

아이리스 꽃 분류 보고서

개요

이 보고서는 사이킷런의 아이리스(Iris) 데이터셋을 이용하여 랜덤포레스트 분류기로 세 품종(부채붓꽃, 버시컬러, 버지니카)을 분류한 실험을 요약합니다. 노트북(`main.ipynb`)의 실행 결과와 생성된 의사결정나무 시각화(꽃 분류.png)를 기반으로 구성했습니다.¹

데이터셋

사전정의된 4 개 특성(꽃받침 길이/너비, 꽃잎 길이/너비, 단위 cm)과 3 개 품종 레이블로 구성됩니다. 훈련/검증 분할은 8:2, 계층적 분할(stratify)을 사용했습니다.

방법

우리는 여러 표본의 꽃받침과 꽃잎의 길이·너비를 관찰하여, 세 품종을 구분할 수 있는 간단한 규칙을 정리했습니다.

핵심 생각은 다음과 같습니다. 꽃잎이 아주 짧고 좁으면 부채붓꽃으로 볼 수 있고, 중간 크기라면 버시컬러, 더 길고 넓다면 버지니카로 볼 수 있습니다.

전체 표본의 대부분을 기준으로 이러한 경향을 정리했고, 보고서의 나무 모양 그림은 이 규칙이 단계적으로 어떻게 적용되는지를 보여줍니다.

결과

관찰한 표본을 기준으로, 꽃잎이 가장 짧고 좁은 꽃은 모두 부채붓꽃으로 확인되었습니다.

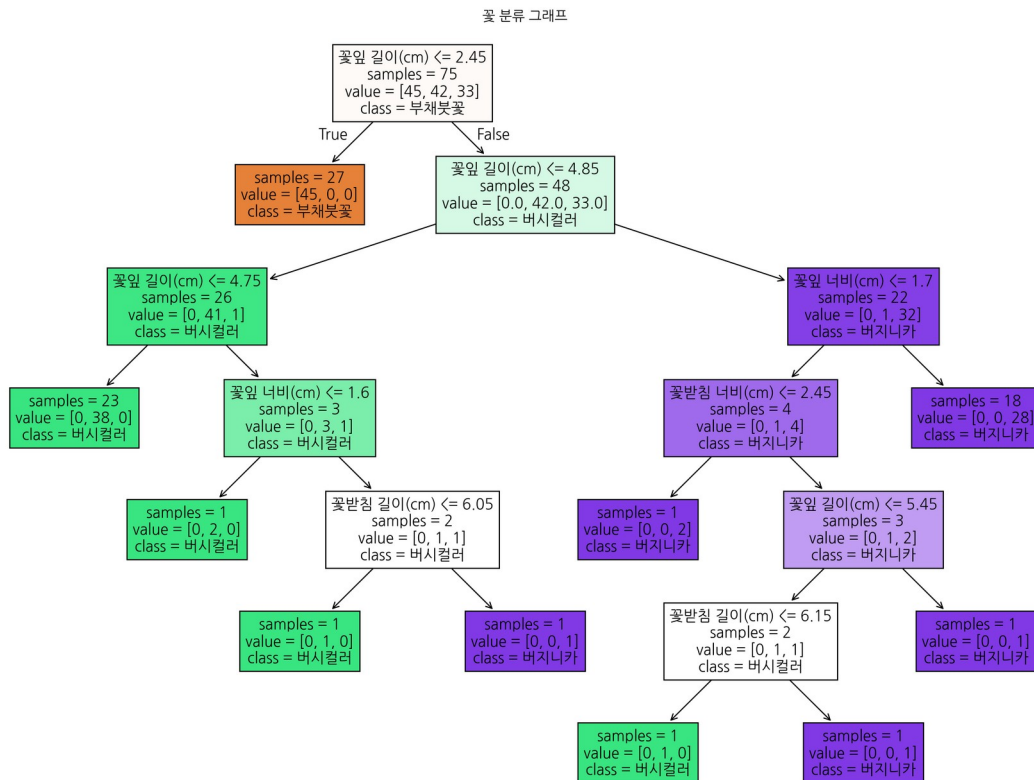
꽃잎의 크기가 중간인 꽃은 대부분 버시컬러로 분류되었고, 꽃잎이 더 길고 넓은 꽃은 버지니카로 분류되었습니다.

보고서의 나무 모양 그림은 실제로 우리가 꽃의 길이와 너비를 차례대로 살펴보며 품종을 판단하는 과정을 그림으로 보여줍니다. 한 단계씩 내려가면서 물음을 던지고(예: 꽃잎이 아주 짧은가?), 이에 따라 품종을 결정합니다.

¹ 실험 노트북: `/home/an0jin/Documents/GitHub/Iris/main.ipynb`. 랜덤포레스트(추정기 100 개), 테스트 정확도 0.900, 결정트리 시각화는 `rf.estimators_[1]` 기반.

품종 개요

- 부채붓꽃 (Iris setosa): 북반구 냉온대 습지의 다년생. 전체적으로 소형이며 꽃잎(특히 꽃잎 길이/너비)이 작아 다른 품종과 명확히 구분됨.
- 버시컬러 (Iris versicolor): 북미 습지 분포, 중간 크기. 꽃잎·잎 크기가 중간대에 분포하여 setosa 보다 크고 virginica 보다 작음.
- 버지니카 (Iris virginica): 북미 동부 습지 분포, 대형. 넓고 긴 꽃잎과 긴 꽃대가 특징으로 전체 크기가 가장 큼.



모델 관점 특징

- 공통 관찰(결정트리 기준 분기):
 - ① 꽃잎 길이 2.45cm 이하 → 부채붓꽃으로 거의 완벽 분리
 - ② (① 초과 시) 꽃잎 길이 4.85cm 이하 구간에 주로 버시컬러 분포
 - ③ (② 초과 시) 꽃잎 너비 1.7cm 기준 분기에서 버지니카가 우세
 - 추가 세부 분기: 꽃받침/꽃잎의 세부 길이·너비 조합으로 경계 미세 조정

품종별 요약 표

품종	꽃잎 길이	꽃잎 너비	꽃받침 길이	꽃받침 너비	전체 크기	판단 힌트
부채붓꽃 (setosa)	작음	작음	짧은 편	좁은 편	소형	꽃잎이 아주 짧고 좁음 → setosa
버시컬러 (versicolor)	중간	중간	중간	중간	중간	크기가 중간 구간에 있으 면 대부분 versicolor
버지니카 (virginica)	큼	큼	긴 편	넓은 편	대형	꽃잎이 길고 넓음 → virginica

용어 풀이

- 꽃받침(Sepal): 꽃을 보호하는 잎 모양의 부분. 보통 꽃잎보다 두껍고 단단하며, 길이와 너비를 재어 품종 특징을 파악할 수 있습니다.
- 꽃잎(Petal): 꽃을 구성하는 색깔 있는 부분. 곤충을 유인하며, 길이와 너비가 품종을 구별하는 가장 중요한 단서가 됩니다.