**기계학습 1차 과제**

학번: 18011780

이름: 김민준

**코드**

import numpy as np #numpy를 불러오고 np로 축약한다

import pandas as pd #pandas를 불러오고 pd로 축약한다

import seaborn as sns #seaborn을 불러오고 sns로 축약한다

iris = sns.load\_dataset('iris') #iris라는 변수명으로 iris 데이터를 가져온다

print(iris.head()) #최초의 5개의 값을 출력한다

print(iris.shape) #iris data의 행과 열의 수를 출력한다

X=iris.drop('species', axis =1) #'species'열을 없애고 x값을 정의한다

print(X.shape) #x의 행과 열을 출력한다

y=iris['species'] # 'species'열을 y로 정의한다

from sklearn.preprocessing import LabelEncoder #LabelEncoder를 불러온다

classle=LabelEncoder()

y=classle.fit\_transform(iris['species'].values) #species 열의 문자열을 categorical 값으로 바꿔준다

print('species labels:', np.unique(y)) #중복되는 y값을 하나로 정리하여 출력한다

yo=classle.inverse\_transform(y) #원래의 species문자열로 바꿔준다

print('species:',np.unique(yo)) #중복되는 y값을 하나로 정리하여 출력한다

from sklearn.model\_selection import train\_test\_split #train\_test\_split을 불러온다

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=22) #x와 y의 test크기를 20%로 지정해주고 random state를 22로 지정해준다

print(X\_train.shape) #x train의 행과 열의 수를 출력해준다

print(X\_test.shape)#x test의 행과 열의 수를 출력해준다

print(y\_train.shape)#y train의 행과 열의 수를 출력해준다

print(y\_test.shape)#x test의 행과 열의 수를 출력해준다

from sklearn.neighbors import KNeighborsClassifier #KNN을 불러온다

knn=KNeighborsClassifier(n\_neighbors=3) #3개의 인접한 값으로 정의해준다

knn.fit(X\_train,y\_train) #모델을 fitting해준다

y\_train\_pred=knn.predict(X\_train) #y값을 예측치를 정의해준다

y\_test\_pred=knn.predict(X\_test) #모델을 적용한 y값을 정의해준다

print('Misclassified training samples: %d' %(y\_train!=y\_train\_pred).sum()) #오분류 데이터의 갯수를 출력해준다

print('Misclassified test samples: %d' %(y\_test!=y\_test\_pred).sum()) #오분류 데이터 갯수를 출력해준다

from sklearn.metrics import accuracy\_score #정확도 계산을 위한 accuracy\_score를 불러온다

print(accuracy\_score(y\_test, y\_test\_pred)) #정확도를 출력해준다

**실행 결과**



