

2022년 2월 졸업대상자 물리학 졸업논문 신청 및 발전보고서

제출자	성명 : 김광민	학번 : 12183121	
지도교수	이재우 (인) <2021년 9월 17일까지 지도교수 확인을 받을 것>		
논문제목	Python으로 linear regression 문제 해결		
논문개요	<p>특정한 데이터 집합과 linear regression을 이용해 식을 갖고 함수 그래프를 그린다.</p> <p>이후 Python으로 코드를 구현해보고 데이터 집합에 적용해 본다.</p>		
<div style="display: flex; flex-direction: column; align-items: center;"> <div>단계별</div> <div>발전</div> <div>보고</div> </div>	1단계 (9/24)	<p>선형 회귀의 개념 작성</p> <p>선형 회귀, linear regression에 대한 대략적인 설명과 사용되는 용어, 파라미터 노치 (weight, bias) 부분, linear regression의 이점, 데이터셋의 라이브러리 설명, 내 코드의 설명</p> <p>- 1. linear regression을 구현할 수 있는 방법들 (최소제곱법을 중심으로) 소개</p> <p>2. 최소제곱법의 구체적인 이론과 사용될 공식</p> <p>3. linear regression의 해석</p> <p>해석식만 작성 (anscombe's quartet)</p> <p>결과: 비교, 그래프</p>	지도교수 확인
	2단계 (10/15)	<p>random, numpy, matplotlib과 같은 문헌 공부와 함께 regression과 관련된 통계 강의 수강</p>	지도교수 확인
	3단계 (11/5)	<p>linear regression의 이론 보충 (통계학 개론 참조)</p> <p>라이브러리에 대한 공부 & 추가로 이차 곡선 구현 예정</p> <p>dataset의 후보들을 정하고 있음. (예: 일방적인 옳고됨 random down / 수석 과감함)</p>	지도교수 확인
	4단계 (11/26)	<p>상관 계수 코딩 완료. 해당 코드를 이용해 regression 함수를 만들 것임. (코드는 몇 개나 설명에 못미침) 데이터셋은 난수를 발생시켜 랜덤하게 불 대칭이며 최종적으로는 Kaggle에서 가져온 dataset인 종고과 관련 자료를 이용해 가장 예측오차를 줄일 수 있는 코드를 확인하고 비교해 볼 예정</p>	지도교수 확인
결과 및 결론 포스터 발표 포함> (12/9)	<p>Python을 이용해서 기존에 생각한 라이브러리가 아닌 코드를 구현하여 Linear Regression 분석을 진행했고 성능상에 큰 차이가 없는 것을 보았다. 그와함께 Simple Linear Regression의 한계도 명확히 파악할 수 있었다.</p>		지도교수 확인