Python Dataframe

pandas library

Read from a CSV file

- import pandas as pd
-
- df = pd.read_csv('data/iris.csv')
- print(type(df))
- ---
- <class 'pandas.core.frame.DataFrame'>
-

dataframe

- Print first 5 rows
  - print(df.head())
- Print name of the columns
  - print(df.columns)
- Print Number rows and columns
  - print(df.shape)

dataframe

- print(df.index) - rows index
- print(df.iloc[1]) - index a row
- print(df.iloc[1,1]) – index a entry
- print(df.dtypes) – types of the columns
  - columns all have the same type

Input and Output dataframes

- X = df[["sepal_length","sepal_width","petal_length","petal_width"]]
- y = df["species"]
- print(type(X))
- print(type(y))
- print(X.shape)
- print(y.shape)

Classification

- The iris dataset is typically used for classification.
- Input the values of the columns, "sepal_length","sepal_width","petal_length","petal_width"
- Output – classification (setosa, versicolor, verginica)
- X used as input (multiple columns)
- y output (single column)