



2주차

📅 날짜	@2024년 10월 8일
🌟 상태	진행 중

3장 - 회귀 알고리즘

2진 분류 - 타깃값 0

회귀 - 임의의 타깃값

콜턴

동년배의 평균 키를 구하는 과정 중 회귀 용어 사용

회귀

임의의 숫자를 예측하는 것

k-최근접 이웃 회귀

- 가까운 위치의 타깃 찾기
- 타깃의 평균 내기

분류 - 이웃 클래스의 다수 클래스

회귀 - 이웃 클래스의 타깃 평균

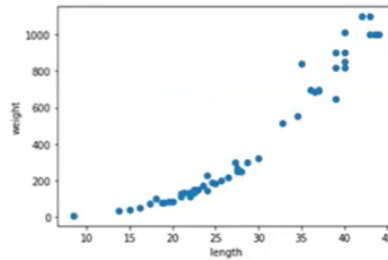
농어의 길이

perch_weight : 타깃

회귀는 임의의 숫자 - 일반적으로 특성 하나의 산점도 표현

```
import matplotlib.pyplot as plt

plt.scatter(perch_length, perch_weight)
plt.xlabel('length')
plt.ylabel('weight')
plt.show()
```



훈련 세트 - train_test_split 사용

(회귀에서 stratify 매개변수는 사용 x, 분류는 사용 o)

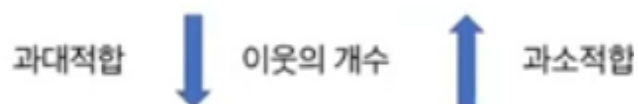
회귀 모델 - KNeighborsRegressor, mean_absolute_error 사용

$R^2 = 1 - \frac{(\text{타깃}-\text{예측})^2 \text{의 합}}{(\text{타깃}-\text{평균})^2 \text{의 합}}$

과대적합 : 편향은 낮지만 분산이 높은 경우

과소적합 : 분산은 낮지만 편향이 높은 경우

→ k-최근접 이웃 : 이웃 개수 조정



아주 큰 농어 - LinearRegression

농어 무게(y) = 기울기 * 농어 길이(x) + 절편

선형 회귀

산점도 : scatter

방정식 : lr.coef_+lr.intercept_

다항 회귀

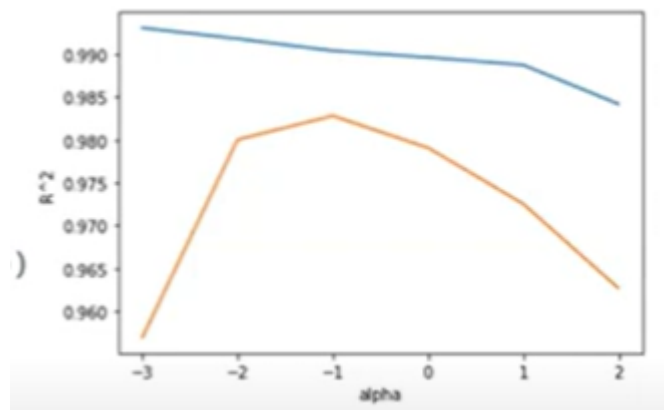
농어 무게(y) = $a * \text{길이}^2 + b * \text{길이} + c$

다중 회귀

특성(degree) 만들기 - PolynomialFeatures(변환기)

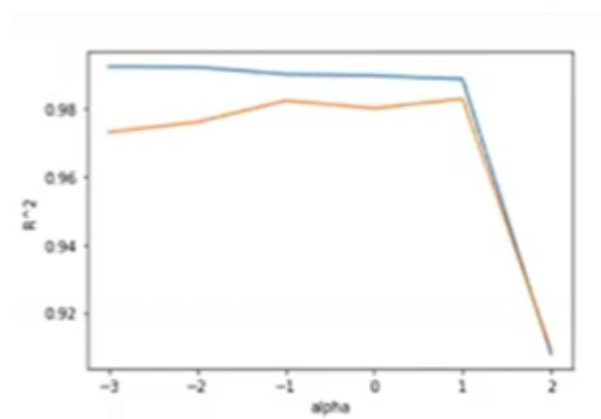
표준화 - StandardScaler

릿지 회귀- Ridge



→ 규제 조절

라쏘 회귀 - Lasso



→ 덜 중요한 가중치 제거