**Classification 머신러닝 참고자료**

머신러닝자료: <https://github.com/Youngpyoryu/SD_academy/blob/732f017936ae6e7c8295ffa7dbbcdf4fe21302c1/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D/%EB%B6%84%EB%A5%98(Classification)/Classification_210413.ipynb>

2021/04/13

KNN

특정구간에서는 1, 특정구간에서는 0 이런식으로 특정 구간 별로 값이 나뉨.

통계학파 크게 두 가지로 나뉨. 1. 빈도주의(실제로 일어난 일) 2. Bayesian

P(A|B) = P(B|A)P(A)/P(B)

여기서 P(B|A)는 사전확률

P(A)/P(B)는 Likelihood

P(A|B)는 사후 확률을 의미한다.

Ex) P(비|맑은날) -> 맑은 날 중에서 비가 올 확률

Entropy 가 높다 라는 것은 화합물(혼잡도)이 많다는 것.

Entropy 가 낮다 라는 것은 화합물(혼잡도)이 적다는 것

-log 함수는 부호가 (-)이므로 정보량이 증가할수록 확률이 감소함.

log함수가 x축을 기준으로 대칭 되는게 -log함수

지니계수가 낮으면 데이터가 잘 분리 되었다.

지니계수가 높으면 데이터가 잘 분리 되지 않았다라는 의미

DecisionTree Classifier.html

<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>