

강정호 성적 예측

03.26

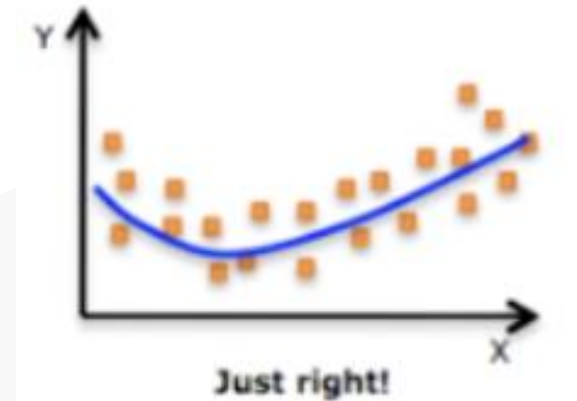


머신러닝(ML) 모델이란 ?

- Machine Learning

- 데이터 속에서 패턴을 발견하고 예측을 수행하는 알고리즘의 표현식

		닭다리	닭날개	닭가슴	판매량
학습		2	2	2	130
		3	1	4	190
		1	4	4	180
테스트		3	1	2	?

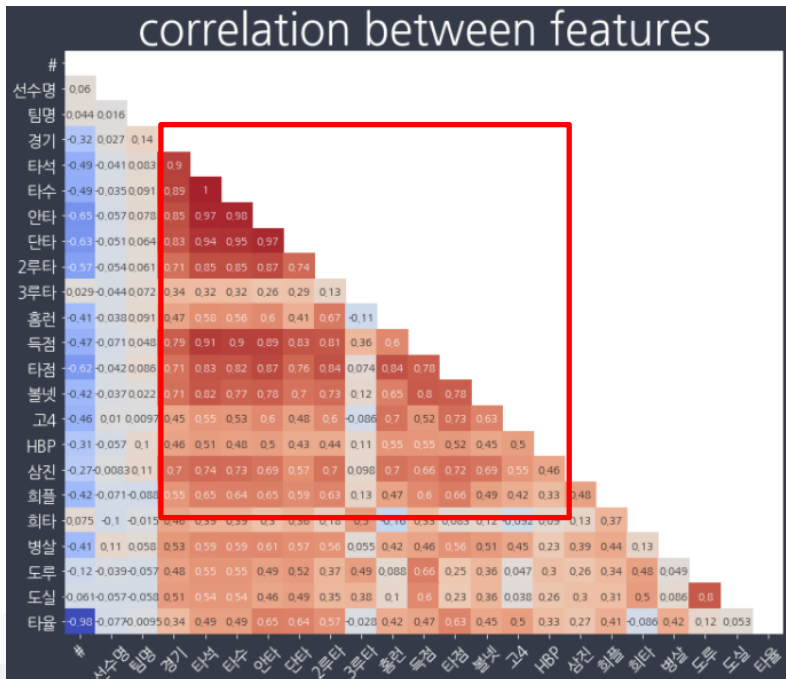


$$40x + 30y + 10z = \text{판매량}$$

데이터 살펴보기

■ KBReprot 21년 기록

#	선수명	팀명	경기	타석	타수	안타	단타	2루타	3루타	홈런	득점	타점	볼넷	고4	HBP	삼진	희플	희타	병살	도루	도실	타율
1	이정후	Hero	123	544	464	167	112	42	6	7	78	84	62	1	9	37	9	9	8	10	3	0.360
2	전준우	롯데	144	619	552	192	139	46	0	7	88	92	53	0	5	71	8	9	10	6	3	0.348
3	강백호	KT	142	628	516	179	122	40	1	16	76	102	104	10	0	85	8	8	7	10	5	0.347
4	홍창기	LG	144	651	524	172	140	26	2	4	103	52	109	4	16	95	2	2	6	23	8	0.328
5	양의지	NC	141	570	480	156	95	29	2	30	81	111	69	5	11	60	10	10	12	2	1	0.325



- 변수들간의 상관관계가 높음

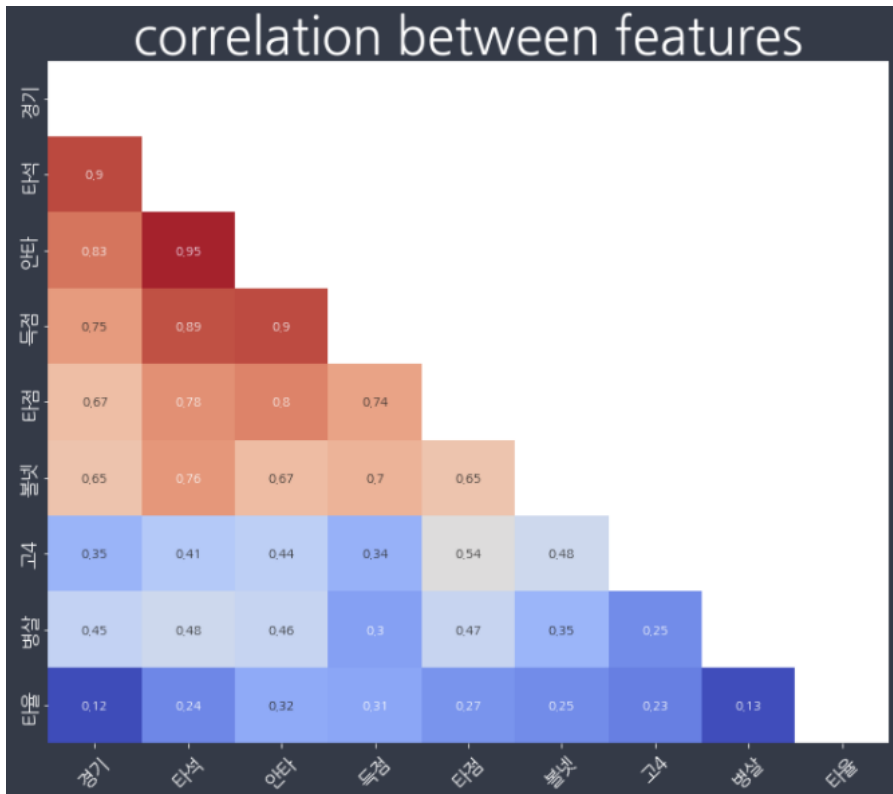
⇒ 다중공선성

⇒ 회귀분석에서 문제가 됨!

- 타석, 안타, 득점, 타점, 볼넷, 고4, 병살 선택

데이터 살펴보기

- KBReprot 21년 기록



- 여전히 다중공선성 문제가 있음
- 학습할 데이터가 많지 않음

데이터 살펴보기

학습	12년 타석	12년안타	12년 득점	12년 타점	12년 볼넷	12년 고4	12년 병살	13년 타율
	126	151	61	80	81	8	11	0.319
	⋮							
	13년 타석	13년안타	13년 득점	13년 타점	13년 볼넷	13년 고4	13년 병살	14년 타율
	430	110	41	52	73	7	14	0.365
	⋮							
테스트	19년 타석	19년안타	19년 득점	19년 타점	19년 볼넷	19년 고4	19년 병살	20년 타율
	568	151	78	63	52	1	6	?

XGBRegressor

■ 모델 검증 결과

```
cross_score.mean()    # 평균적으로 0.025정도의 오차가 있을  
  
-0.024872738930082558
```

$| \text{예측 값} - \text{실제 값} | = 0.025$
($\text{예측 값} - 0.0125 < x < \text{예측 값} + 0.0125$)

■ 강정호 타율 예측

- 강정호의 19,21년 KBO 성적은 김현수와 비슷할 거라 판단
- 강정호의 22년 KBO 성적은 21년 오재일과 비슷할 거라 판단

	타석	안타	득점	타점	볼넷	고4	병살	예측타율
19년	595	160	75	82	54	9	11	0.331(20년)
21년	595	144	70	96	77	5	7	0.288(22년)
22년	484	119	64	97	58	8	5	0.313(23년)

$0.3185 < x < 0.3435$

$0.2755 < x < 0.3005$

$0.3005 < x < 0.3255$

한계점

- 레이블을 내년 시즌 타율로 놓을 수 밖에 없어 성능이 잘 안나옴
- 데이터의 수와 쓸만한 피처들이 너무 부족함
 - 다중공선성 해결 못함
 - KBO에서 메이저리그에 갔다 온 타자들이 너무 적어 분석하기 힘들
 - KBO는 타구 속도, 발사 각도 등 트랙맨 데이터가 공유되지 않음
- 타고투저에 따른 가중치 반영하지 못함
- 강정호의 공백기간 능력 저하를 반영하지 못함