

homework7

December 13, 2022

1. Find a classification problem at UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>) or similar, and compare the classification performance of a trained k -NN learner with that of a trained SVM (support vector machine). Explore using the *tune* function in the library *e1071* which contains an implementation of SVMs. Write a paragraph discussing your results.

```
[1]: from sklearn.model_selection import train_test_split
from sklearn.datasets import make_classification

X, y = make_classification(n_samples = 1000, n_features = 10, n_informative = 2,
    ↪ n_redundant = 0, random_state = 0, shuffle = False)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2,
    ↪ random_state = 1234)
```

```
[2]: from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC

clf = make_pipeline(StandardScaler(), SVC(gamma='auto'))
clf.fit(X, y)
clf.score(X_test, y_test)
```

[2]: 0.965

```
[3]: from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
neigh.fit(X, y)
neigh.score(X_test, y_test)
```

[3]: 0.955

The trained SVM scored 96.5% accuracy on a classification problem similar to the UCI Machine Learning Repository. The k -NN learner scored 95.5% accuracy. Coding is cool!