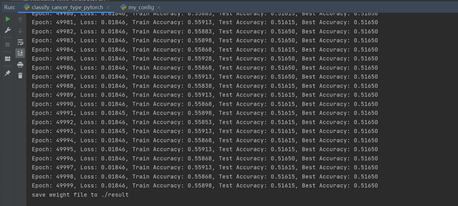
1. **What I have done:**

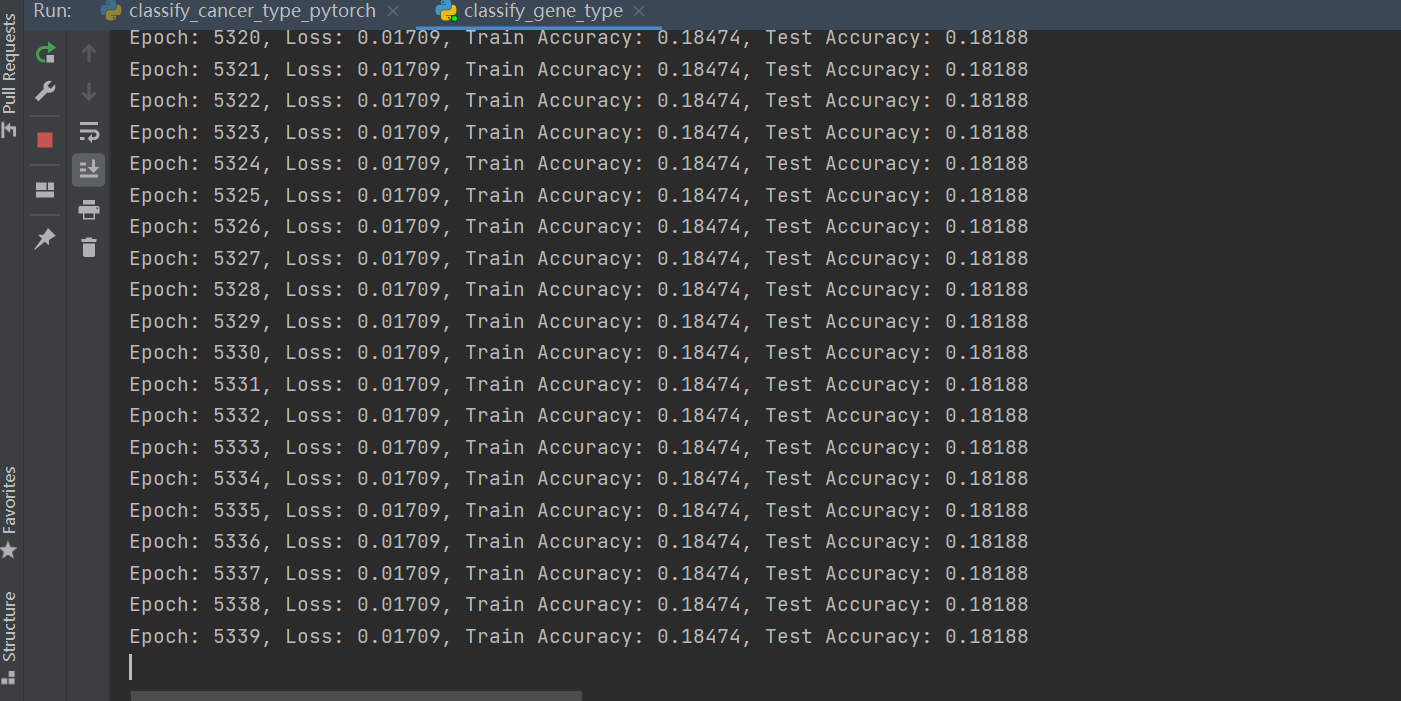
1.I generated all of the (sbs x sample id) matrix for all the cancer types from the maf file I downloaded before.

2.combined sbs x sample id maxtrix with gene mutation status for each of the samples

3.Did multiclass classification on the cancer types with features set as sbs signatures, using the softmax to construct the BPNet to perform the training and testing of the model ,however, the result of the classification is not really good as it has only the accuracy of 51.658



4.The classification of the gene mutation status with sbs signatures applied as feature is also not optimistic as it has only the accuracy of 18 %



5.Wrote the data and data preprocessing part of dissertation but it’s still in modification

**2.The goal:**

1. we look into whether the sbs signatures could be used to determine the cancer types
2. we then find the most powerful sbs signatures for determine such cancer (eg. Lung cancer), then we set those sbs signatures as input and then set the gene mutation status as labels and classify the gene mutation status.
3. To be able to know the mutation status of the genes without sequencing the tumor, we could build recommendation system to predict which of the gene are mutated in different cancer types.
4. **The problems:**
5. The classification is multiclass classification and the data is not huge enough for training, the data is also not representative, thus the classification is not really good.
6. Also the gene mutation status classification using sbs signatures as features is also not optimistic as there are only 65 sbs signatures and 187 genes ,also, in order to classify every mutation status in all of the gene in a cancer type, the 187 gene mutation status was formed as one-hot encoding and that has also undermined the accuracy.
7. After the bad classification result, should we still build a recommendation system on it?