

Linear Regression

변재현 2022년 07월 05일

목차

- 선형회귀모델의 정의
- 회귀 모델의 종류
- 선형회귀 모델의 문제 : 다중공선성
- 선형회귀 모델 오차항
- 파라미터 추정 방법
- Least Squares Estimation Algorithm
- 데이터 설명 : 마케팅 데이터
- EDA(Exploratory Data Analysis)
- 실제값 예측값 시각화

선형회귀 모델의 정의

- 여러 개의 독립 변수와 한 개의 종속변수 간의 상관관계를 모델링 하는 기법
- 예측값과 실제 값의 RSS (Residual of Sum of Square)를 최소화해 OLS (Ordinary Least Square) 추정방식으로 구현한 클래스이다(파라미터 추정법에서 자세히)
- $RSS = \sum (y_i - \hat{y}_i)^2$ (실제값과 회귀 직선의 예측값의 차이) 제곱의 합
- RSS가 최소화가 되는 직선으로 회귀 계수들이 예측이 되는 것을 OLS라고 한다.

회귀 모델의 종류

1. 선형: Y 가 직선으로 표현
2. 비선형: Y 가 직선외의 선으로 표현
3. 단순: 독립변수가 1개
4. 다중: 독립변수가 2개 이상

선형회귀 모델의 문제 : 다중공선성

OLS 기반의 회귀계수 계산은 입력 피처의 독립성에 많은 영향을 받는다.

- 일반적으로 선형회귀는 피쳐들이 독립되어있다고 가정한다.
- 변수들 간의 상호작용은 일절 무시된다.
- 피쳐들 간의 상관관계가 높은 경우 분산이 매우 커져서 (연쇄작용 때문) 오류에 민감해진다. 이를 다중 공선성이라 한다.
- Ex) 변수들간의 양의 영향을 준다고 하면 회귀를 했을때 독립변수들 간의 영향이 포함이 되어 있어 예측값 분산이 더 커진다.v원래 1만큼 올라야 되는데 변수들 간의 상호작용이 포함되어 있어 2로 예측하는거. 상관관계가 높은 경우 예측값 분산이 커진다.
- 상관관계= $\frac{\text{예측값 분산}}{\text{실제값 분산}}$ 커짐

선형회귀 모델 오차항

오차항이 있는 경우와 없는 경우

- X변수만으로 Y를 표현

$$Y=f(x)$$

$$\text{회사이익} = f(\text{회사 영업이익} + \text{회사 자본이익})$$

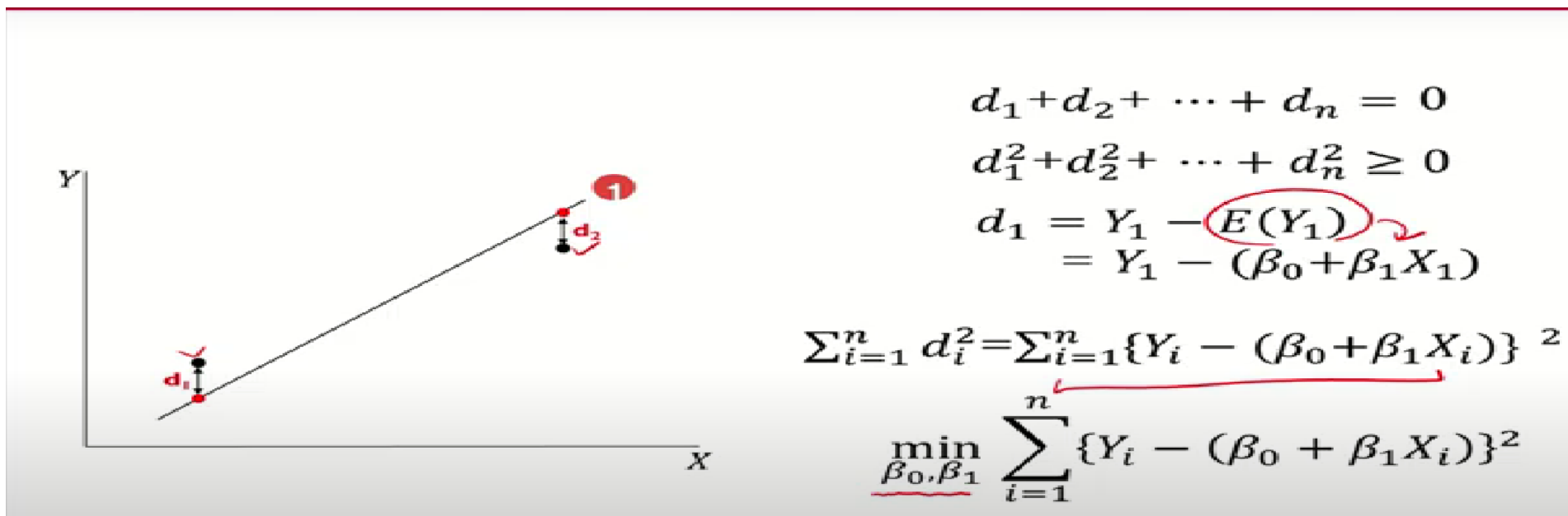
- X변수와 오차항(Y절편)으로 Y를 표현

$$Y=f(x)+\text{오차항}$$

$$\text{회사이익} = f(\text{물가} + \text{수요} + \text{공급}) + \text{오차항}$$

파라미터 추정 방법

Least Squares Estimation Algorithm



김성범.(2019.01.28)

예측값과 실제 값의 RSS를 최소화해 OLS 추정방식으로 구현한 클래스이다의 의미
점들과 수직으로서 직선과의 차이를 최소화 시키는 직선

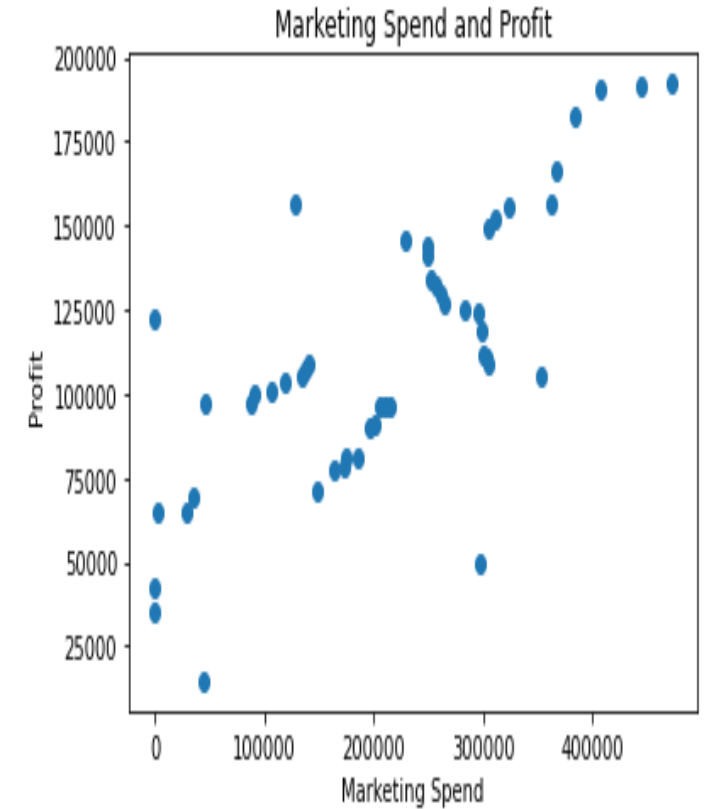
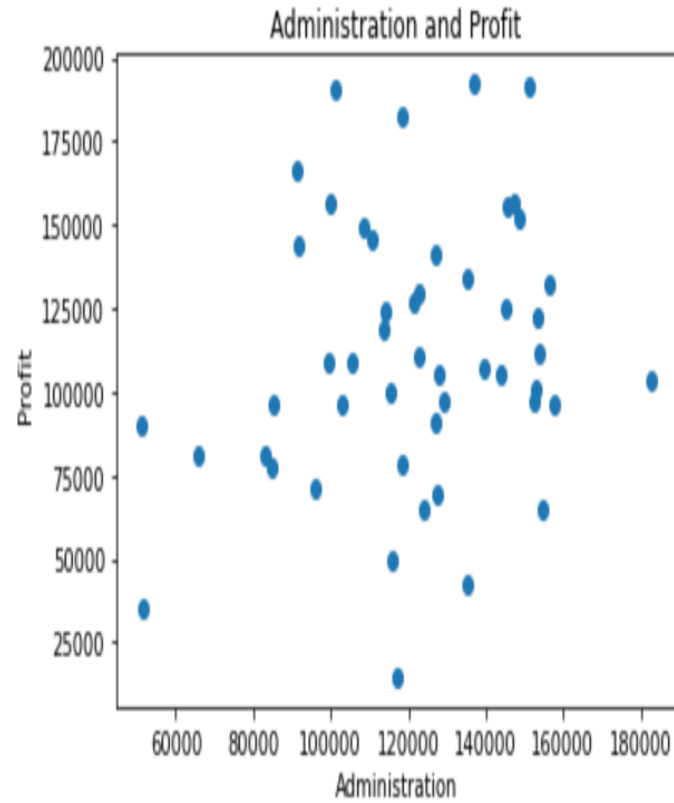
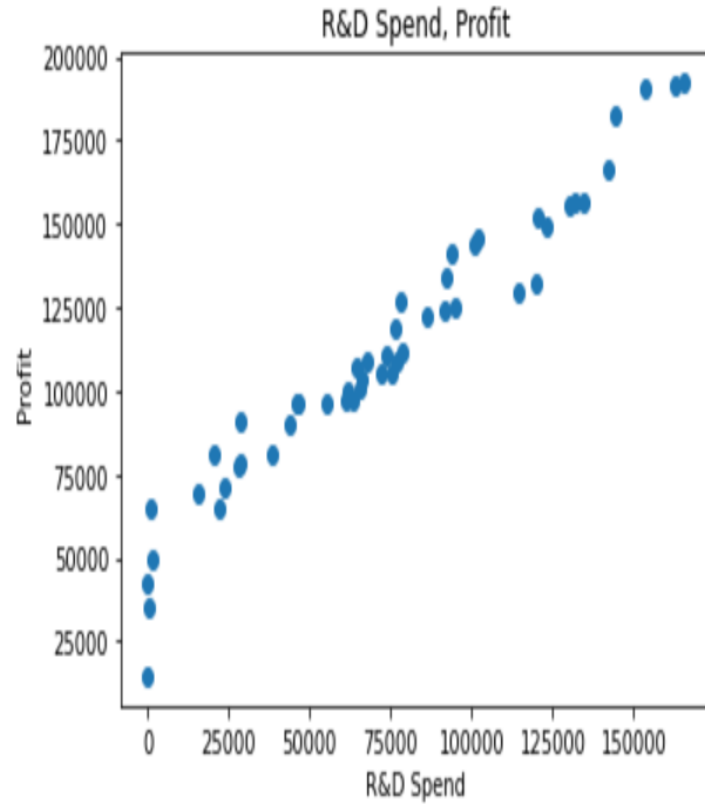
데이터 설명 : 마케팅 데이터

독립변수(원인 변수)

종속변수(결과 변수)

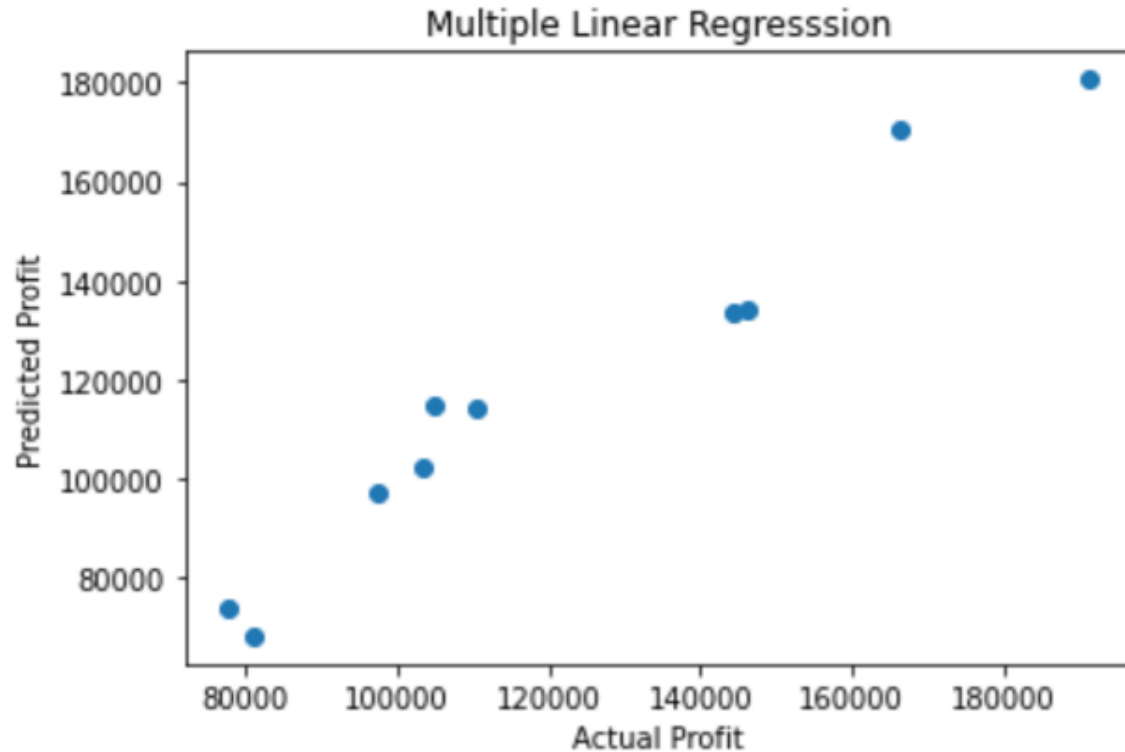
	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

EDA(Expiatory Data Analysis)



주요 변수는 R&D 비용, 마케팅 비용을 확인

실제값 예측값 시각화



MAE : 6886.5945882465185

MSE: 67220275.37568131

RMSE:8198.797190788495

R제곱: 0.9474386447268488

pvalue=array([[0.12782268],

X:실제값
Y:예측값

감사합니다

출처

김성범.[김성범소장 / 인공지능공학연구소].(2019. 1. 28.). [핵심 머신러닝] 선형회귀모델 1 (개요, 모델가정).[영상].유튜브. <https://www.youtube.com/watch?v=4Yo297HQyAk>

권철민, 파이썬 머신러닝 완벽가이드,위키북스,p,320,p321