**교육일지**

**교육 제목 : 머신러닝**

**교육 장소 : YGL C6 강의실**

**교육 일시 : 2021/10/15**

**import pandas as pd # 매트릭스 연산**

**import numpy as np # 엑셀**

**import matplotlib.pyplot as plt # plot**

**## 도미**

**bream\_length = [25.4, 26.3, 26.5, 29.0, 29.0, 29.7, 29.7, 30.0, 30.0, 30.7, 31.0, 31.0, 31.5, 32.0, 32.0, 32.0, 33.0, 33.0, 33.5, 33.5, 34.0, 34.0, 34.5, 35.0, 35.0, 35.0, 35.0, 36.0, 36.0, 37.0, 38.5, 38.5, 39.5, 41.0, 41.0]**

**bream\_weight = [242.0, 290.0, 340.0, 363.0, 430.0, 450.0, 500.0, 390.0, 450.0, 500.0, 475.0, 500.0, 500.0, 340.0, 600.0, 600.0, 700.0, 700.0, 610.0, 650.0, 575.0, 685.0, 620.0, 680.0, 700.0, 725.0, 720.0, 714.0, 850.0, 1000.0, 920.0, 955.0, 925.0, 975.0, 950.0]**

**print(type(bream\_length))**

**print(np.shape(bream\_length))**

**plt.scatter(bream\_length, bream\_weight, s=100)**

**plt.title("Bream, Length vs Weight", pad = 20)**

**plt.xlabel("Lenght (cm)")**

**plt.ylabel("weight (g)")**

**빙어**

**smelt\_length = [9.8, 10.5, 10.6, 11.0, 11.2, 11.3, 11.8, 11.8, 12.0, 12.2, 12.4, 13.0, 14.3, 15.0]**

**smelt\_weight = [6.7, 7.5, 7.0, 9.7, 9.8, 8.7, 10.0, 9.9, 9.8, 12.2, 13.4, 12.2, 19.7, 19.9]**

**print(np.shape(smelt\_length))**

**plt.scatter(bream\_length, bream\_weight)**

**plt.scatter(smelt\_length, smelt\_weight)**

**plt.title("Bream, Smelt")**

**plt.xlabel("Lenght (cm)")**

**plt.ylabel("weight (g)")**

**plt.legend()**

**## knn을 이용한 분류**

**length = bream\_length + smelt\_length**

**weight = bream\_weight + smelt\_weight**

**fish\_data = [[l, w] for l, w in zip(length, weight)]**

**print(type(fish\_data))**

**print(np.shape(fish\_data))**

**fish\_data[:5]**

**fish\_target = [1]\*35 + [0]\*14**

**print(fish\_target)**

**## sklearn knn module import**

**from sklearn.neighbors import KNeighborsClassifier**

**##객체 생성**

**kn = KNeighborsClassifier(n\_neighbors =5)**

**## knn 모델 fitting**

**kn.fit(fish\_data, fish\_target)**

**## knn model metrics**

**kn.score(fish\_data, fish\_target)**

**## predction**

**kn.predict([[30, 600]])**

**plt.scatter(bream\_length, bream\_weight, s = 100, label = "bream =1")**

**plt.scatter(smelt\_length, smelt\_weight, s = 100, label = "bream =0")**

**plt.scatter(30, 600, s =100, label = "new")**

**plt.title("Bream vs Smelt")**

**plt.xlabel("Lenght (cm)")**

**plt.ylabel("weight (g)")**

**plt.legend()**

**## 샘플링 편향**

**from sklearn.neighbors import KNeighborsClassifier**

**kn = KNeighborsClassifier()**

**from sklearn.model\_selection import train\_test\_split**

**#train\_test\_split(**

**# \*arrays,**

**# test\_size=None,**

**# train\_size=None,**

**# random\_state=None,**

**# shuffle=True,**

**# stratify=None,**

**#)**

**train\_input, test\_input, train\_target, test\_target = train\_test\_split(**

**fish\_data, fish\_target, test\_size =0.25,**

**stratify = fish\_target, random\_state = 42)**

**#Feature scaling**

**mean =np.mean(train\_input, axis = 0)**

**std = np.std(train\_input, axis =0)**

**print('mean = ', mean.round(2))**

**print('std = ', std.round(2))**

**print(train\_input[:5])**

**#표준화**

**train\_scaled = (train\_input -mean)/ std**

**train\_scaled[:5]**