Or Samimi Golan TBD

Shajaf Gonen 204067268

Unless specified otherwise in an alternative, we generated our classifiers in the following form:

1. SNP-wise. Meaning, that we generated a classifier for each missing SNP in the test set (300 classifiers in total in each alternative).
2. Each sample in the train (test) set, is the 100 SNPs before and after the SNP to be learned (200 feature in the feature vector in total). The matching label of a sample, is of course the value of the SNP. (Known in the train set = {0, 1, 2}, unknown in the test set = {-1}). These details are generated out of 'extracted\_train' and 'extracted\_test'
3. The success rate is calculated using 10-fold validation, and 3-fold validation.

Alternative 1 - SVM – one vs one using libsvm

Libsvm handles mutli-class classification, using one-vs-one technique (learned in class).

By running: svmtrain, using the following options.

Success rates

Linear SVM

C = 1: 69.47%

C = 0.1: 72.62%

C = 0.01: 70.69%%

Polynomial SVM

C = 1, degree = 4: 73.25%

C = 0.1, degree = 10: 72.37%

C = 1, degree = 10: 72.30%

C = 100, degree = 10: 72.30%

Radial basis SVM

C = 1, gamma = 1/features\_number (default value)= 67.51%

C = 100000, gamma = 1/features\_number (default value) = 73.69%

Alternative 2 – Adaboost - TBD

Summary

|  |  |  |
| --- | --- | --- |
| Script Name | Description Summary | Success Rate (10-fold-validation) |
|  |  |  |
|  |  |  |
|  |  |  |