

## 4.3 Multinomial logistic regression

A simple generalization of logistic regression is to predict a discrete (but non-binary) label $y \in \{1, 2, \ldots, m\}$ from a real-valued vector $\vec{x} \in \mathcal{R}^d$. For the graphical model shown below, consider the following parameterized conditional probability table (CPT):

$$P(y|\vec{x}) \;=\; \frac{e^{\vec{w}_y \cdot \vec{x}}}{\sum_{i=1}^{m} e^{\vec{w}_i \cdot \vec{x}}}.$$

The parameters of this CPT are the weight vectors $\vec{w}_y$. Note that in this model, a weight vector must be learned for each possible label $\{1, 2, \ldots, m\}$. The sum in the denominator normalizes this distribution so that $\sum_{y=1}^{m} P(y|\vec{x}) = 1$.

Consider a training set of $T$ labeled examples $\{(\vec{x}_t, y_t)\}_{t=1}^{T}$. As shorthand, let $\hat{y}_{it} \in \{0, 1\}$ denote the *target* assignment matrix defined by:

$$\hat{y}_{it} \;=\; \begin{cases} 1 & \text{if} \quad y_t = i, \\ 0 & \text{otherwise.} \end{cases}$$

Also, let $p_{it} = P(Y = i|\vec{x}_t)$ denote the probability that the model maps the $t$th example to the $i$th label. The weight vectors can be obtained by maximum likelihood estimation. Show that a necessary condition for the log-likelihood $\mathcal{L} = \sum_t \log P(y_t|\vec{x}_t)$ to be maximized as a function of the weight vector $\vec{w}_i$ is that:

$$\sum_t (\hat{y}_{it} - p_{it})\, \vec{x}_t \;=\; 0.$$