# Viikko 44 -tehtävät

## Tehtävä 1

import matplotlib.pyplot as plt

import pandas as pd

*# Importing the dataset*

df = pd.read\_csv('./work/viikko9/datasets/customers.csv')

X = df.iloc[:, [3, 4]].values

*# Using the elbow method to find the optimal number of clusters*

from sklearn.cluster import KMeans

wcss = []

for i in range(1, 11):

    model = KMeans(n\_clusters = i, init = 'k-means++', random\_state = 0)

    model.fit(X)

    wcss.append(model.inertia\_)

plt.plot(range(1, 11), wcss)

plt.title('The Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('WCSS')

plt.show()

Kuva, joka sisältää kohteen teksti, kuvakaappaus, viiva, diagrammi

Kuvaus luotu automaattisesti

*# Training the K-Means model on the dataset*

model = KMeans(n\_clusters = 5, init = 'k-means++', random\_state = 0)

y\_kmeans = model.fit\_predict(X)

*# Visualising the clusters*

plt.scatter(X[y\_kmeans == 0, 0], X[y\_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster 1')

plt.scatter(X[y\_kmeans == 1, 0], X[y\_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')

plt.scatter(X[y\_kmeans == 2, 0], X[y\_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster 3')

plt.scatter(X[y\_kmeans == 3, 0], X[y\_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')

plt.scatter(X[y\_kmeans == 4, 0], X[y\_kmeans == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')

plt.scatter(model.cluster\_centers\_[:, 0], model.cluster\_centers\_[:, 1], s = 300, c = 'yellow', label = 'Centroids')

plt.title('Clusters of customers')

plt.xlabel('Annual Income (k$)')

plt.ylabel('Spending Score (1-100)')

plt.legend()

plt.show()

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Värikkyys, diagrammi

Kuvaus luotu automaattisesti

## Tehtävä 2

import matplotlib.pyplot as plt

import pandas as pd

import seaborn as sns

*# Importing the dataset*

df = pd.read\_csv('./work/viikko9/datasets/iris.csv')

X = df[['petal\_length', 'petal\_width']]

*# Using the elbow method to find the optimal number of clusters*

from sklearn.cluster import KMeans

wcss = []

for i in range(1, 11):

    model = KMeans(n\_clusters = i, init = 'k-means++', random\_state = 0)

    model.fit(X)

    wcss.append(model.inertia\_)

plt.plot(range(1, 11), wcss)

plt.title('The Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('WCSS')

plt.show()

Kuva, joka sisältää kohteen teksti, diagrammi, viiva, Tontti

Kuvaus luotu automaattisesti

iris\_map = {0: 'versicolor', 1: 'setosa', 2: 'virginica'}

df\_pred = pd.DataFrame(y\_kmeans)

df\_pred.columns = ['predict']

df\_pred['predict'] = df\_pred['predict'].map(iris\_map)

df\_pred['real'] = df['species']

pd.crosstab(df\_pred['predict'], df\_pred['real'])

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Fontti, numero

Kuvaus luotu automaattisesti

*# Visualising the clusters*

plt.figure(figsize=(10, 7))

sns.scatterplot(x = 'petal\_length', y = 'petal\_width', hue=df\_pred['predict'], data=df, palette=['red', 'green', 'blue'])

sns.scatterplot(x = model.cluster\_centers\_[:, 0], y = model.cluster\_centers\_[:, 1], color = 'black', marker='x', s = 150, label = 'Centroids')

plt.title('Clusters of Iris Flowers')

plt.xlabel('Petal length')

plt.ylabel('Petal width')

plt.legend()

plt.show()

Kuva, joka sisältää kohteen teksti, kuvakaappaus, näyttö, diagrammi

Kuvaus luotu automaattisesti