사이킷 런을 활용한 머신러닝

1. 정리

피쳐 = 독립변수(입력값)

레이블을 0, 1, 2,로 둘 것

독립변수: 꽃

Classification vs Regression

Classification: 특정 기준에 따라서 데이터를 분류해서 레이블 값을 지정한다.

Regression: 특정 데이터에 대한 레이블 값을 추정할 수 있는 기준을 구한다.

-학습데이터와 테스트 데이터의 차이

학습데이터: 알고리즘의 학습을 위한 대상이 되는 데이터

테스트 데이터: 알고리즘의 정확도를 파악하기 위한 목적의 데이터

k교차 검증

실제 다양한 알고리즘에 대해서 적절성을 파악하기 위한 방법으로 사용되었으나

기존 데이터 세트의 양을 늘려주는 효과를 가져오기 때문에

데이터의 양이 적을 때, 사용하기도 한다.

Kfold 예제 숙제

2. 궁금증 정리사항