CS167: Machine Learning

Scikit-learn Practice:

kNN (classifier and regressor)
weighted kNN (classifier and regressor)
Decision Tree (classifier and regressor)

Wednesday, October 9th, 2024

Drake
UNIVERSITY

Announcements

- Notebook #3 Cross Validation with kNN and Vehicle Fuel Efficiency
  - Due tonight 10/09 by 11:59pm
  - To submit, download the `ipynb` file from Colab

Introduction to Scikit Learn Library

- `scikit-learn` is one of Python's main Machine Learning Libraries.

  "It is an open source machine learning library that supports supervised and unsupervised learning. It also provides various tools for model fitting, data preprocessing, model selection, model evaluation, and many other utilities."

  - built on NumPy, SciPy, and matplotlib
  - plays nicely with `pandas`
  - https://scikit-learn.org/stable/

Introduction to Scikit Learn 'Algorithm'

- When working in Scikit Learn (`sklearn`), there is a general pattern that we can follow to implement any supported machine learning algorithm. It goes like this:

  - Load your data using `pd.read_csv()`
  - Split your data `train_test_split()`
  - Create your classifier/regressor object
  - Call `fit()` to train your model
  - Call `predict()` to get predictions
  - Call a metric function to measure the performance of your model

Today's Agenda

- More exercises using sklearn library:
  - StandardScalar() for normalization
  - kNN from sklearn
    - Classifier vs. Regressor
  - weighted kNN from sklearn
    - Classifier vs. Regressor
  - DecisionTree from sklearn
    - Classifier vs. Regressor
- These exercises will be useful for your next two assignments (Notebook#4 as well as Project#1)