File Edit View Run Kernel Tabs Settings Help

03_Ins_K-Means.ipynb

Code Python 3

Instructor Do: The K-Means Algorithm

[1]:
```
# Initial imports
import pandas as pd
import hvplot.pandas
from pathlib import Path
from sklearn.cluster import KMeans
```

[2]:
```
# Loading data
file_path = Path("data/new_iris_data.csv")
df_iris = pd.read_csv(file_path)
df_iris.head(10)
```

[2]:

	sepal_length	sepal_width	petal_length	petal_width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
5	5.4	3.9	1.7	0.4
6	4.6	3.4	1.4	0.3
7	5.0	3.4	1.5	0.2
8	4.4	2.9	1.4	0.2
9	4.9	3.1	1.5	0.1

Using K-Means

[3]:
```
# Initializing model with K = 3 (since we already know there are three classes of iris plants)
model = KMeans(n_clusters=3, random_state=5)
```

[4]:
```
# Fitting model
model.fit(df_iris)
```

[4]:
```
KMeans(n_clusters=3, random_state=5)
```

[5]:
```
# Get predictions
predictions = model.predict(df_iris)
print(predictions)
```

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 2 2 2 2 0 2 2 2 2
 2 2 0 0 2 2 2 2 0 2 0 2 0 2 2 0 0 2 2 2 2 2 0 2 2 2 2 0 2 2 2 0 2 2 2 0 2
 2 0]
```

[6]:
```
# Add a new class column to df_iris
df_iris["class"] = model.labels_
df_iris.head()
```

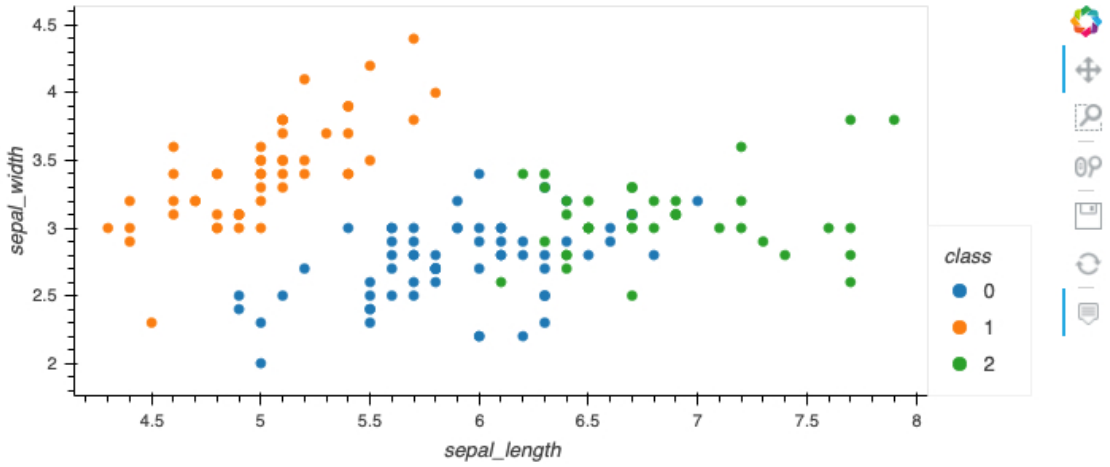
[6]:

	sepal_length	sepal_width	petal_length	petal_width	class
0	5.1	3.5	1.4	0.2	1
1	4.9	3.0	1.4	0.2	1
2	4.7	3.2	1.3	0.2	1

3	4.6	3.1	1.5	0.2	1
4	5.0	3.6	1.4	0.2	1

[7]:
```
# Plotting the clusters with two features
df_iris.hvplot.scatter(
    x="sepal_length",
    y="sepal_width",
    by="class")
```

[7]:



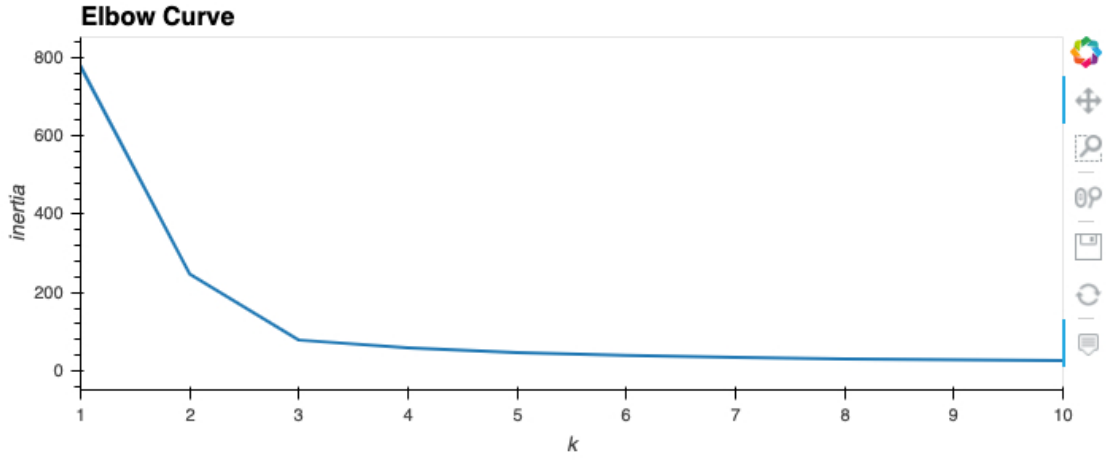
Finding the best value for *k* using the Elbow Curve

[8]:
```
inertia = []
k = list(range(1, 11))

# Looking for the best k
for i in k:
    km = KMeans(n_clusters=i, random_state=0)
    km.fit(df_iris)
    inertia.append(km.inertia_)

# Define a DataFrame to plot the Elbow Curve using hvPlot
elbow_data = {"k": k, "inertia": inertia}
df_elbow = pd.DataFrame(elbow_data)
df_elbow.hvplot.line(x="k", y="inertia", title="Elbow Curve", xticks=k)
```

[8]:



[ ]:


0
1
Python 3 | Idle
Mode: Command
Ln 1, Col 1
03_Ins_K-Means.ipynb