Loading data

TYPES OF DATA- CSV-COMMA SEPARATED VALUE, EXCEL, JASON

The size of class is 10 bcz we are working with digits so numbers are from 0-9 so the total number of class is 10.

scilearn digit dataset (reference):https://scikit-learn.org/stable/datasets/toy_dataset.html

```
from sklearn import datasets #load the dataset
digits = datasets.load_digits() #load digit dataset
features=digits.data #create features matrix
target=digits.target #create target vector
features[0] #view first observation
```

```
array([ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.,  0.,  0., 13., 15., 10.,
       15.,  5.,  0.,  0.,  3., 15.,  2.,  0., 11.,  8.,  0.,  0.,  4.,
       12.,  0.,  0.,  8.,  8.,  0.,  0.,  5.,  8.,  0.,  0.,  9.,  8.,
        0.,  0.,  4., 11.,  0.,  1., 12.,  7.,  0.,  0.,  2., 14.,  5.,
       10., 12.,  0.,  0.,  0.,  0.,  6., 13., 10.,  0.,  0.,  0.])
```

scilearn iris dataset (reference):https://scikit-learn.org/stable/datasets/toy_dataset.html

```
from sklearn import datasets #load the dataset
iris = datasets.load_iris() #load iris dataset
features=iris.data #create features matrix
target=iris.target #create target vector
features[0] #view first observation
```

```
array([5.1, 3.5, 1.4, 0.2])
```

MAKE REGRESSION Reference:

Double-click (or enter) to edit

```
from sklearn.datasets import make_regression
features,target,coefficients = make_regression(n_samples=100,
                                               n_features=3,
                                               n_informative=3,
                                               n_targets=1,
                                               noise=0.0,
                                               coef=True,
                                               random_state=1)

print("Feature Matrix:\n {}".format(features[:3]))
print("Target:\n {}".format(target[:3]))
```

```
Feature Matrix:
 [[ 1.29322588 -0.61736206 -0.11044703]
 [-2.793085    0.36633201  1.93752881]
 [ 0.80186103 -0.18656977  0.0465673 ]]
Target:
 [-10.37865986  25.5124503   19.67705609]
```

Make dataset for classification

```
from sklearn.datasets import make_classification
features, target=make_classification(n_samples=100,
                                     n_features=3,
                                     n_informative=3,
                                     n_redundant=0,
                                     n_classes=2,
                                     weights=[.25,.75],
                                     random_state=1)

print("Feature matrix\n {}".format(features[:3]))
print("Target matrix\n {}".format(target[:3]))
```

```
Feature matrix
 [[ 1.06354768 -1.42632219  1.02163151]
 [ 0.23156977  1.49535261  0.33251578]
 [ 0.15972951  0.83533515 -0.40869554]]
Target matrix
 [1 0 0]
```

CLUSTERING TECHNIQUES

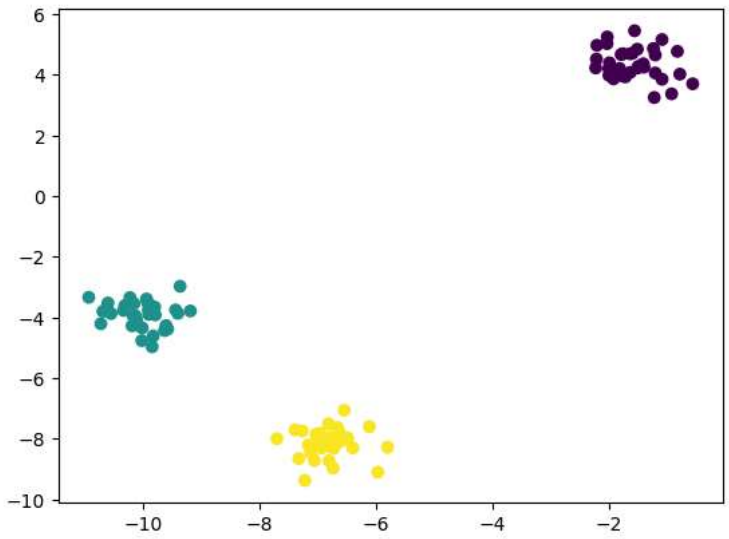
```
from sklearn.datasets import make_blobs
features, target=make_blobs(n_samples=100,
                            n_features=2,
                            centers=3,
                            cluster_std=0.5,
                            shuffle=True,
                            random_state=1)

print("Feature Matrix\n {}".format(features[:3]))
print("Target Matrix\n {}".format(target[:3]))
```

```
Feature Matrix
 [[ -1.22685609   3.25572052]
 [ -9.57463218  -4.38310652]
 [-10.71976941  -4.20558148]]
Target Matrix
 [0 1 1]
```

PLOTTING

```
import matplotlib.pyplot as plt
plt.scatter(features[:,0], features[:,1],c=target)
plt.show()
```



LOADING A CSV FILE

```
import pandas as pd
import os
df = pd.read_csv("/content/sample_data/california_housing_train.csv")
df.head(2)
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
0	-114.31	34.19	15.0	5612.0	1283.0	1015.0	472.0	1.4936	66900.0
1	-114.47	34.40	19.0	7650.0	1901.0	1129.0	463.0	1.8200	80100.0

Next steps: Generate code with df View recommended plots

LOADING AN EXCEL FILE

```
import pandas as pd
import os
df = pd.read_excel("/file_example_XLSX_50.xlsx")
df.head(2)
```

	0	First Name	Last Name	Gender	Country	Age	Date	Id
0	1	Dulce	Abril	Female	United States	32	15/10/2017	1562
1	2	Mara	Hashimoto	Female	Great Britain	25	16/08/2016	1582

Next steps: Generate code with df View recommended plots

LOADING A JASON FILE

```
!pip install wget
```

```
Collecting wget
  Downloading wget-3.2.zip (10 kB)
  Preparing metadata (setup.py) ... done
Building wheels for collected packages: wget
  Building wheel for wget (setup.py) ... done
  Created wheel for wget: filename=wget-3.2-py3-none-any.whl size=9656 sha256=a3b338e56b7e3bba907252e620765d0b07635689b888bfaaf113d7b2eb
  Stored in directory: /root/.cache/pip/wheels/8b/f1/7f/5c94f0a7a505ca1c81cd1d9208ae2064675d97582078e6c769
Successfully built wget
Installing collected packages: wget
Successfully installed wget-3.2
```

```
!wget https://raw.githubusercontent.com/LearnWebCode/json-example/master/animals-1.json -O animals-1.json
```

```
--2024-04-22 06:43:38--  https://raw.githubusercontent.com/LearnWebCode/json-example/master/animals-1.json
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.109.133, 185.199.110.133, 185.199.111.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.109.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 420 [text/plain]
Saving to: ‘animals-1.json’

animals-1.json      100%[===================>]     420  --.-KB/s    in 0s

2024-04-22 06:43:38 (5.81 MB/s) - ‘animals-1.json’ saved [420/420]
```

```
import pandas as pd
df = pd.read_json("animals-1.json", orient="columns")
df.head(2)
```

	name	species	foods
0	Meowsy	cat	{'likes': ['tuna', 'catnip'], 'dislikes': ['ha...
1	Barky	dog	{'likes': ['bones', 'carrots'], 'dislikes': ['...

Next steps: Generate code with df View recommended plots

QUERING A SQL DATABASE

```
import pandas as pd
import sqlite3

con=sqlite3.connect("/student.db")
df=pd.read_sql_query("SELECT * from students ",con)
df.head()
```

	registration_no	name	class	gender	dob	date_of_registration	religion	email_id	father_name	mother_name
0	1	ywgrrrrrr	Sycs	Male	5/20/05	24/02/2024	Christian	wsgn	wsayghj	qsaj
1	2	swygauh	Fycs	Male	5/27/05	24/02/2024	Christian	tsfxyghjbn	gdsgcuysc	gsujbnx
2	3	yghsqab	Sycs	Male	5/26/05	25/02/2024	Hindu	qswax	sxyugzhj	edytsyiuhjk
3	4	sgyauhj	Tycs	Male	5/13/05	25/02/2024	Muslim	pillaiabincs232414@gmail.com	dsyuxgzhj	dsyxguhjk
4	5	Sahil	Fycs	Male	6/16/05	25/02/2024	Christian	atysghzj	stxyugzh	6styuzh

Next steps: Generate code with df View recommended plots