Documentation for Classification

We constructed five classifiers using Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (kNN), Decision Tree (DT) and Neural Net (NN). This document describes the details of the classifiers, include range of parameter grid, the parameters and functions we used for each classifier. The classifiers are constructed using scikit-learn (http://scikit-learn.org/stable/), a machine learning library in Python.

Range of parameter grid

We use *GridSearchCV* to find the optimal parameters, the parameter ranges of grid search are:

- SVM: ‘C’: [1:1:10] / ‘kernel’: linear, rbf / ‘decision_function’: ovo, ovr
- RF: ‘n_estimators’: [5:1:20] / ‘criterion’: gini, entropy / ‘max_features’: auto, sqrt, log2, None
- kNN: ‘n_neighbors’: [5:1:15] / ‘weights’: uniform, distance / ‘algorithm’: auto, ball_tree, kd_tree, brute / ‘p’: [1:1:3]
- DT: ‘criterion’: gini, entropy / ‘splitter’: best, random / ‘max_features’: auto, sqrt, log2, None
- NN: ‘hidden_layer_size’: [5:1:25] / ‘activation’: identity, logistic, tanh, relu / ‘solver’: lbfgs, sgd, adam / ‘learning_rate’: constant, invscaling, adaptive

Parameters and functions

- SVM: SVC (C=1, kernel = ‘linear’, decision_function_shape = ‘ovo’, probability = 1, class_weight = w)
- RF: RandomForestClassifier (n_estimators = 19, criterion = ‘entropy’, max_features = ‘auto’, class_weight = w)
- kNN: KNeighborsClassifier (n_neighbors = 10, weights = ‘uniform’, algorithm = ‘auto’, p = 1)
- DT: DecisionTreeClassifier (criterion = ‘entropy’, max_features = None, splitter = ‘random’, class_weight = w)
- NN: MLPClassifier (hidden_layer_sizes = 15, activation = ‘tanh’, solver = ‘adam’, learning_rate = ‘constant’)