



IRIS DATASET with KNN

패기2조-VariouseOne

박주언	손상원	이동현	이항주	이창대	임단비
-----	-----	-----	-----	-----	-----

What is Iris Dataset?



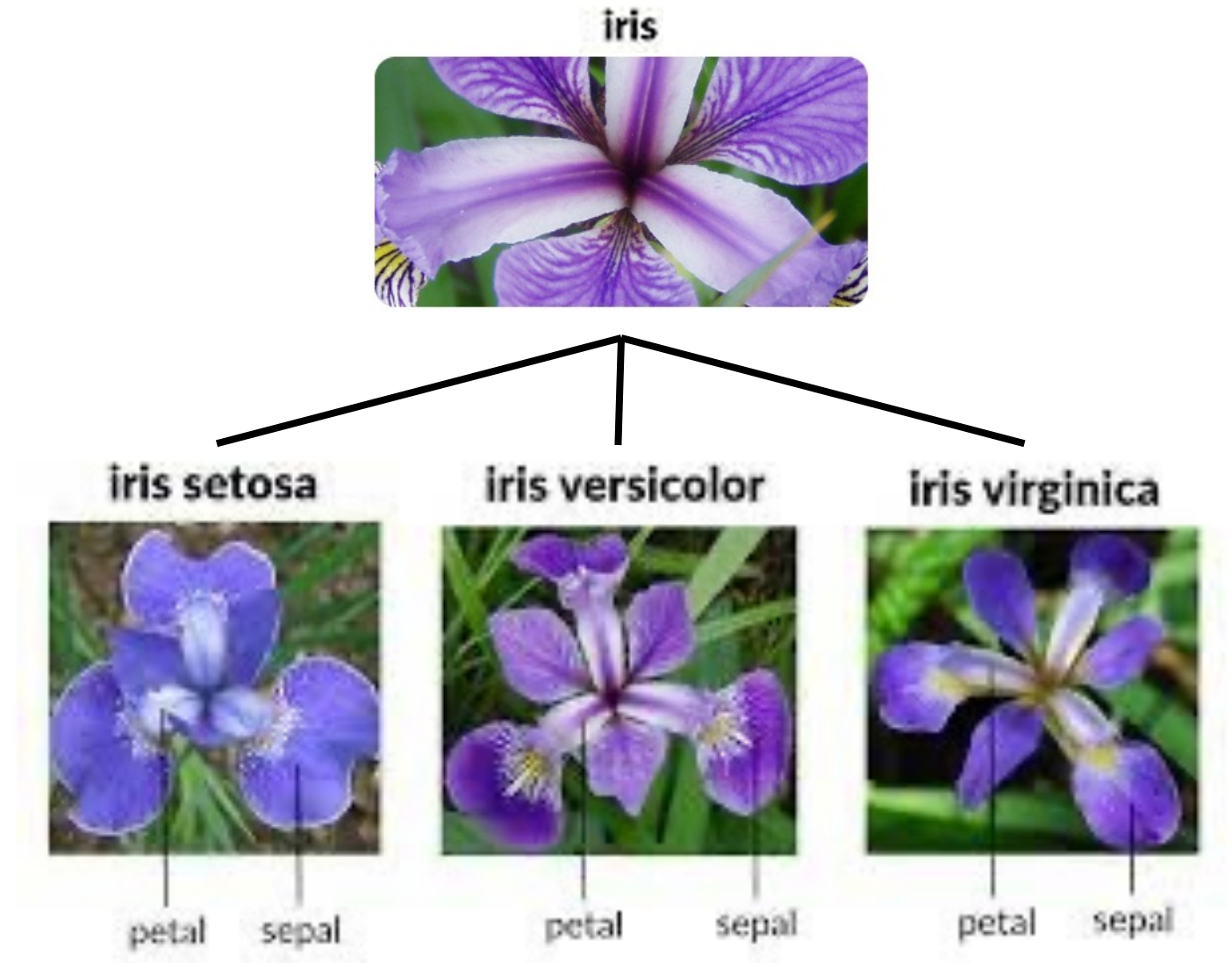
kaggle

Iris Species

Classify iris plants into three species in this classic dataset

The columns in this dataset are:

- Id
- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species



What should we do?



- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm

Prediction



with KNN

- Species

Dataset of IRIS



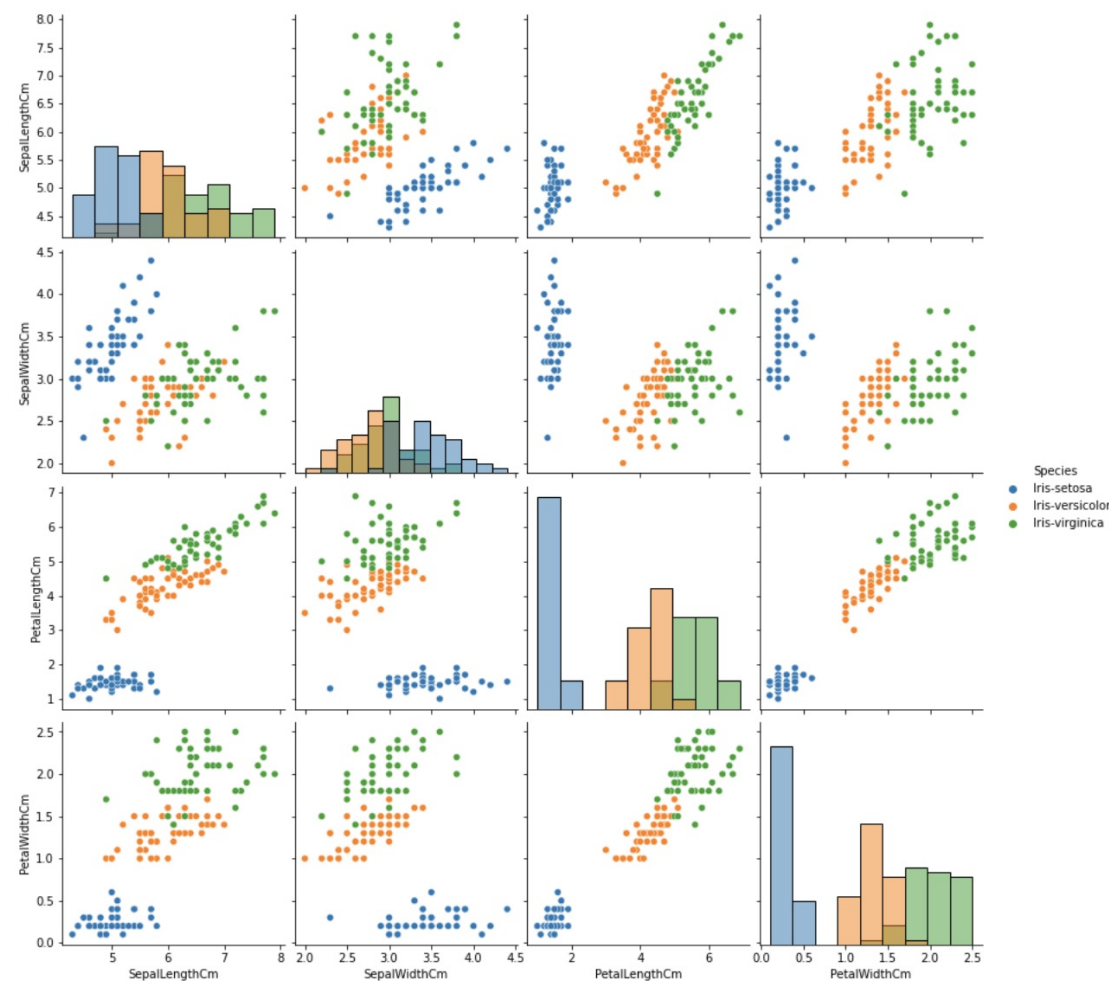
```
In [13]: 1 iris = pd.read_csv('./Iris.csv')
          2 iris
```

Out[13]:

	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

```
In [18]: 1 import seaborn as sns
          2 sns.pairplot(iris, hue="Species", size=3, diag_kind='hist')
```



Algorithm of KNN



KNNClassifier

```
def my_dist(train_x, test_x):  
    test = train_x - test_x  
    test = np.power(test, 2)  
    #print(test)  
    distance = []  
    for i in range(len(train_x)):  
        temp_error = np.sqrt(np.sum(test.iloc[i]))  
        distance.append([temp_error, df.iloc[i]['Species']])  
    return distance
```

```
[ ] K = 5
```

```
[ ] distance_record = my_dist(x_train, x_test)  
    for i in range(10):  
        print(distance_record[i])  
    distance_record.sort()  
  
    result = np.array(distance_record)  
    result = result[:K, 1]  
    print(result)  
    st.mode(result)
```

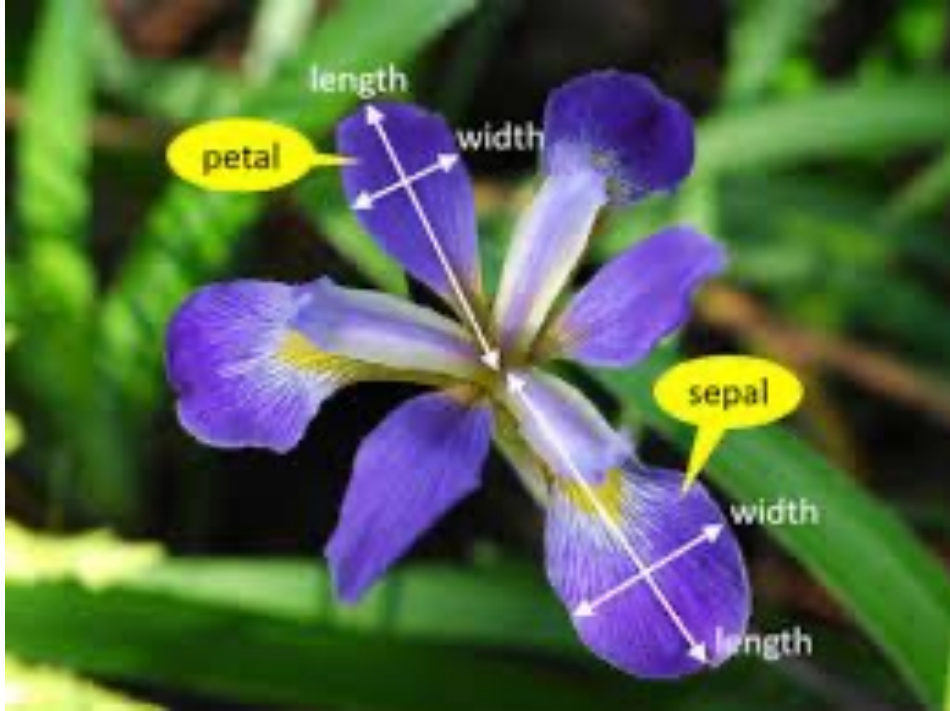
설계고안

- 1) 알고자 하는 값과
모든 데이터 간의 거리 구하기
- 2) 가장 짧은 거리 구해서 sorting
- 3) K 값 만큼 자르고 그 중 최빈값 도출

Result



[TEST]



SepalLengthCm : 5.1
SepalWidthCm : 3.5
PetalLengthCm : 5.1
PetalWidthCm : 0.2

[Result] – 최소거리 중 K=5 값 내 최빈값 구하기

[0.0, 'Iris-setosa']
[0.09999999999999998, 'Iris-setosa']
[0.1414213562373093, 'Iris-setosa']
[0.14142135623730964, 'Iris-setosa']
[0.14142135623730995, 'Iris-setosa']

[0.14142135623730995, 'Iris-setosa']
[0.17320508075688743, 'Iris-setosa']
[0.17320508075688762, 'Iris-setosa']
[0.22360679774997896, 'Iris-setosa']
[0.300000000000000016, 'Iris-setosa']

['Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa' 'Iris-setosa']

Iris-setosa ← **Our Prediction**