



Università degli Studi di Torino - M.Sc. in Stochastic and Data Science - A.Y. 2021/2022

Final project of Statistical Machine Learning (MAT0043)

Gene selection for cancer type classification

In recent years, medicine has made a great step forward in finding new and efficient therapies for different diseases. Thanks to up-to-date technologies, collecting huge amount of data is no longer an issue, so that one can exploit them to define personalised treatments for patients. In particular, cancer genome scale screens are just one example of these applications. In particular, they provide valuable information about the role of genes in driving cancer growth. Thus, researches has developed a Cancer Dependency Map in order to identify genetic and pharmacologic dependencies. However, this is quite a challenging aim: the dataset is not at all easy to handle (high-dimensional, over than ~ 17.000 features) and the picked drug-targetable genes should only rely on a specific cancer type, thus imbalanced classes.

In this project, we apply sophisticated Statistical Machine Learning algorithms to classify different cancer types and select the most relevant genes. After a quick exploratory analysis with PCA, we try Random Forest, Lasso-SVM and Neural Network classifiers and see how the same technique performs differently according to which tumor we are focusing on. In fact, our classification accuracies range from 45% to 98%.
Aggiungere altre conclusioni

1 Introduction

Scrivere che il cancro e' una malattia molto brutta e brevemente come funziona (aka le cellule impazziscono: sviluppano mutazioni, diventano imprevedibile e causano problemi negli individui...). Il compito della medicina e' trovare delle cure (banana). E qua entra in gico DepMap: DNA arrays che raccolgono mutationi dei geni delle cellule cancerogene.

2 Dataset

3 Methods

4 Results

5 Conclusion and future works

References

- [1] L. Baum, T. Petrie, *Statistical Inference for Probabilistic Functions of Finite State Markov Chains*, in Ann. Math. Stat., 37, 1554-1563, 1966.