Decision Tree (Gini,Entropy)

# 라이브러리 링크

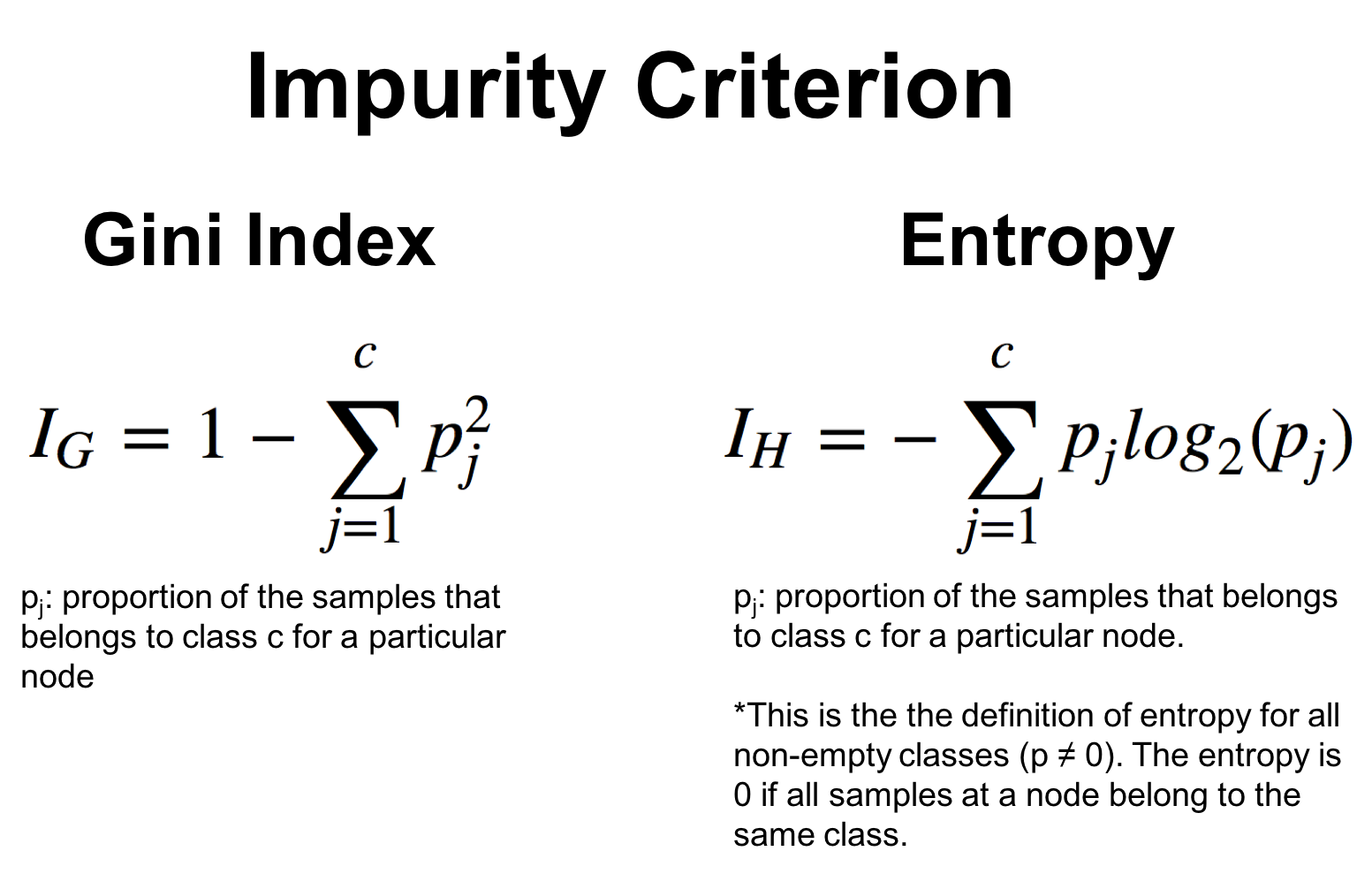
**install :**

<https://scikit-learn.org/stable/install.html>

**github :**

<https://github.com/scikit-learn/scikit-learn/blob/82df48934eba1df9a1ed3be98aaace8eada59e6e/sklearn/tree/_classes.py>

# 기초 설명



Decision Tree 학습법은 어떤 항목에 대한 관측값과 목표값을 연결시켜주는 예측 모델이며.

이에 대해 결정 트리를 사용하며 이는 시각적이고 명시적인 방법으로 의사 결정 과정과 결정된 의사를 보여주는데 사용된다.

# 버전 정보

* NumPy >= 1.14.6 (pip install numpy)
* Scipy >= 1.1.0 (pip install scipy)
* Joblib >= 0.11 (pip install joblib
* Threadpoolctl >= 2.0.0 (pip install threadpoolctl)
* pandas >= 1.2.4 (pip install pandas)
* matplotlib == 3.22 (pip install matplotlib)
* sklearn == 1.0.2 (pip install sklearn)
* graphviz == 0.10.1 (pip install graphviz)

# 데이터셋 설명 및 출처

* diabetes.scv: This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.
* Sources : https://www.kaggle.com/uciml/pima-indians-diabetes-databas

# 코드 설명

* 당뇨병에 대한 환자의 케이스 데이터 diabetes.scv 를 데이터 셋으로 받아 이를 Decision Tree 모델에 학습시킨후 이에대한 모델의 검증 정확도를 나타내고 만들어진 Decision Tree 를 pdf 형식으로 저장 및 출력한다.

# 검증 방법

* 코드 내부에서 데이터셋을 학습용 데이터셋과 검증용 데이터셋으로 나누어 이를 검증함.

(test\_size = 0.2, random\_state = 0)

* 추가적으로 sklearn의 accuracy\_score 함수를 사용하여 실제값과 예측값 사이의 적합도를 평가.