



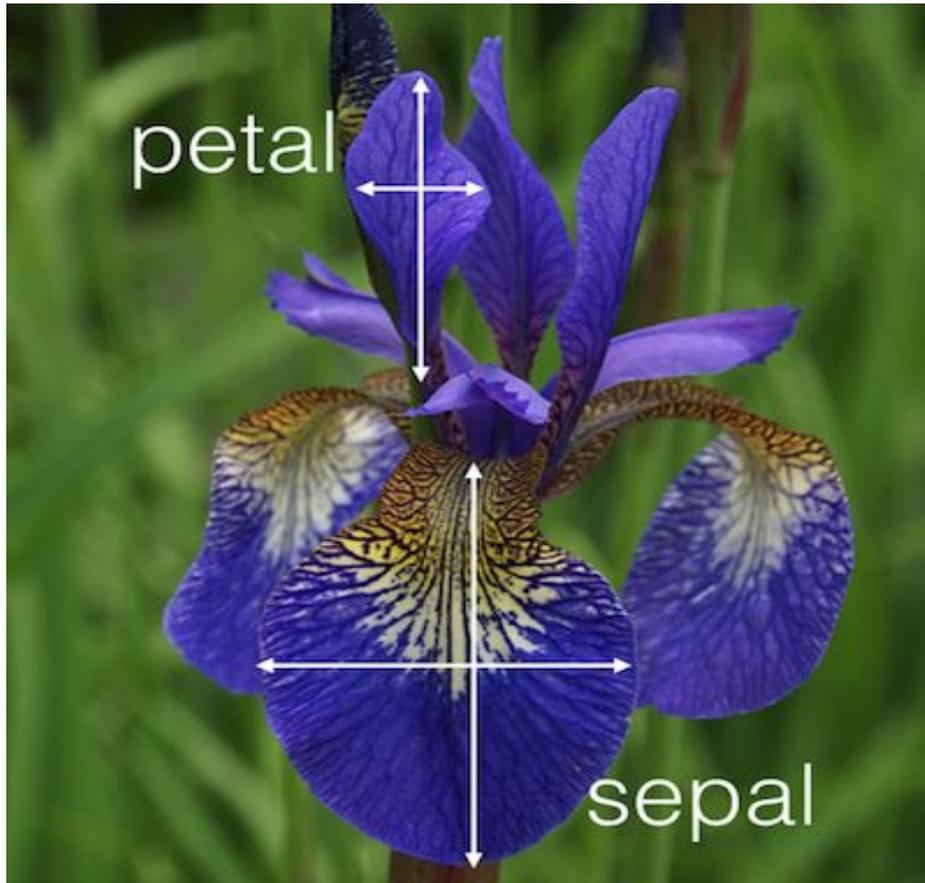
הטכניון
מכון טכנולוגי
לישראל

יישום KNN באמצעות ספריית SKLearn

דוגמא: סיווג אירוסים

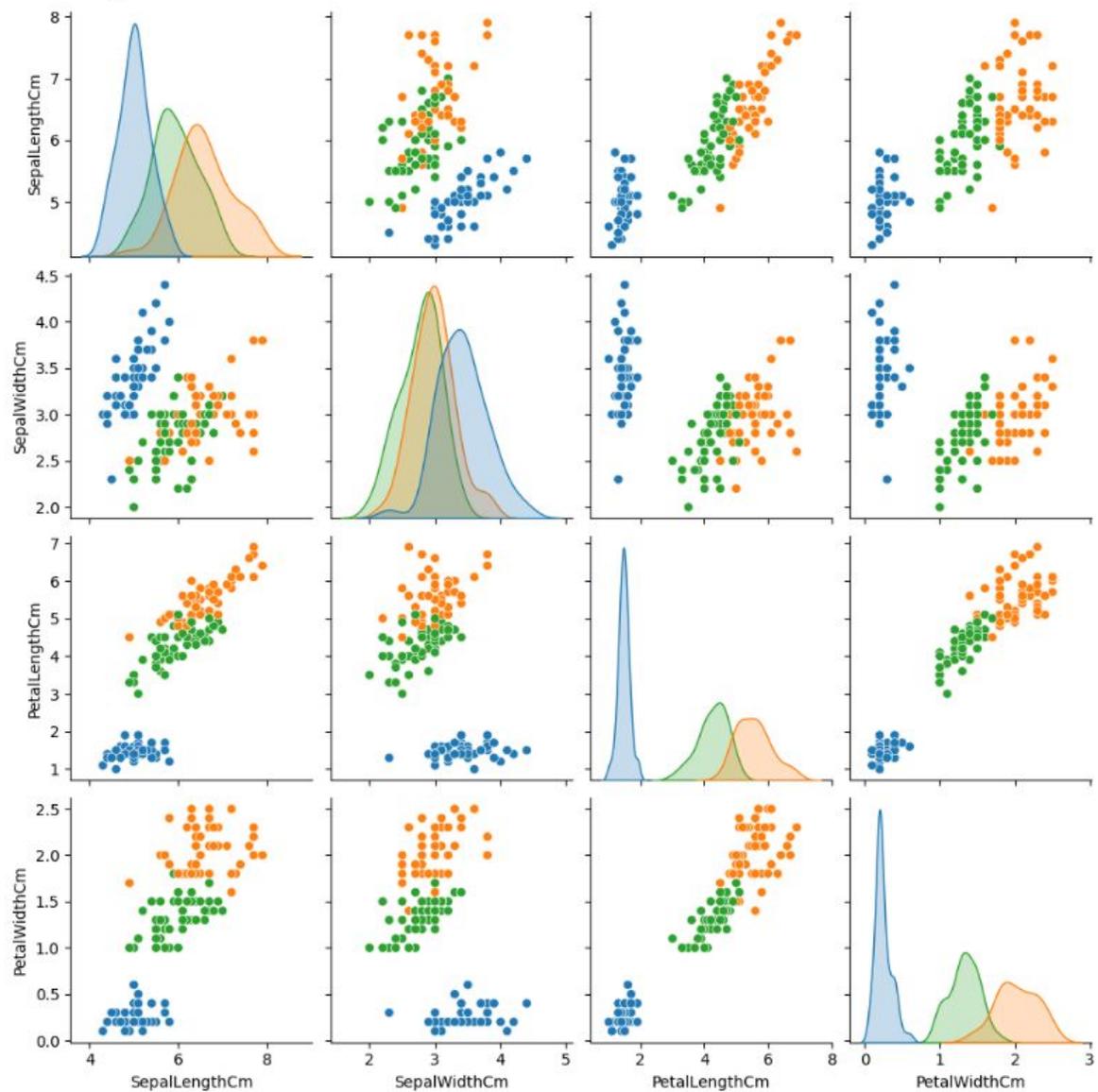


מאפיינים של פרח



משמעות	מאפיין
אורך עלי גביע	Sepal length
רוחב עלי הגביע	Sepal width
אורך כלי הכותרת	Petal length
רוחב עלי הכותרת	Petal width

```
sns.pairplot(df,hue='Species')
```



פיזור המאפיינים

ייבוא ספריות

```
import pandas as pd
import numpy as np
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
```

קריאת נתונים עם ספריית pandas

```
#load dataset
df = pd.read_csv('https://raw.githubusercontent.com/arielb30/datasets/main/Iris.csv')
df
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

קריאת נתונים עם ספריית pandas

```
#drop ID column
df.drop('Id',axis=1,inplace=True)
#shuffle rows
df = df.sample(frac = 1)
#reset index
df.reset_index(drop=True,inplace=True)
df
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	6.3	2.3	4.4	1.3	Iris-versicolor
1	4.7	3.2	1.6	0.2	Iris-setosa
2	5.1	3.7	1.5	0.4	Iris-setosa
3	6.2	3.4	5.4	2.3	Iris-virginica
4	5.8	2.6	4.0	1.2	Iris-versicolor
...
145	5.0	2.3	3.3	1.0	Iris-versicolor
146	7.2	3.2	6.0	1.8	Iris-virginica
147	4.6	3.6	1.0	0.2	Iris-setosa
148	5.5	2.3	4.0	1.3	Iris-versicolor
149	6.9	3.1	5.1	2.3	Iris-virginica

150 rows x 5 columns

בחירת מאפיינים (X) ומטרה (y)

```
#Chose attributes  
X=df.drop(['Species'],axis=1).to_numpy()  
X.shape
```

(150, 4)

```
#chose target  
y = df['Species'].to_numpy()  
y.shape
```

(150,)

יצירת אובייקט KNN טעינת הנתונים לאובייקט

```
# create KNN classifier  
knn1 = KNeighborsClassifier(n_neighbors = 3)  
# fit to data  
knn1.fit(X,y)
```

```
▼ KNeighborsClassifier  
KNeighborsClassifier(n_neighbors=3)
```

פרדיקציות

```
#predict on 3 samples
knn1.predict([[1,1,1,1],[4,4,4,4],[5,2,3,1]])

array(['Iris-setosa', 'Iris-virginica', 'Iris-versicolor'], dtype=object)
```

```
#predict on all samples
y_pred = knn1.predict(X)
y_pred.shape
```

```
(150,)
```

```
#show 5 predictions
y_pred[:5]
```

```
array(['Iris-versicolor', 'Iris-setosa', 'Iris-setosa', 'Iris-virginica',
       'Iris-versicolor'], dtype=object)
```

מדד accuracy

יחס הדוגמאות שסווגו נכון לסך הדגימות

```
#accuracy using classifier,X,y  
knn1.score(X=X,y=y)
```

0.96

```
#accuracy using y and y_pred  
accuracy_score(y_true=y, y_pred = y_pred)
```

0.96

מבצע פרדיקציה ולאחר מחשב את המדד

מקבל פרדיקציה ומחשב את המדד

KNN exercise 1

[KNN iris exercise 1](#)



הטכניון
מכון טכנולוגי
לישראל

תודה על ההשתתפות