The University of British Columbia

Data Science 581 Modelling and Simulation II

Lab Assignment 2

One of several lessons in R, this lab is about matrices, looping, recursion and apply, and it is a warm-up to the upcoming material on Markov chains.

I uploaded another textbook (*Linear Regression and Generalized Linear Models*) . Chapters 4 and 8 of this book covers what we discuss for lecture 3 and 4. Questions 3 through 5 are from that book.

Please submit only Q3, Q5 and Q6 through canvas.

1. Consider the following 2×3 matrix X,

```
X <-  matrix(seq(1, 6), nrow=3)
X

##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

   (a) Obtain the matrix $H = X(X^TX)^{-1}X^T$, where T denotes matrix transpose and the multiplication is matrix multiplication. In R, you can transpose `X` using `t(X)`. You can multiply matrices `A` and `B` using `A%*%B`.

   (b) Compute H^2, using matrix multiplication. How does the result compare with H?

   (c) Calculate the eigenvalues and eigenvectors of H. Use the `eigen()` function, and see the help file for further information.

   (d) Calculate the trace of the matrix H, and compare with the sum of the eigenvalues. The trace is the sum of the diagonal elements. You can extract the diagonal elements of a matrix using the `diag()` function.

   (e) Calculate the determinant of the matrix H, and compare with the product of the eigenvalues. You can compute the determinant of a matrix using `det()`.

   (f) Using the definition of eigenvector, verify that the columns of X are eigenvectors of H.

2. Consider the following matrix.

$$P = \begin{bmatrix} 0.5 & 0.2 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.7 \\ 0.1 & 0.2 & 0.1 & 0.6 \\ 0.1 & 0.3 & 0.1 & 0.5 \end{bmatrix}$$

   It can be entered into R in a number of ways, including

```
P <- matrix(c(.5, .1, .1, .1, .2, .1, .2, .3, .1, .1, .1, .1, .2, .7, .6, .5), nrow=4)
```

   (a) P is an example of a stochastic matrix, meaning that the sum of the elements of each row is 1. Use the `apply()` function to verify that the row sums add to 1, as in

```r
apply(P, 1, sum)
```

(b) Compute P^n for $n = 2, 3, 5, 10$. Is a pattern emerging?
For example, with $n = 2, 3$ and 5, we would use

```r
P2 <- P%*%P
P3 <- P2%*%P
P5 <- P2%*%P3
```

3. Exercise # 4 from chapter 4 (multiple regressions).
4. Exercise # 15 from chapter 4 (multiple regressions).
5. Exercise #2 from chapter 8 (generalized linear model).
6. Read about the `epil` dataset using `? MASS::epil`. Inspect the dependency of the number of seizures (y) in the age of the patient (age) and the treatment (trt).
   (a) Fit a Poisson regression with glm.
   (b) Are the coefficients significant?
   (c) What is the 95% confidence interval for the estimates of the coefficients.
   (d) Does the treatment reduce the frequency of the seizures?
   (e) According to this model, what would be the number of seizures for 20 years old patient with progabide treatment?