Assignment 8, due May 2

Corrections: (April 29, posted a corrected version of `pca.R`.
May 1, posted another correction to `pca.R`. This one makes the `U` and `v` objects matrices instead of named lists. Now, the expression `U[3,2]` will gives the third component of the second left principal component vector.)

1. **True/False**.

   (a) Suppose Z_k is a family of independent Gaussian random variables with mean zero and variance 1. Suppose that random variables are "explained" by the independent factors Z_k as

   $$X_j = \sum_{k=1}^{m} w_{jk} Z_k \ .$$

   Then the weight factors w_{jk} are uniquely determined by the covariance matrix of the random variables X_j.

   (b) Computer calculations are done exactly. If `y = sqrt(x)` and then `z = y*y`, then `z - x` will be exactly zero.

   (c) Let S_{tj} be the closing price of asset j for day t. Let $X_{t,j}$ be the corresponding daily returns. If we multiply each number S_{tj} by two then the singular values of X do not change. But if we record one of the series $S_{t,j}$ twice (add column $n+1$ equal to column n, say), then the singular values will change.

   (d) An $n \times n$ matrix Q is *orthogonal* if $QQ^t = I$. The singular values of X and XQ are the same, if X is a $T \times n$ matrix.

2. (*Warning: this involves quite a lot of arithmetic and calculation with fractions and square roots. Please make your calculations readable.)* Let X be the matrix

   $$X = \begin{pmatrix} 1 & 2 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}$$

   (a) Find the right principal component vectors v_1 and v_2 and the corresponding singular values σ_1 and σ_2 and the corresponding left principal component vectors U_1 and U_2.

   (b) Verify that $X = \sigma_1 U_1 v_1^t + \sigma_2 U_2 v_2^t$.

(c) (*You will need to convert some fraction or square root expressions to decimal values to do this.*) Sketch in the v plane the contour curve $f(v) = \|Xv\|^2 = 1$. For this, it might be helpful to use (x, y) notation

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

Write $f(v) = ax^2 + 2bxy + cy^2$ and find the major and minor axes of the ellipse $f(v) = f(x, y) = 1$. (The *major axis* is the vector that points to the point on the ellipse farthest from the origin. The *minor axis* is the vector that points to the point closest to the origin.) Which right principal component (v_1 or v_2) corresponds to the major axis? What is the relation between σ_1 and σ_2 to the lengths of the major and minor axes? Please choose the scale on the axes so that the ellipse fills much of the plot.

(d) Verify your algebra using the `svd()` function in `R`. The script `pca.R` does the computation. You can run it (using `source("pca.R")` in the `R` app) and then just print the results. You can add to the script the formulas from part (a) to verify that the SVD calculation agrees with your paper and pencil/pen calculations from part (a).

3. Suppose there is a list of energy sector assets and an investor wants to put fraction p of her wealth in these energy sector stocks, with the remaining $1-p$ allocated to stocks from other sectors. Show that there is a three fund theorem. That is, there are three collections of weights w_1, w_2, and w_3 so that every efficient allocation subject to these constraints is a combination of these funds.

4. Suppose U is a $T \times n$ matrix whose columns are ortho-normal vectors. Show that $U^t U = I$. This is the $n \times n$ identity matrix.

5. Let $X = U\Sigma V^t$ be the SVD. Show that $C = X^t X$ has n eigenvalues $\lambda_j = \sigma_j^2$ and corresponding eigenvectors v_j being the columns of V.

6. Let A be a square $n \times n$ matrix with entries a_{ij}. The *trace* of A is the sum of the diagonal entries:

$$\mathrm{Tr}(A) = \sum_{j=1}^{n} a_{jj} \ .$$

(a) If X is a $T \times n$ matrix, verify this formula for the total sum of squares:

$$SS_{\mathrm{tot}} = \mathrm{Tr}(X^t X) \ .$$

(b) Show that if A and B are $n \times n$ matrices, then, even if $AB \neq BA$,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) \ .$$

(c) Show that if $X = U\Sigma V^t$ is the SVD of X, then

$$SS_{\text{tot}} = \sum_{j=1}^{n} \sigma_j^2 \; .$$

Hint: use the trace formula for sum of squares.

(d) Let R be the residual from the approximation with m principal components. Show that the residual sum of squares is

$$SS_{\text{res}} = \sum_{j=m+1}^{n} \sigma_j^2 \; .$$

Computing exercise.

Use principal component analysis to look for factors in your daily return data from Assignment 7. Experiment with different numbers of assets and different T (from one to ten years) to see what fraction of the total sum of squares is explained by 1, 2, 3, factors, etc. Describe the relative sizes of the first few singular values. Use the `R` function `svd` to find the principal components. Do not include the index SPY in the collection of assets. Comment on how well the first factor (given by the first principal component) tracks SPY. Asked another way, how close is SPY to being the "market factor"?