

기계학습 기말고사 정리 2 (20241519 조예성)

질문에 따라 데이터를 구분하는 알고리즘: 결정트리

- 노드 구분 3가지: Root Node, Intermediate Node, Leaf Node
- 최대 깊이 제한: max_depth
- 각 분할에 사용할 속성 개수: max_features
- 분할 위한 최소 샘플 수 조건: min_sample_split
- 리프 노드가 가져야할 최소 샘플 수: min_sample_leaf
- 가중치가 부여된 전체 샘플에서의 비율: min_weight_fraction_leaf
- 리프 노드의 최대 수: max_leaf_node
- 분할로 얻어질 최소한의 불순도 감소량: min_impurity_decrease
- 객체 명: DecisionTreeRegressor

트리 분할의 기준: Criterion

- 두 종류: 지니 불순도, 엔트로피

간접 투표 위하여 변경: voting_clf.voting = "soft"

훈련세트 샘플링 방식 2가지: 배깅, 페이스팅

랜덤포레스트에서 각 특성이 얼마나 중요한지 측정한 점수: 특성 중요도

여러 모델 사용 학습: 앙상블 학습

- 합친 결과에 한 번 더 학습: 스택킹
 - 블렌더, 메타 학습기

차원을 2차원으로 줄이는 코드: PCA(n_components=2)

PCA 2 종류: 랜덤 PCA, 점진적 PCA

샘플 간 거리 보존하며 차원축소: MDS; 다차원 스케일링

각 샘플을 가장 가까운 이웃과 연결: Isomap

비슷한 샘플은 가까이, 비슷하지 않은 샘플은 멀리: t-SNE

훈련 과정에서 클래스 사이를 가장 잘 구분하는 축 학습: 선형 판별 분석

k-평균에서 중앙 기준: 센트로이드

최적 클러스터 개수 찾는 그래프 2종류: 이너셔그래프, 실루엣 점수

- 엘보: 이너셔 그래프
- 높을수록 좋은거: 실루엣 계수

능동학습 서술: 전문가가 학습 알고리즘과 상호 작용하여, 알고리즘이 요청 시, 특정 샘플의 레이블을 제공하는 학습 방법