

# ST 495 Advanced computing for statistical methods

## Homework problem set 2

January 24, 2023

**No R packages are permitted for use in this assignment.**

1. Using the real data set that you have found as part of Homework 1, to be used for your midterm project, propose a statistical model with population features that may be appropriate to estimate from your data. For example, if your data set includes housing prices in US along with a variety of other housing related covariates, then you might be interested in learning regression parameters for explaining the variation in housing prices. Specifically, if  $Y$  denotes the price of a given house,  $X_1$  is the square footage of the house,  $X_2$  is the city where the house is located, etc., then perhaps you could formulate the linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + U,$$

where  $U$  is a random error variable, and  $\beta_0, \beta_1, \beta_2, \dots$  are the coefficient parameters (i.e., the population features) that you will estimate with your real data. Note that your statistical model need not be a linear model; this is simply an example for illustration.

2. Let  $M \in \mathbb{R}^{p \times p}$  and  $b \in \mathbb{R}^p$ . Then for any solution  $x \in \mathbb{R}^p$ , the system of linear equations

$$Mx = b$$

is said to be in *row echelon form* if the matrix  $M$  is upper triangular. Recall from your linear algebra prerequisite course a system of linear equations can always be re-expressed in row echelon form via a series of row operations. Write a function in R that takes as input (  $M$ ,  $b$  ) where  $M$  is a  $p \times p$  matrix and  $b$  is a  $p$ -dimensional column vector and returns the row echelon form of the arguments  $M$  and  $b$ . This can be done via the Gaussian elimination algorithm ([https://en.wikipedia.org/wiki/Gaussian\\_elimination](https://en.wikipedia.org/wiki/Gaussian_elimination)).

3. Write an algorithm to solve for at least one solution to a row echelon form expression of a system of linear equations (if a solution exists). More specifically, assume you begin

as in question 2 with an arbitrary system  $Mx = b$ . Then passing this system through the function that you wrote for question 2 will produce the row echelon form of the system,  $\widetilde{M}x = \widetilde{b}$ . Your function for this question will be able to solve for at least one solution (if a solution exists) to the system  $\widetilde{M}x = \widetilde{b}$ , which you know is in row echelon form. This can be done via the back substitution algorithm ([https://en.wikipedia.org/wiki/Triangular\\_matrix#Forward\\_and\\_back\\_substitution](https://en.wikipedia.org/wiki/Triangular_matrix#Forward_and_back_substitution))

4. Use the function you wrote for question 3 to finish writing the function in `lecture_code_2.r` for determining an eigenvector associated with an eigenvalue of a given triangular matrix  $A \in \mathbb{R}^{p \times p}$ . Your function should take as an input ( `A`, `lambda`), where `lambda` is a user-supplied eigenvalue of  $A$ , and return an eigenvector associated with `lambda`.