|  |  |
| --- | --- |
| 교육제목 | 데이터 기반 인공지능 시스템 엔지니어 양성 과정 |
| 교육일시 | 211014 |
| 교육장소 | 자택(디스코드 사용 온라인 학습) |
| **교육내용** | |
| 1. 기계학습 모델 평가 - Bootstrapping   Bootstrapping 샘플은 복원추출을 이용한 데이터의 무작위 샘플임.  선택한 샘플을 기반으로 모델을 구축하고 OOB(Out-of-Bag) 샘플을 이용하여 모델을 평가함.  캡처.JPG   1. 초매개변수 조절   초매개변수는 학습과정을 제어하는 데 사용되는 매개 변수를 의미한다.  초매개변수는 모델 학습과정이 아닌 모델 개발자에 의해서 지정된다.  캡처.JPG   * K가 2일 경우 overfitting, 데이터가 추가될때마다 문제 생김. * K가 100일경우 데이터와 맞지 않게 됨.  1. K-nearest neighbors classification   지도학습으로서 분류(classification)나 회귀(regression)에 사용되는 비모수적 방법.  파라메터 학습을 위한 훈련과정이 없으나 훈련집합은 필요  각 데이터 간에 거리를 계산하기 위한 거리척도가 필요  초매개변수 k를 설정해야 함.  거리에 대한 가중치.  캡처.JPG   1. 기하학적 거리   캡처.JPG  예)  캡처.JPG   1. Log transformation   데이터에 log를 씌우면 오른쪽으로 치우친 분포(Right skewed)가 정규 분포로 변환됨.  캡처.JPG   1. Feature 표준화(Standardization)   각 feature의 측정 단위에 대한 보정  예) 아파트 값을 추정하기 위한 feature들 중 평수와 주변지역의 땅값  Centering and scaling을 통해서 평균이 0, 표준편차가 1이 되도록 변환해줌.  캡처.JPG   1. 결측치 대체   결측치를 최상의 추측값으로 대체.  Estimated statistic ( e.g., Mean, Median, Mode, Regression)  K-nearest neighbor  Tree-based   1. 결측치 대체 방법에 따른 비교   캡처.JPG   1. 중요하지 않은 Feature 제거 (filtering)   의미없는 변수들을 포함했을 때 RMSE의 변화  캡처.JPG   1. Lumping   매우 작은 빈도를 갖는 범주들을 모아서 하나의 범주로 재범주화 하는 것.  캡처.JPG   1. One-hot & dummy encoding   각 범주를 1 또는 0으로 표시  캡처.JPG   1. Label encoding   캡처.JPG   1. 차원 축소   캡처.JPG   1. 모델평가 지표   캡처.JPG  캡처.JPG  캡처.JPG | |