

Biostat 561: Final Homework

Instructor: Amy Willis, Biostatistics, UW

December 5, 2017

Homework due Thursday 14 December, 2 p.m. Strictly no extensions.

Office hours Monday 11 December, 2 p.m., and Wednesday 13 December, 3 p.m.

Link to Final Homework submission: <https://classroom.github.com/a/199ETbeb>

Be sure to upload a `.cpp` file, a `.py` file, and a `.R` file along with a `.pdf` file showing screenshots of your output along with your commentary.

Question 1: C++

Write a C++ function that takes arguments `beta` (a vector of dimension `p`) and `n` (a scalar), generates data from the following model

$$\begin{aligned}X &\in R^{n \times p}, y \in R^n, \epsilon \in R^n \\y &= X\beta + \epsilon \\X_{i1} &= 1 \\X_{i2} &\sim \text{Bernoulli}(0.7) \\X_{i3} &\sim \text{Uniform}(-1, 1) \\\epsilon_i &\sim N(0, 1),\end{aligned}$$

writes X and y to output files, and returns the following estimate of `beta`:

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Choose a vector `beta` of dimension 3, and confirm the output of your function using R's native matrix multiplication function.

You should be using Rcpp to interface the script with R.

Question 2: Python

Repeat Question 1, this time writing a Python script to perform the same task.