Project Plan KB8024 Shuhan Xu

Week 1

* Complete script for data parsing
* Complete script for data prepressing
* Read scikit learn documentation on Support Vector Machines (1.4)
* Read scikit learn documentation on cross validation (3.1)
* Read scikit learn documentation on model evaluation (3.3)
* Work on script for k-fold cross-validation

Week 2

* Complete script for k-fold cross-validation
* Plot learning curve of predictor to diagnose bias and variance
* Brainstorm more features to include in predictor
* Use cross validation to optimize C and gamma parameters
* Use the optimum parameters to train the SVM
* Test the SVM on test set

Week 3

* Read the paper on using PSSM to predict protein secondary structure by Jones DT (1999)
* Use psi-blast to extract evolutionary information for my SVM
* Train my SVM based on multiple sequence alignment
* Use cross validation to optimize parameters for the improved SVM
* Use the optimum parameters to train the improved SVM
* Test the SVM on test set
* Prepare and present paper

Week 4

* Read scikit learn documentation on random forest and decision tree
* Create random forest and decision tree predictors
* Compares these predictors to my SVM predictor
* Test performance of my predictors on 50 new proteins
* Review my predictor based on current literature
* Write report for the project

Week 5

* Submit report