# Stat243: Problem Set 7, Due Fri. November 17

November 2, 2017

This covers Units 9 and 10.
It's due **on paper** and submitted via Github at the start of class
Some general guidelines on how to present your problem set solutions:

1. Please use your Rtex/Rnw/Rmd solution from PS1, problem 4 as your template for how to format your solutions (only non-Statistics students are allowed to use R Markdown).

2. As usual, your solution should mix textual description of your solution, code, and example output. And your code should be commented.

3. Your paper submission should be the printout of the PDF produced from your Rtex/Rnw/Rmd file. Your Github submission should include the Rtex/Rnw/Rmd file, any R or bash code files containing chunks that you read into your Rtex/Rnw/Rmd file, and the final PDF.

4. Use functions as much as possible, in particular for any repeated tasks. We will grade in part based on the modularity of your code and your use of functions.

5. Please note my comments in the syllabus about when to ask for help and about working together.

6. Please give the names of any other students that you worked with on the problem set.

## Problems

1. Suppose I have a statistical method that estimates a regression coefficient and its standard error. I develop a simulation study and have $m = 1000$ simulated datasets that each give me an estimate of the coefficent and its standard error. How would I determine if the standard error properly characterizes the uncertainty of the estimated regression coefficient? Note your answer could be as simple as a sentence or two describing what quantities to consider.

2. Show that $\|A\|_2$ is the largest of the absolute values of the eigenvalues of $A$ for symmetric $A$. To do so, find the following quantity,

$$\|A\|_2 = \sup_{z:\|z\|_2=1} \sqrt{(Az)^\top Az}.$$

If you're not familiar with the notion of the supremum (the 'sup' here), just think of it as the maximum.

Hints: when you get to having the quantity $\Gamma^\top z$ for orthogonal $\Gamma$, set $y = \Gamma^\top z$ and show that if $\|z\|_2 = 1$ then $\|y\|_2 = 1$. Finally, if you have the quantity $y^\top Dy$ for diagonal matrix $D$, express this as a sum and think intuitively about how to maximize it if $\|y\|_2 = 1$.

3. Some practice with matrix manipulations.

   (a) Consider a rectangular matrix, $X$, with dimensions $n \times p$ and $n > p$. Show that the right singular vectors of $X$ are the eigenvectors of the matrix $X^\top X$ and that the eigenvalues of $X^\top X$ are the squares of the singular values of $X$. Also show that $X^\top X$ is positive semi-definite (which is good because $X^\top X$ is essentially an empirical covariance matrix, up to scaling and shifting). (Sidenote: as I mentioned in class, since the condition number of X is the ratio of the largest and smallest magnitude singular values, this shows why the condition number for using the Cholesky for regression is the square of the condition number for using the QR, and hence why the QR might be preferred to the Cholesky for doing OLS.)

   (b) Consider an $n \times n$ positive semi-definite matrix $\Sigma$ and assume you have already computed the eigendecomposition of $\Sigma$. How can you compute the eigenvalues of $Z = \Sigma + cI$ in $O(n)$ arithmetic calculations (including any additions or multiplications), where $c$ is a scalar and $I$ is the identity matrix?

4. The following calculation arises in solving a least squares regression problem where the coefficients are subject to an equality constraint, in particular, we want to minimize $(Y - X\beta)^\top (Y - X\beta)$ subject to the $m$ constraints $A\beta = b$ for an $m$ by $p$ matrix $A$. (Each row of $A$ represents a constraint that that linear combination of $\beta$ equals the corresponding element of $b$.
   Solving this problem is a form of optimization called quadratic programming. Some derivation using the Lagrange multiplier approach gives the following solution:

   $$\hat{\beta} = C^{-1}d + C^{-1}A^\top (AC^{-1}A^\top)^{-1}(-AC^{-1}d + b)$$

   where $C = X^\top X$ and $d = X^\top Y$. $X$ is $n$ by $p$.

   (a) Describe how you would implement this in pseudo-code.

   (b) Write an R function to efficiently compute $\hat{\beta}$, taking account of the principles discussed in class in terms of matrix inverses and factorizations. Note: you can use any of R's matrix manipulation functions that you want - I'm not expecting you to code up any algorithms from scratch.
   Note: in reality an efficient solution is only important when the number of regression coefficients, $p$, is large.

5. Two-stage least squares (2SLS) is a way of implementing an causal estimation method called instrumental variables that is commonly used in economics. Consider the following set of regression equations:

   $$\begin{aligned} \hat{X} &= Z(Z^\top Z)^{-1}Z^\top X \\ \hat{\beta} &= (\hat{X}^\top \hat{X})^{-1}\hat{X}^\top y \end{aligned}$$

   which can be intepreted as regressing $y$ on $X$ after filtering so that we only retain variation in $X$ that is correlated with the instrumental variable $Z$. An economics graduate student asked how he could compute $\hat{\beta}$ if $Z$ is 60 million by 630, $X$ is 60 million by 600, and $y$ is 60 million by 1, but both $Z$ and $X$ are sparse.

   (a) Describe briefly why I can't do this calculation in two stages as given in the equation, even if I use the techniques for OLS discussed in class for each stage.

(b) Figure out how to rewrite the equations such that you can actually calculate $\hat{\beta}$ on a computer without a huge amount of memory. You can assume that any matrix multiplications involving sparse matrices can be done on the computer (e.g., using the spam package in R). Describe the specific steps of how you would do this and/or write out in pseudo-code.

6. (Extra credit) For this problem, your task is to empirically explore the condition number of the eigen-decomposition. First create a set of eigenvectors. An easy way to do this is to find the eigenvectors of $A = Z^\top Z$ for arbitrary $Z$, throwing away the eigenvalues. Now explore creating positive eigenvalues of different magnitudes and creating $\Gamma \Lambda \Gamma^\top$ from the $\Gamma$ you got for $A$ and your chosen eigenvalues. Use *eigen()* and see how close the computed eigenvalues are to the actual eigenvalues. I would use $n = 100$ and let the eigenvalues vary between all being equal and having a range of values from very large to very small. At what condition number do you empirically see that your matrix is not numerically positive definite? How does the error in the estimated eigenvalues relative to the known true values vary with the condition number and the magnitude of the eigenvalues?