

# Lab2: Introduction to GitHub

Your Name

September 4, 2019

## Introduction

In this lab we will work with github on a local and remote repository for version control of your work. We will use the problem set from HW1 as a starting point.

1. Exercise 1.11 Christensen (review notes on expectations and covariance; recall that  $E[\mathbf{AY}] = \mathbf{AY}$  and  $\text{Cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{ACov}(\mathbf{Y})\mathbf{B}^T$ )
2. Exercise 1.5.2 Christensen
3. We showed that  $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$  was an orthogonal projection on the column space of  $\mathbf{X}$  and that  $\hat{\mathbf{Y}} = \mathbf{P}_\mathbf{X}\mathbf{Y}$ . While useful for theory, the projection matrix should never be used in practice to find the MLE of  $\boldsymbol{\mu}$  due to computational complexity (inverses and matrix multiplication) and instability. To find  $\hat{\boldsymbol{\beta}}$  for  $\mathbf{X}$  of full column rank we solve  $\mathbf{X}\boldsymbol{\beta} = \mathbf{P}_\mathbf{X}\mathbf{Y}$  which leads to the *normal equations*

$$(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Y}$$

and solving the system of equations for  $\boldsymbol{\beta}$ . Instead consider the following for  $\mathbf{X}$  ( $n \times p, p < n$ ) of rank  $p$

- (a) Any  $\mathbf{X}$  may be written via a singular value decomposition as  $\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$  where  $\mathbf{U}$  is a  $n \times p$  orthonormal matrix, ( $\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$  and columns of  $\mathbf{U}$  form an orthonormal basis (ONB) for  $C(\mathbf{X})$ ),  $\boldsymbol{\Lambda}$  is a  $p \times p$  diagonal matrix and  $\mathbf{V}$  is a  $p \times p$  orthogonal matrix ( $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_p$ ). Note the difference between *orthonormal* and *orthogonal*. Show that  $\mathbf{P}_\mathbf{X}$  may be expressed as a function of  $\mathbf{U}$  only and provide an expression for  $\hat{\mathbf{Y}}$ . Similarly, find an expression for  $\hat{\boldsymbol{\beta}}$  in terms of  $\mathbf{U}$ ,  $\boldsymbol{\Lambda}$  and  $\mathbf{V}$ . Your result should only require the inverse of a diagonal matrix!
- (b)  $\mathbf{X}$  may be written in a (reduced or thinned) QR decomposition as a matrix  $\mathbf{Q}$  that is a  $n \times p$  orthonormal matrix (which forms an ONB for  $C(\mathbf{X})$ ) and  $\mathbf{R}$  which is a  $p \times p$  upper triangular matrix (i.e all elements below the diagonal are 0) where  $\mathbf{X} = \mathbf{QR}$ . The columns of  $\mathbf{Q}$  are an ONB for the  $C(\mathbf{X})$ . Show that  $\mathbf{P}_\mathbf{X}$  may be expressed as a function of  $\mathbf{Q}$  alone. Show that the normal equations reduce to solving the triangular system  $\mathbf{R}\boldsymbol{\beta} = \mathbf{Z}$  where  $\mathbf{Z} = \mathbf{Q}^T\mathbf{Y}$ . Because  $\mathbf{R}$  is upper triangular, show that  $\hat{\boldsymbol{\beta}}$  may be obtained by back-solving thus avoiding the matrix inverse of  $\mathbf{X}^T\mathbf{X}$ .

- (c) Any symmetric matrix  $\mathbf{A}$  may be written via a Cholesky decomposition as  $\mathbf{A} = \mathbf{L}\mathbf{L}^T$  where  $\mathbf{L}$  is lower triangular. If  $\mathbf{Z} = \mathbf{X}^T\mathbf{Y}$  show that we can solve two triangular systems  $\mathbf{L}\mathbf{L}^T\boldsymbol{\beta} = \mathbf{Z}$  by solving for  $\mathbf{w}$  using  $\mathbf{L}\mathbf{w} = \mathbf{Z}$  using a forward substitution and then for  $\hat{\boldsymbol{\beta}}$  using  $\mathbf{L}^T\boldsymbol{\beta} = \mathbf{w}$  avoiding any matrix inversion. Write this out in psuedo-code.
- (d) Prove that the two projection matrices obtained by the SVD and the QR method are the same. (review Theorems in Christensen Appendices about uniqueness of projections and cite appropriate results to show that they are the same.)
- (e) Use R to find  $\mathbf{Q}$  and  $\mathbf{U}$  for the matrix in Example 1.0.2 in Christensen using the QR and SVD methods respectively. Does  $\mathbf{Q}$  equal  $\mathbf{U}$ ? See help pages via `help(qr)` and `help(svd)` for function documentation. (Use Sweave/knitr to write up the solution!)
- (f) Verify that the numerical solutions for the projections using  $\mathbf{Q}$  and  $\mathbf{U}$  are equal to the expression in problem 1.5.8 (b).  
 Note: The Cholesky method is the fastest in terms of  $O(np^2 + p^3/3)$  floating point operations (flops), but is numerically unstable if the matrix is poorly conditioned. R uses the QR method ( $O(2np^2 - 2p^3/3)$  flops) in the function `lm.fit()` (which is the workhorse underneath the `lm()` function. The SVD method is the most expensive  $O(2np^2 + 11p^3)$  but can handle the rank deficient case. There are generalized Cholesky and QR methods for the rank deficient cases that involve pivoting the columns of  $\mathbf{X}$  which avoids some of the instability used in the BAS package `bas.lm` for Bayesian regression and `lm`
4. Suppose  $\boldsymbol{\Sigma}$  is a real  $p \times p$  positive semi-definite matrix. Then the Cholesky decomposition of  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$  where  $\mathbf{L}$  is a lower triangular matrix with real, non-negative elements on the diagonal. Suppose you can generate standard normal random variates,  $\mathbf{Z} = (z_1, \dots, z_p)^T$  with  $z_i \text{ i.i.d. } N(0, 1)$ .
- (a) What is the distribution of  $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{L}\mathbf{Z}$  for  $\boldsymbol{\mu} \in \mathbb{R}^p$ ?
- (b) Explain why it does not matter for generating  $\mathbf{Y}$  that the Cholesky decomposition is not unique when  $\boldsymbol{\Sigma}$  is not positive definite. Does it matter how we choose a *matrix square root*  $\mathbf{A}$  of  $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$ ? Explain.
- (c) In R use the `chol` function to find the Cholesky factorization of  $\mathbf{V}$  in exercise 1.5.2.
- (d) Using the Cholesky factorization, generate 5,000 samples for the distribution of  $\mathbf{Y}$ . Given the factorization, matrix multiplication `%*%`, and `sweep`, can you do this without a loop in one line of code?!?
- (e) In R create a histogram for the marginal distribution of  $Y_1$  and overlay the actual density that you found in 1.5.2 (a).

- (f) Using the samples of  $\mathbf{Y}$ , use matrix multiplication and `sweep` to generate samples of  $\mathbf{Z}$ . Compare the empirical estimates of means, variances, and covariances to the values you obtained in part (g) of 1.5.2.

For the parts that involve R write up using  $\text{\LaTeX}$  and knitr with a Rnw file to generate a PDF. Upload your Rnw file and pdf to Sakai. The other parts may be written using  $\text{\LaTeX}$ /knitr or by hand and scanned in. Suggestion see macro definitions in this repo for help.