DS201/DSL253: Statistical Programming

Assignment 04

13/02/2025

Instructions for Submission: You can submit your solution as a Jupyter Notebook/Matlab file with comments and discussions on the results obtained in each step.

1. Follow Standard Report Format: Include sections like Introduction, Data, Methodology, Results, Discussion, and Conclusion.
2. File Naming Convention: Adhere to the specified naming convention for each file you submit (e.g., RollNumber FirstName Asg1).
3. Refrain from using zip files. If necessary, submit multiple files.
4. Include comments in the code explaining the logic and any assumptions made.
5. Include References: Cite any external sources or references used in your assignment.
6. Code Quality: Ensure your code follows best practices, is well-organized, and avoid plagiarism as a plagiarism check will be conducted.
7. Be aware that late submissions are not permitted; ensure timely submission.
8. Coding can be done in any language.

Question 1

You have been provided with two datasets containing time series signals collected from different regions of the brain. Each dataset contains time series data from 50 brain regions, recorded over 190 time points.

(i) Compute the coactivation (correlation) matrix for the 50 brain regions of `data_1` and `data_2` time series signals.

(ii) Normalize the data in both `data_1` and `data_2` to bring the values between -1 and +1.

(iii) Compute the coactivation (correlation) matrix again after normalization for both datasets.

(iv) Perform Principal Component Analysis (PCA) on the normalized signals to reduce their dimensionality. Compute the coactivation (correlation) matrix for the transformed signals.

(v) Compare the correlation matrices of `data_1` and `data_2` before and after normalization, and after PCA. Interpret the changes in the correlation matrix at each step.

Question 2

If a random variable X is $N(\mu, \sigma^2)$, where $\sigma^2 > 0$, then the random variable $V = \frac{(X-\mu)^2}{\sigma^2}$ follows a chi-squared distribution with 1 degree of freedom $\chi^2(1)$.

Your task is to empirically verify this theorem using coding. You will generate n random samples from a normal distribution, compute the transformed variable V, and compare its distribution to the theoretical $\chi^2(1)$ distribution. Observe the distribution for different values of n.

Question 3

You are given a dataset of numbers that follows a Gaussian (normal) distribution but contains some noise. Your task is to:

(i) Compute the mean μ and variance σ^2 of the dataset.

(ii) Verify the empirical rule (68-95-99.7 rule) by calculating the percentage of data points that fall within $1\sigma, 2\sigma$ and 3σ from the mean.

(iii) Compute the Cumulative Distribution Function (CDF) for the dataset and determine the probability of data points falling beyond 2σ.