SK Learn - Decision Tree

Exercise 1:

https://www.kaggle.com/datasets/pablomgomez21/drugs-a-b-c-x-y-for-decision-trees

Exercise 2:

https://www.w3schools.com/python/python_ml_decision_tree.asp

Exercise 3:

A common practice exercise for the sklearn.tree.DecisionTreeClassifier involves `sklearn.datasets` - for sample datasets - load_iris

SK Learn - SVM

Exercise 4:

Sklearn Buildin dataset - breast_cancer

Exercise 5:

A common practice exercise for the sklearn.tree.DecisionTreeClassifier involves `sklearn.datasets` - for sample datasets - load_iris

==========

General Exercises for SK-Learn classification/Clustering datasets

Scikit-learn (sklearn) provides a variety of datasets suitable for classification tasks, categorized into three main types:

- **Toy Datasets:**

  These are small, standard datasets embedded within the sklearn.datasets module and can be loaded directly without downloading. Examples include:

- load_iris(): The classic Iris flower dataset for multi-class classification.
- load_digits(): A dataset of handwritten digits for multi-class classification.
- load_wine(): A dataset for classifying wines based on chemical analysis.
- load_breast_cancer(): A dataset for binary classification of breast cancer malignancy.
- **Real-world Datasets (Fetched Datasets):**

  These are larger datasets that need to be downloaded from the internet the first time they are accessed. Examples include:

- fetch_olivetti_faces(): A dataset of human faces for face recognition.
- fetch_20newsgroups(): A text dataset for document classification.
- **Generated Datasets:**
  These functions allow you to create synthetic datasets with controlled properties for specific machine learning tasks, useful for understanding algorithm behavior or testing. For classification, key functions include:
- make_classification(): Generates a random N-class classification dataset with controllable features, clusters per class, and other parameters.
- make_blobs(): Generates isotropic Gaussian blobs for clustering, which can be adapted for classification by assigning labels to the blobs.
- make_gaussian_quantiles(): Generates data for a quantile-based classification problem.

Part 1 - "All classification SK Learn" algo -To DO:

Take a dataset above and apply "all classification SK Learn" algo(s) studied and follow – normal Machine learning workflow

Part 2 - "All Clustering SK Learn" algo To DO:

Take a dataset above and apply "all clustering SK Learn" algo(s) studied and follow – normal Machine learning workflow

Put code in your personal Public code GitHub repository.