|  |  |
| --- | --- |
| 교육 제목 | 머신러닝 |
| 교육 일시 | 2021-10-13 |
| 교육 장소 | 집(비대면) |
| **교육 내용** | |
| 오전 | 선형과 비선형의 차이   * 파라미터의 값이 선형적으로 연결 되어 있는 경우를 선형이라 한다     손실함수    오차의 제곱을 통해 정확도를 판단.  제곱을 사용하여 부호를 없앰. (방향은 의미가 없고 크기만 의미가 있기 때문)  이때 손실함수의 최소값을 찾는 것이 가장 적절한 모델을 찾는 것  미분을 통하여 기울기의 값이 0이 되면 최소값을 찾을 수 있다. |
| 오후 | 기계학습의 분류    Label : 결과가 가져야 하는 답의 값, 학습 결과에 대한 답이 존재  PCA : 정보화 손실이 최소화 되도록 Feature를 줄이는 방법  데이터의 상관성을 고려하여 Feature를 제거  준지도 학습 (지도 학습과 비지도 학습이 섞여있는 경우)  라벨이 존재하는 데이터와 존재하지 않는 데이터가 섞여있는 것을 학습해야 하는 경우      Training set : 모델 알고리즘 학습. 모델에 사용 될 feature들을 결정, 초매개변수 조절 (약 전체 자료의 70%)  Test set : 최종 선택된 모델의 성능평가 (약 전체 자료의 30%, 자료의 수가 적을 경우 생략 가능)  일반적인 ML 예측 과정  학습세트(Training set) : 머신러닝 모델을 학습할 때 사용  검증세트(Validation set) : 하이퍼 파라메타 결정할 때 사용  테스트세트(Test set) : 학습 된 모델 평가할 때 사용    ML모델의 치우침과 분산  바이어스와 분산은 반비례 관계이다.    과소적합(Underfit)   * 모델이 너무 단순하여 에러가 발생한다 * 바이어스가 너무 높다 * 실제로는 잘 생기지 않는다 -> 현실에서는 모델이 복잡한 경우가 많음.     과적합(Overfit)   * 모델이 학습 데이터에 지나치게 치우쳐 새로운 데이터에 관하여는 정확한 값을 내놓지 못한다. * 모델이 너무 복잡하며 학습 데이터에 대한 에러는 없다 * 바이어스가 너무 낮다 (한쪽으로 쏠림) |