MATH 637: Mathematical Techniques in Data Science
Lab

Dominique Guillot

Departments of Mathematical Sciences
University of Delaware

April 17, 2020

```
from sklearn.svm import SVC
```

1. Load the zip dataset.
2. Train a support vector machine model on the given training set (use default parameters).
3. Compute the accuracy of your prediction on the given test set.
4. Attempt to improve your model by optimizing over the "C" parameter.
5. Try other kernels other than the default rbf kernel to see if they improve the accuracy of the model.

Decision trees – Titanic dataset

```
from sklearn.tree import DecisionTreeClassifier
            sklearn.tree.plot_tree
```

1. Load the titanic dataset and have a look at the description.
2. Construct a decision tree to predict whether or not a given passenger survived. (You may want to limit the depth of the tree (`max_depth` option).
3. Plot your estimated decision tree.

Exporting the tree

```
from sklearn import tree

tree.plot_tree(tree_model)

fn = ["Pclass", "Sex", "Age", "SibSp", "Parch", "Fare",
        "Embarked"]
cn = ["NotSurvive", "Survive"]

fig, axes = plt.subplots(nrows = 1,ncols = 1, dpi=1000)
tree.plot_tree(clf, feature_names = fn, class_names = cn)
fig.savefig('titanic_tree.png')
```