MAIS 202 - PROJECT DELIVERABLE 2

Roland T. Killian

Mais202

1. I will be conducting a machine learning project to determine brain tissue malignancy from methylation profiles. I will first use univariate selection to find the most variant positions in the profiles for malignant vs non malignant, and then train a model to take inputted profiles and output malignancy. Depending on how hard it is to acquire good data, this may turn into a tissue identification project.

2. I will be using the NIH GPL13534 database containing hundreds of thousands of samples, select around a thousand containing brain data, and break them down into csv files. Samples can contain over 400,000 features, and this is why feature selection will be so vital. After selecting some 100-10,000 features, I will prepare a training, validation, and test set. By importing CSV's and breaking them down into vectors with the appropriate labels.

3. The model I will be using is a multilayer perceptrons neural network. I have chosen MLP because it is great at classification problems. I also want to understand neural networks better, and what better way to learn than building one! I am still in the data collection phase, as this is not very easy data to get working. I will have progress for the next deliverable.

4. So my next step is to run the feature selection techniques and finish compiling my data. Then to commence the neural network!