CST 383: Intro to Data Science
Dr. Glenn Bruns

Hyperparameter tuning

Purpose. Many machine learning algorithms have "hyperparameters" that need to be tuned. In this assignment you will deepen your understanding of the hyperparameter concept and how to tune hyperparameters.

Instructions. The following reading is recommended but not required. Please at least skim the readings and read what you find most interesting or important. The books for our course can be found at this playlist.

- *Python Data Science Handbook*
  - Section 'Hyperparameters and Model Validation' in Chapter 5, up to, but not including subsection 'The bias-variance trade-off'. (about 5 pages)
- *Introduction to Machine Learning with Python*
  - Section 5.2, up to, but not including subsection 'Search over spaces that are not grids'. (about 10 pages)

Answer the following questions by downloading and editing hyperparams.txt.

1. In a classification problem, the goal with hyperparameter tuning is to achieve:
   a. high training accuracy (in other words, high accuracy on the training data)
   b. high test accuracy
   c. low training accuracy
   d. low test accuracy
2. If there are 6 hyperparameters, each with 3 possible values, how many hyperparameter value combinations are there?
3. (T/F) Hyperparameters have quantitative values, not categorical values.
4. (T/F) In KNN classification, the best choice of hyperparameter values will depend on the data set being used (for example, heart disease data or college data).
5. In the following call to GridSearchCV, how many hyperparameter value combinations will be tested?

```
param_grid = {'n_neighbors': np.arange(3, 12, 2),
              'weights': ['uniform', 'distance']}
clf = GridSearchCV(KNeighborsClassifier(), param_grid)
clf.fit(X_train, y_train)
```

Submission. Submit hyperparams.txt on Canvas.