Coding Assignment
Aug 11, 2020

Deliverables for all assignments:

- RMarkdown document + compiled HTML version that contains all steps of the assignment.
- If solved in Python, please report results and code in Jupyter Notebook.

The data:

In this coding assignment, we want to analyse a drug high-throughput screen of four anonymised drugs across hundreds of cancer cell lines. The cancer cell lines have been molecular characterised and belong to different cancer types.

We provide you with three different files.

- The cell line information file contains the cell line name and matching COSMIC ID (hint: COSMIC ID is used in all other files to identify the cell lines). This file contains additional meta data about the cell line.
- The pan-cancer binary event matrix (BEM) containing the mutational and copy number status (1=mutated, 0=wild type) of all cell lines across key cancer genes.
- The drug response data contains log10(IC50) values, which is the concentration to reduce cell viability by half across. This file contains drug response data for 4 drugs across all cell lines (note data is sparse).

Assignment 1: Visualise the data

- Visualise the frequencies of cancer types represented in this study.
- The binary event matrix (BEM) contains mutational and copy number features of the cancer cell lines. Visualise the data e.g. the mutational frequencies, oncoprint, etc.
- Visualise the distribution of IC50 values for drug A, B, C and D. Describe what you observe.

Assignment 2: Drug response biomarkers (statistics)

- What are the drug response biomarkers for drug A, B, C and D? Which mutations cause drug sensitivity, which are drug resistance biomarkers?
- What is the difference between a pan-cancer and cancer type specific analysis?
- Be creative in visualising biomarker results.

Assignment 3: Predict drug response (machine learning)

- Predict the drug response of all four drugs based on the mutational and copy number data.
- Which drugs can you predict better than random?
- What are the important features?
- How could you improve your predictions?

Good luck!