Tip1 mainly refers the importance of preprocessing for machine learning study. This operation removes any possible trend related to the order the data instances were collected, data cleaning, that is discarding all the data which have corrupt, inaccurate, inconsistent, or outlier values. In addition, this paper refers to the normalization by feature. These procedures can be conducted by python. And the paper told that I should check that I have enough data to solve this computational biology problem with machine learning.

Tips2 refers to splitting your input dataset into three independent subsets (training set, validation set, test set) to calculating optimized parameter and evaluate performance of machine learning.

Tip3 refers to selecting machine leaning algorithm category such as supervised or unsupervised learning depending on the type of dataset. I studied several algorithms and characteristics of them in this class. The knowledges about these are useful for doing this task.

Tip4 describes which algorithm to select from the categories selected by Tip 3. Here, it is recommended to select the simplest algorithm. I learned several simple algorithms listed in this paper in this class. If these algorithms are inconvenient, I need to learn more complicated algorithm.

Tip5 describes about how to deal with the dataset whose sample labels are imbalanced. This paper recommended to use MCC scores for examining whether your prediction evaluation is going well or not, rather than other metric accuracy or F1 score. This topic was not addressed in this class, so I may have to study this if I handle imbalanced dataset.

Tip6 describes about optimizing the hyper-parameters of a machine learning algorithm. Hyper parameter is a parameter which is user defined value and not learned by the algorithm directly. This paper refers to grid search as one of the methods for deciding hyper parameter of a machine learning algorithm.

Tip7 is associated with bias-variance trade of, which is the topic we study in detail in this class. “Minimizing overfitting”, the title of Tip7, corresponds the effort for reducing the variance of the model learned by a machine learning algorithm.

Tip8 describes about how to evaluate algorithm performance, index such as MCC, acuracy, F1 score or precision-recall curve. These can be calculated by drawing the value of the

contingency table. There were several indices, and I learned a part of them in this class.

Tip9 recommends that I use open source software. In my machine learning research, I use python, R, shell script on bash, all of which are open source

Tip10 describes about importance of asking for feedback and help to experts.

I am in a very privileged environment in this respect, so I am very grateful to my teachers and colleagues.