

Machine Learning

Lecture 8: Create Your First Project

COURSE CODE: CSE451

2021

Course Teacher

Dr. Mrinal Kanti Baowaly

Associate Professor

Department of Computer Science and
Engineering, Bangabandhu Sheikh
Mujibur Rahman Science and
Technology University, Bangladesh.

Email: baowaly@gmail.com



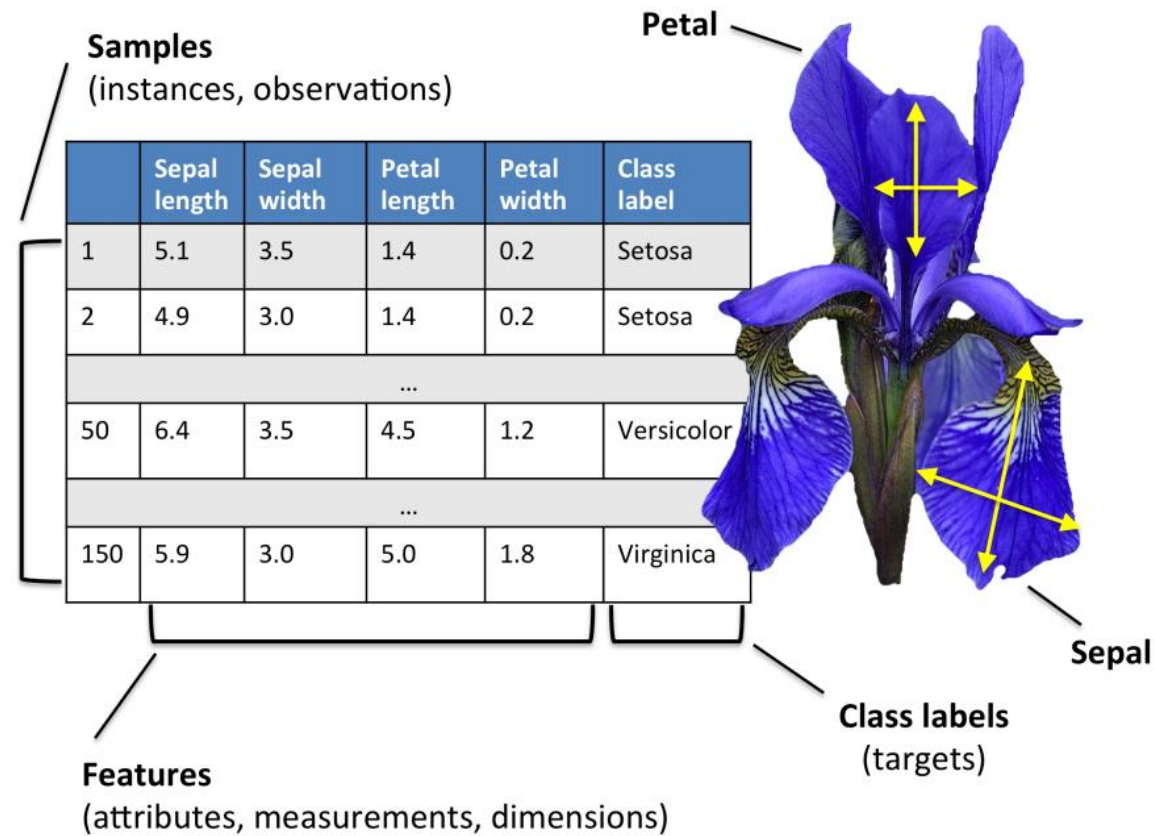
Iris flower classification



Iris Flower Dataset

- 150 samples
- 3 labels/categories: Species of Iris (Iris setosa, Iris virginica and Iris versicolor)
- 4 features: Sepal length, Sepal width, Petal length, Petal Width in cm
- Download Link: [Kaggle](#)

Iris dataset instances



Import libraries

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import tree
from sklearn.metrics import accuracy_score
```

Load the dataset

```
iris_data = pd.read_csv('IRIS.csv')
```

Summarize the dataset

dimensions (no. of rows & columns)

```
print(iris_data.shape)
```

list of columns/features

```
print(iris_data.columns)
```

peek some data

```
print(iris_data.head(10))
```

statistical summary

```
print(iris_data.describe())
```


Specify the target variable and its distribution

target variable

```
target = iris_data['species']
```

distribution of class labels or categories

```
print(pd.value_counts(target))
```

Specify the target variable and its distribution

target variable

```
target = iris_data['species']
```

distribution of class labels or categories

```
print(pd.value_counts(target))
```

alternative of finding class distribution

```
print(iris_data.groupby('species').size())
```

Split dataset into training and test data

```
seed = 7
```

```
train_data, test_data = train_test_split(iris_data, test_size=0.3,  
random_state= 7)
```

```
# shape of the datasets
```

```
print('\nShape of training data :',train_data.shape)
```

```
print('\nShape of testing data :',test_data.shape)
```

```
# class distribution of the training data
```

```
print(pd.value_counts(train_data['species']))
```

```
# class distribution of the test data
```

```
print(pd.value_counts(test_data['species']))
```

Balanced split of the dataset

seed = 7

```
train_data, test_data = train_test_split(iris_data, test_size=0.3,  
random_state=seed, stratify=target)
```

Why do we use random state in splitting dataset?

Find in on [stackoverflow](#).

Separate the independent and target variables

separate the independent and target variables from training data

```
train_x = train_data.drop(columns=['species'],axis=1)
```

```
train_y = train_data['species']
```

separate the independent and target variables from test data

```
test_x = test_data.drop(columns=['species'],axis=1)
```

```
test_y = test_data['species']
```

Build the model

create a classifier object/model

```
model=tree.DecisionTreeClassifier()
```

train the model with fit function

```
model.fit(train_x, train_y)
```

Make predictions

make predictions on training data

```
predictions_train = model.predict(train_x)
print('\nTraining Accuracy :', accuracy_score(train_y,
predictions_train))
```

make predictions on test data

```
predictions_test = model.predict(test_x)
print('\nTest Accuracy :', accuracy_score(test_y, predictions_test))
```

Home work for the Lab.

- ✓ Apply some preprocessing tasks
 - Normalization
 - Standardization
- ✓ Apply different classifiers and compare their performances
 - Logistic Regression (LR)
 - K-Nearest Neighbors (KNN)
 - Support Vector Machines (SVM)
- ✓ Compute training accuracy, testing accuracy for each model
- ✓ Find the best model for the prediction task

Some example projects

Iris classification [[Link](#)]

Your First Machine Learning Project in Python Step-By-Step [[Link](#)]

24 Data Science Projects To Boost Your Knowledge and Skills [[link](#)]

6 Complete Machine Learning Projects [[Link](#)]