Machine Learning Package

Portfolio of Machine Learning Algorithms

Assignment 1
Datasets

Degree in Applied Data Science

2024/2025

Class Dataset

- In the "**data**" folder, add the "**dataset.py**" module, which will contain the "**Dataset**" class.

- class Dataset:
  - Attributtes:
    - X – matrix/table of features (independent variables)
    - y – vector of the dependent variable
    - features – vector of feature names
    - label – name of the dependent variable

  - Methods:
    - shape – returns the dimensions of the dataset (X)
    - has_label – returns whether the dataset has a label (y)
    - get_classes – returns the classes of the dataset (possible values of y)
    - get_mean, get_variance, get_median, get_min, get_max – returns mean variance, median, minimum, and maximum value for each feature/dependent variable
    - summary – returns a pandas DataFrame with all descriptive metrics
    - dropna – removes all samples containing at least one null value (NaN).
    - fillna - replaces all null values with another value or the mean or median of the feature.
    - remove_by_index - removes a sample by its index.

io sub-package

- Now, add another sub-package named "io" with two modules called "csv_file.py" and "data_file.py." We will add functions to read and write datasets.

- def read_csv
  - arguments:
    - filename – name/path of the file
    - sep – value separator
    - features – boolean. Does the file have feature names?
    - label – boolean. Does the file have a label (y)? (for simplicity, if yes, assume it's the last column)
  - expected output:
    - Dataset object
    - Reads the specified file and returns a Dataset object
    - Hint: you can use the panfas package

io sub-package

- def write_csv
  - arguments:
    - filename – name/path of the file
    - dataset – Dataset object to eroite to a csv
    - sep – value separator
    - Features - boolean. Does the file have feature names?
    - label – boolean. Does the file have a label (y)?
  - expected output:
    - Writes the specified file with the provided arguments.
    - Hint: you can use the panfas package

io sub-package

- def read_data_file
  - arguments:
    - filename – name/path of the file
    - sep – value separator
    - label – boolean. Does the file have a label (y)?
  - expected output:
    - Dataset object
    - Reads the specified file and returns a Dataset object.
    - Hint: You can use modules from other packages like numpy.genfromtxt

io sub-package

- def write_data_file
  - arguments:
    - filename – name/path of the file
    - dataset – Dataset object to eroite to a csv
    - sep – value separator
    - label – boolean. Does the file have a label (y)?
  - expected output:
    - Writes the specified file with the provided arguments.
    - Hint: You can use modules from other packages like numpy.savetxt