Homework 4: Solutions

Problems to turn in individually

1. Let M, N, and O be $n \times n$ matrices and v be an n-vector. Rewrite

$$\sum_{i,j,k} M_{ik} O_{kj} N_{jl} v_i$$

in matrix-vector notation ending with the vector v as in class. Now write the expression in matrix-vector notation but starting, as opposed to ending, with the vector v. Notice that when you start with a vector, it must be the transpose of the vector for the multiplication to make sense. Compare your two expressions, which are the transpose of each other. This comparison should make clear how to take the transpose of a product of matices and vectors. Use this to rewrite $(w^T B C^T A^T u)^T$ in terms of the product of the matrices B, C, and A and their transposes and the vectors w and u and their transposes. This last question should take you under 30 seconds to do if you understood how to generalize the comparison correctly.

ANS:

$$\begin{aligned}
\sum M_{ik} O_{kj} N_{jl} v_i &= \sum N^T_{lj} O^T_{jk} M^T_{ki} v_i \Rightarrow N^T O^T M^T v \\
&= \sum v_i M_{ik} O_{kj} N_{jl} \Rightarrow \left(v^T M O N\right) = \left(N^T O^T M^T v\right)^T \\
\text{so } \left(w^T B C^T A^T u\right)^T &= \boxed{u^T A C B^T w}
\end{aligned}$$

2. Let A be a matrix of constants and x be a vector. Determine the gradient of $x^T A x$ with respect to x. (That is determine the derivative of $x^T A x$ with respect to each component in the x vector.) If A is symmetric, how can you simplify your results? (You should be able to combine the two terms in your answer when A is not symmetric into a single term.)

ANS:

$$
\begin{aligned}
\nabla_x \left(x^T A x\right) &\Rightarrow \frac{\partial}{\partial x_i}\left(\sum_{j,k} x_j A_{jk} x_k\right) = \sum_{j,k}\left(\frac{\partial x_j}{\partial x_i} A_{jk} x_k + x_j A_{jk} \frac{\partial x_k}{\partial x_i}\right) \\
&\qquad \text{(Product rule)} \\
&= \sum_{j,k} \left(\delta_{ij} A_{jk} x_k + x_j A_{jk} \delta_{ik}\right) \\
&= \sum_k A_{ik} x_k + \sum_j x_j A_{ji} = \sum_j A_{ij} x_j + A_{ij}^T x_j \\
&\Rightarrow \boxed{Ax + A^T x} \quad \boxed{(= 2Ax \text{ if } A \text{ is symmetric })}
\end{aligned}
$$

3. In this question, you will continue to use R with the data you extracted from HW2. Recall that you had saved a time series of stock returns for your favorite stock. Now do the following:

   (a) Sort the returns and find the 99th percentile value.

   (b) Find the 1st percentile value.

   (c) What is the mean of all returns greater than 5%?

   (d) Assuming that the returns are normally distributed, find the mean and variance of the return distribution using maximum-likelihood. Compare your results to the actual mean and variance.

   (e) How many sigma away from the mean is the lowest return?

   (f) Fit different GARCH models to your data. Which one seems to fit the best?

   (g) Plot the volatility time series.

   (h) What is the mean volatility you extracted from the data?

   ANS: Here is the R program (answers vary depending on data)

```
stkret = read.table("Data.txt",header=TRUE)
n = length(stkret)
sortret = sort(stkret)

pct99 = sortret[floor(0.99*n)]
pct01 = sortret[floor(0.01*n)]
```

```
For MLE, see the code in the notes.

sigma_away = (sortret[1]-mean(stkret))/sd(stkret)

For GARCH, use the code in the notes.
```

Problems to turn in as a group

1. Let X be a random variable. Let's say that we take n samples x_i, where $i = 1, 2, ..., n$, from the distribution of X. Similarly, take n samples y_i, where $i = 1, 2, ..., n$, from the distribution of a random variable Y. These sample data values can be used to create *point estimators* (i.e., best approximations) to various parameters for the random variables. For example, the point estimator (i.e., the best approximation) of the mean of X (i.e., $E[X]$) from the sampled data is called the *sample mean* $\bar{x}$:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

The point estimator of $V[X]$, the variance of X, is the *sample variance:*

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n-1} = \frac{\sum_{i=1}^{n}(x_i^2) - n\bar{x}^2}{n-1}.$$

The point estimator of $Cov[X, Y]$, the covariance of X and Y, is the *sample covariance:*

$$\frac{\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{n-1} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{n-1}.$$

If we now consider the n data points (x_i, y_i) and try to best fit these data points with the line

$$y = a + bx,$$

show that the value of b that comes from the regression calculation happens to equal the sample covariance of the x_i and y_i divided by the sample variance of the x_i.

ANS:

The regression coefficients are $\beta = \left(x^T x\right)^{-1} x^T y$, where, for this case, $\beta = \begin{bmatrix} a \\ b \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$, and $x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}$. Therefore, $x^T y = \begin{bmatrix} \sum y_i \\ \sum(x_i y_i) \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \sum(x_i y_i) \end{bmatrix}$ (where all sums are from $i = 1$ to n) and $x^T x = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{bmatrix}$, so $(x^T x)^{-1} = \frac{1}{n\sum x_i^2 - n^2\bar{x}^2} \begin{bmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$. Therefore, multiplying $\left(x^T x\right)^{-1}$ and $x^T y$, gives us that b is equal to $\frac{-n^2\bar{x}\bar{y} + n\sum(x_i y_i)}{n\sum x_i^2 - n^2\bar{x}^2} = \frac{\text{sample covariance}}{\text{sample variance of } x}$.

2. Let's say the monthly returns over the past seven months for stock in Quail Egg, Inc. are the following: .00852, .01045, .00573, .01578, -.01223, .00682, .00452. In general, we do not expect that these returns are correlated with time, but given statistical variation and the small number of data points, we expect there to be some slope to the line that best fits these points. (Although the likelihood of the slope being positive or negative is the same.)

   (a) Determine the line $a + bt$ (where t is in months) that minimizes the sum of the square differences between the returns and the line.

   (b) Determine the constant a that minimizes the sum of the square differences between the returns and the constant. Compare this to the average value of the returns. Do you expect the minimizing constant and the average value to be close? the same? Explain.

   (c) Determine the parabola $a + bt + ct^2$ that minimizes the sum of the square differences between the returns and the parabola. (Hint: think of t^2 as x_2 from class.)

ANS:

$$y = \begin{bmatrix} .00852 \\ . \\ . \\ . \\ .00452 \end{bmatrix}$$

Use $\beta = \left(x^T x\right)^{-1} x^T y$

(a) $$x = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ . & . \\ . & . \\ . & . \\ 1 & 7 \end{bmatrix} \Rightarrow \boxed{.01097 - .00133t}$$

(b) $$x = \begin{bmatrix} 1 \\ . \\ . \\ . \\ 1 \end{bmatrix} \Rightarrow \boxed{a = .00566}$$

The average value is the same. Why?

$$\begin{aligned} a \text{ minimizes } \sum_i (y_i - a)^2 \quad &\Rightarrow \quad 0 = \frac{d}{da} \sum_i (y_i - 1)^2 \\ &\Rightarrow \quad 0 = 2 \sum_i (y_i - a) \\ & \quad 0 = \sum_i y_i - a \sum_i 1 \\ &\Rightarrow \quad a = \frac{\sum_i y_i}{7} = \text{avg.value!} \end{aligned}$$

(c) $$x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ . & . & . \\ . & . & . \\ 1 & 7 & 49 \end{bmatrix} \Rightarrow \boxed{.01405 - .00338t + .000257t^2}$$