Bayesian Methods
Final Report

Group 4

2022/2023

Instructions

1. Download the file `group4.zip`, containing files:

   - `group4_rep_en.pdf`
   - `group4_rep_pt.pdf`
   - `group4_data1.csv`
   - `group4_data2.csv`

2. Submit a PDF with the report and all the code used (`R` and `Stan`) on the device *Report Submission*.

3. **Submission deadline:** June 9, 2023

Questions

1.

Consider the data in `group4_data1.csv` depicting information on real estate transactions. The variables are the following:

X1 transaction date

X2 house age (years)

X3 distance to the nearest metro station (feet)

X4 number of nearby convenience stores

X5 latitude

X6 longitude

Y house price per square feet (dollars)

1. Choose an adequate bayesian model to predict `Y`, the house price per square feet, using the other variables as predictors.

2. Fit the aforementioned model using Monte Carlo Markov Chains.

3. Perform a diagnostics check to ensure the convergence of the generated chains.

4. Summarise the posteriors for the model, numerically and graphically.

5. Use post-predictive diagnostics to check the adequacy of the model (including residual analysis).

6. Are there any variables that are not influent on the house price of unit area? Justify thoroughly.

7. Obtain a 95% highest probability density interval for the price of a house with an area of 150 square feet, sold in the middle of 2013, 17 years of age, 1100 feet away from the nearest metro station, with 4 convenience stores in it's vicinity, situated in latitude 24.95 and longitude 121.50.

2.

In the file `group4_data2.csv` there is the registry of a biological experiment, where plants were planted on 7 different sites with different soil characteristics. The height of the plant (in cm) was measured for each plant at 4, 5, 6, 7, 8, 9, 10 and 11 weeks.

1. Choose an adequate bayesian model to model the plant height, justifying your answer.

2. Fit the aforementioned model using Monte Carlo Markov Chains.

3. Perform a diagnostics check to ensure the convergence of the generated chains.

4. Summarise the posteriors for the model, numerically and graphically.

5. Comment on the general behaviour of the plant's growth, justifying your conclusions.

6. Summarise numerically and analitically the predictive distribution for the height of a plant in site 3 at the 5th week.

7. Summarise numerically and analitically the predictive distribution for the height of a plant in an external site at the 5th week.