COM6018 Data Science with Python

Lab 6: Introducing Scikit-Learn

Jon Barker

(c) 2023, 2024, Jon Barker, University of Sheffield

In this lab

Using Scikit-Learn to explore a classification task

- Splitting a dataset into training and test sets

- Building a k-Nearest Neighbours classifier

- Tuning the hyperparameters of a classifier

- Using Leave-One-Out Cross-Validation to evaluate a classifier

- Using a confusion matrix to look at the patterns of errors.

The Task

- We will be building a land-mine detection system.

- Measurements are taken from a metal detector.

- We predict whether the object is a land-mine or not.

Background

- Landmines are a major problem in many parts of the world.

- They can be detected with a metal detector type device.

- The metal detect outputs a voltage that is proportional to the amount of metal in the ground.

- The measurement is effected by the height of the sensor above the ground and the amount of moisture in the soil.

The Dataset

We will be using a dataset that is described in

- Yilmaz, C., Kahraman, H. T., & Söyler, S. (2018). *Passive mine detection and classification method based on hybrid model*. IEEE Access, 6, 47870-47888

The data has been made available on the UCI Machine Learning Repository,

- https://archive.ics.uci.edu/dataset/763/land+mines-1

About the Features

338 samples of measurements from a metal detector. Each has 3 features

- Voltage (V): The output voltage of the magnetic sensor

- Height (H): The height of the magnetic sensor from the ground

- Soil Type (S): A value that corresponds to the amount of moisture in the soil

About the Labels

Each sample belongs to one of 5 classes:

- 1: No Landmine

- 2: Anti-Tank Mine

- 3: Anti-Personnel Mine

- 4: Booby trapped Anti-Tank Mine

- 5: M14 Anti-Personnel Mine

There are roughly equal numbers of samples in each class.

Classification or Detection?

- We will start out by treating this as a 5-class **classification** problem.

- We will then reconsider the problem as a **detection** problem, i.e., with just two classes: Mine or No Mine.

Obtaining the Jupyter Notebook

If you have cloned and pulled the module's GitHub repository then you should see,

```
materials/labs/
├── 060_introducing_scikit_learn.ipynb
├── ... etc
├── data
     ├── data/Mine_Dataset.xls
     ├── ... etc
```

The lab is `060_introducing_scikit_learn.ipynb` and it will need the data file `data/Mine_Dataset.xls`.

Or you can download the notebook and data via links on Blackboard.

Getting Help

- If you are stuck just raise a hand to ask for help.

- Feel free to discuss the lab with your neighbours.

- Re-read the Scikit-Learn tutorial notes

  - In the Git repo at

    `materials/tutorials/060_Introducing_Scikit_Learn.ipynb`

  - or online at https://uos-com-6018.github.io/COM6018

- Use the Scikit-Learn API documentation for reference. https://scikit-learn.org/stable/modules/classes.html