**[오픈소스개발프로젝트 2021F]**

**출석 과제**

주의사항: 본문 글자 크기, 자간, 장평 등 서식 변경 금지

작성자: 홍진우 학번: 2018037044

|  |
| --- |
| 해당 주차의 동영상을 시청하고 내용을 요약하시오. (공백포함 600자 이상) |
| 머신러닝에서 손을 안된 Raw data에 우선 관심없는 변수를 제거를 하고 변수값의 성질을 확인한 다음에의미와 일치하는지 확인을 하고 변수값의 분포를 확인한다. Regression 모델엔 fit된 선이 있고 거기에 에러를 계산할 수 있다. 실제 값에서 예측값을 빼서 제곱한 것이 mse라고 하는데 여기서 오류를 예측할 수 있다. Precision은 전체에서 실제 정답이 얼마나 되는지 recall은 정답 중에서 관련된 정답이 얼마나 있는지 나타낸 비율이다. 0~1사의의 값을 임계점을 갖고 0은 거짓 1을 참이라고 했을 때 confusion matrix를 4칸으로 그려 예측 값과 실제 값을 표현을 한다. 여기서 t.p.r,과 f.p.r을 구현할 수 있고 각축으로 그래프를 그려 기준선을 그려서 어떤 것이 정답에 가까운지 어떤 것이 성능이 더 좋은지 확인할 수 있다. Regression는 에러가 얼마나 나는지 확인하는 것이고 model evaluation metrics는 실제 정답율이 얼마나 될지 예측할 수 있다. Linear regression에 least square이 있다. 데이터 포인터가 있고 추세선이 있을 때 어떻게 간단하게 에러의 차이를 최소화할 수 있는지 확인할 수 있다. Ridge regression은 linear와 식이 거의 비슷하지만 페널티를 추가하여 data에 의도적으로 어느정도 빗나가게 예측하게끔 한다. Lasso는 ridge와 다르게 페널티에 절대값을 씌운다. Lasso에 다양한 해석법을 배웠다. |
| 동영상 내용에서 질문 사항이 있는 경우 작성하시오. |
| 1)  2) |