

14

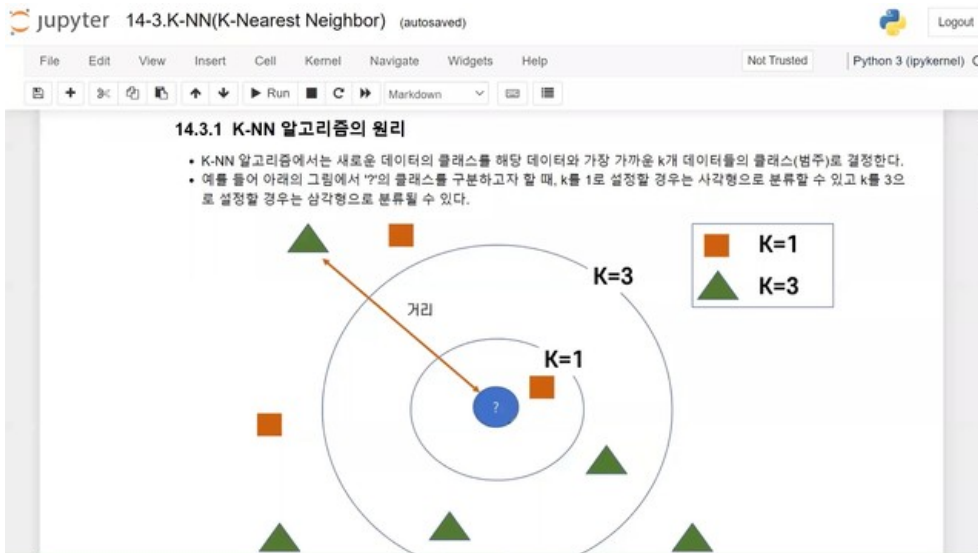
KNN 분류기 개념 & 실습

[KNN 분류 개념]

KNN을 통해 새로운 데이터가 어느 범주에 속할 것인지를 정할 수 있습니다. 이 때, K의 값을 정해주는 것이 중요합니다.

* 이 개념은 결측치를 대체할 때에도 사용됩니다.

00:00



[KNN 분류 실습]

sklearn안에 있는 KNNClassifier로 학습해보겠습니다.

01:50

14.3.2 K의 선택

- K의 선택은 학습의 난이도와 데이터의 개수에 따라 결정될 수 있으며, 일반적으로는 훈련 데이터 개수의 제곱근으로 설정한다. 그리고 짝수로 했을 때, 인접격차의 범주가 동등할 경우가 나오므로 일반적으로 홀수로 설정함

```

In [1]: import pandas as pd
        from sklearn.model_selection import train_test_split

        credit = pd.read_csv("../data/credit_final.csv")
        X = credit[credit.columns.difference(['credit.rating'])]
        y = credit['credit.rating']

        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=321)
        X_train.head()
    
```

Out[1]:

	account.balance	age	apartment.type	bank.credits	credit.amount	credit.duration.months	credit.purpose	current.assets
133	3	26	2	1	1453	18	3	1
464	2	24	1	1	7472	12	4	1
919	1	46	3	2	6331	48	1	4

[최적의 K값을 찾는 방법]

K의 갯수를 바꿔가며, ROC_AUC값이 가장 높은 k의 값을 반복문으로 찾아주면 됩니다.

03:52

