

## 개념 공부 + 필사 과제 보고서

이름	황지현
일시	2021-11-22
개념 공부	<p>&lt;결정 트리&gt;</p> <ol style="list-style-type: none"> <li>1. 결정 트리 학습과 시각화 decisionTreeClassifier 를 이용해 붓꽃 데이터를 분류.</li> <li>2. 예측하기 예측방법: 새로운 붓꽃의 품종을 분류하려 할때, 루트노드부터 시작. 리프노드에 도달하면 그 노드의 각 클래스의 확률이 가장 높은 클래스를 새로운 붓꽃의 품종으로 결정. 노드의 gini 속성은 불순도를 측정. 한 노드의 모든 샘플이 같은 클래스에 속한다면 <math>gini=0</math>, 순수하다고 해석.</li> <li>3. 클래스 확률 추정 새로운 샘플이 리프노드에 도착하면, 그 노드에 있는 클래스 K의 훈련 샘플의 비율을 반환.</li> <li>4. CART 훈련 알고리즘 CART 비용함수를 최소화 시키는 방향으로 훈련세트를 둘로 나눔. 나뉜 두 훈련세트에 대해서 또 이과정을 반복함. 최대 깊이가 되면 중지하거나 아무리 나뉘도 불순도가 줄어들지 않을 때 분할을 멈추게됨.</li> <li>5. 계산 복잡도 대용량 데이터를 분석할 때 다른 알고리즘에 비해 훈련속도가 매우 빠름.</li> <li>6. 지니 불순도 또는 엔트로피 지니 불순도 대신에 엔트로피를 사용하기도 함. 큰 차이는 없음.</li> <li>7. 규제 매개변수 결정 트리는 비파라미터 모델. 모델 구조가 데이터에 맞춰진다는 의미. 하지만 과대적합될 가능성이 크기 때문에 매개변수를 사용함.</li> <li>8. 회귀 결정트리는 회귀문제에도 사용할 수 있음. 주요한 차이는 각 노드에서 클래스를 예측하는 대신 어떤 값을 예측한다는 것. 새로운 샘플이 리프노드에 도달하면 그 리프노드에 있는 훈련 샘플의 평균 타깃값이 예측값이 됨. 회귀에서 CART 알고리즘은 불순도 대신 mse 를 최소화하도록 분할. 이 또한 과대적합되기 쉽기 때문에 규제가 필요.</li> <li>9. 불안정성</li> </ol>

	<p>결정트리는 데이터의 작은 변화에도 매우 민감하기 때문에, PCA 기법과 같이 사용하기도 함.</p>
깃허브 주소	<p><a href="https://github.com/vldzmtnsdl/ESAA-/blob/ESAA/%ED%95%B8%EC%A6%88%EC%98%A8%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D%20_6%EA%B2%B0%EC%A0%95%ED%8A%B8%EB%A6%AC.ipynb">https://github.com/vldzmtnsdl/ESAA-/blob/ESAA/%ED%95%B8%EC%A6%88%EC%98%A8%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D%20_6%EA%B2%B0%EC%A0%95%ED%8A%B8%EB%A6%AC.ipynb</a></p>