7/4 강의노트

텍스트, 도표, 스크린샷이(가) 표시된 사진

자동 생성된 설명

부스팅 : 여러 개의 분류기가 순차적으로 학습을 진행하여 이전 단계에서 예측이 틀린 데이터에 대해서 올바르게 예측하도록 다음 분류기에 가중치를 조절함. 오래걸리고, 과적합이 발생할 가능성이 있음

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

랜덤포레스트 모델

기본 설정은 100개의 결정트리를 사용

n\_jobs : 동시에 몇 개의 학습을 진행하는가, -1이면 컴퓨터 자원이 허용하는 최대한으로 끌어옴 (쓰레드 개수 기반)

텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명

Cross\_validate : 교차검증, 데이터를 5등분하여 각 조각별로 8:2로 데이터를 나눠 학습을 5번 수행하여 교차검증 수행

텍스트, 영수증, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

랜덤포레스트는 의사결정나무의 앙상블이므로 의사결정나무에서 제공하는 매개변수는 대부분 쓸수 있음

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

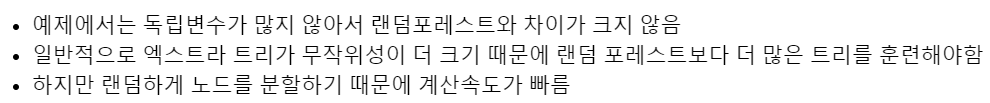
Oob\_score (out of bag) : 부트스트랩 샘플에 뽑히지 않아 훈련에 사용되지 않는 데이터로 검증 데이터처럼 훈련에 사용될 수 있음, 스플릿을 못할 정도로 데이터 개수가 너무 적을 경우 사용할 수 있음

텍스트, 폰트, 화이트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

엑스트라 트리 모델 : 랜덤포레스트와 유사하지만, 샘플 추출에서 부트스트랩(고의성 편중 데이터 추출)을 사용하지 않고 전체 훈련 세트를 사용하며, 각각의 의사결정나무에서 노드 분할 시 무작위 노드로 분할을 하게 됨

엑스트라 트리 모델이 성능은 랜덤 포레스트보다는 아주 약간 떨어지지만, 무작위성 때문에 계산 속도가 빠름



텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

그레디언트 부스팅 : 의사결정나무 모델을 연속으로 돌려서 이전 모델의 오차를 보완하는 방식으로 머신러닝을 진행, 직렬연산이므로 모델은 n\_jobs를 쓰지 않음, 과대적합을 방지하기 위해서 각각의 의사결정나무는 얕음

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

학습율(learning\_rate) : 다음 모델이 오차를 예측할 때 어느만큼 다음 모델에 반영을 할 것인가 (해당 보정도 정확하지 않으므로 보정을 보수적으로 함, 기본값은 0.1)

N\_estimators : 연결하는 의사결정나무 모델 개수

군집알고리즘

비지도 학습법, 데이터에 대핸 정답을 주지 않고(ex. 꽃의 종류) 주어진 데이터를 일정한 특성에 따라서 스스로 분류함 (크기, 색 등으로)

텍스트, 스크린샷, 과일, 사과이(가) 표시된 사진

자동 생성된 설명

Matplotlib의 imshow를 이용하여 이미지 출력

텍스트, 과일, 스크린샷, 사과이(가) 표시된 사진

자동 생성된 설명

색상은 cmap의 gray를 gray\_r로 변경하면 반전 가능

픽셀값을 기반으로 분류 분석

텍스트, 폰트, 스크린샷, 친필이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

각 이미지의 전체 픽셀값 평균

텍스트, 도표, 스크린샷, 그래프이(가) 표시된 사진

자동 생성된 설명