



세션 로딩 중

b a d ā



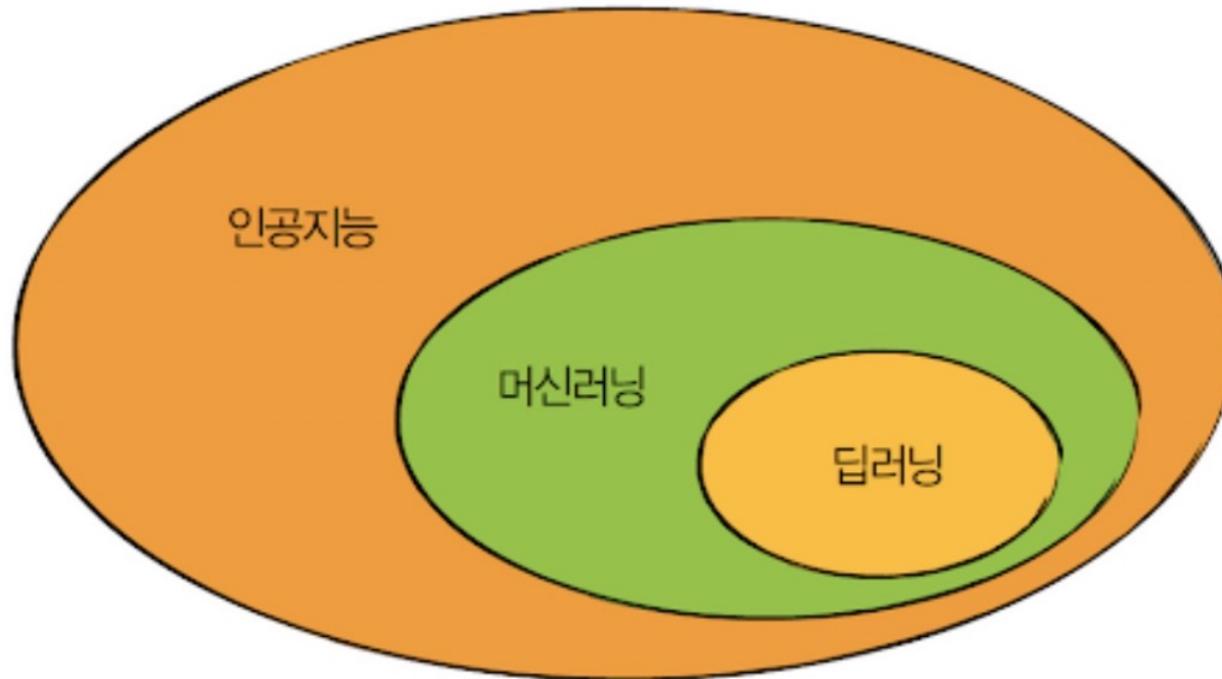
인공지능 기초 / 데이터셋 분리 / 최근접이웃

# | 학습목표

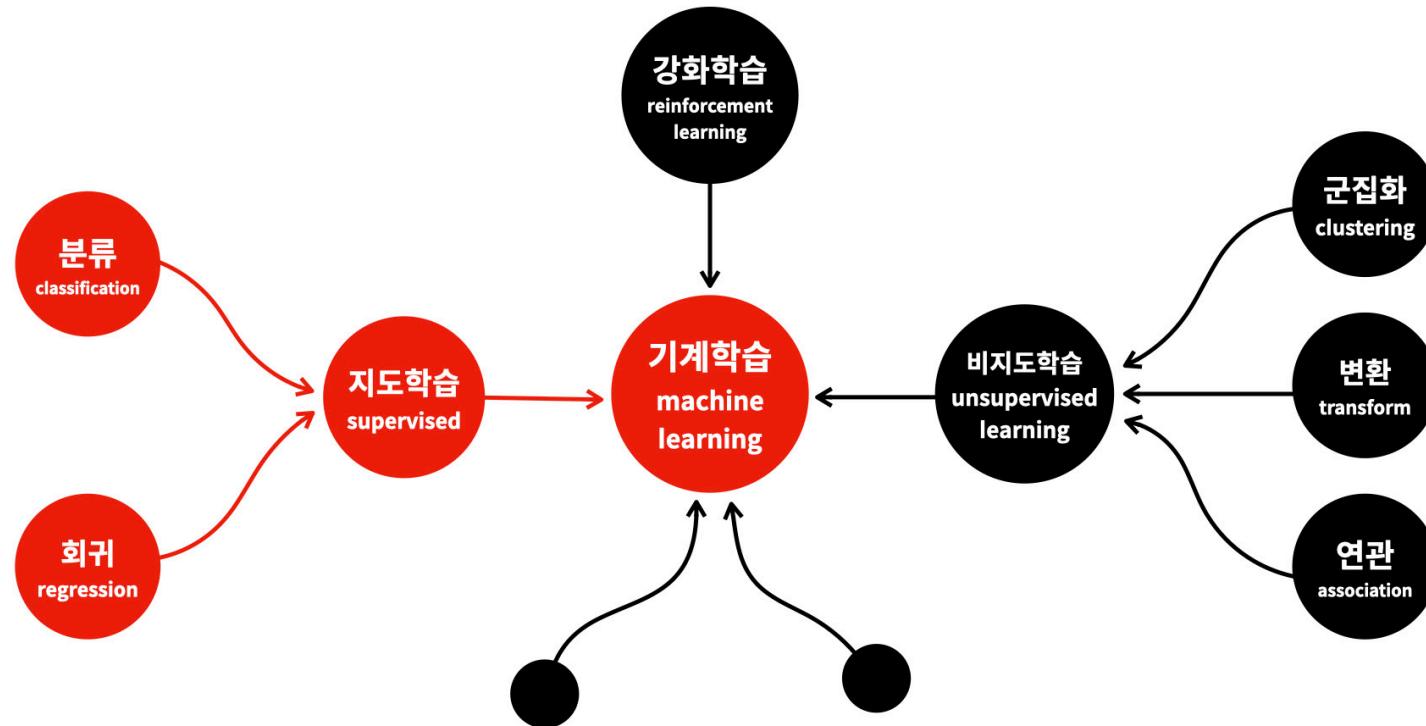
- 지도학습 이해 : 분류 vs 회귀
- 사례 기반의 회귀 vs 모델 기반의 회귀
- 최근접이웃 회귀 vs 선형 회귀
- 특성 공학과 규제

# 0. 지난시간 복습

# | 지난시간 복습

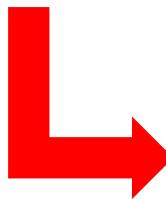


# | 지난시간 복습



# | 지난시간 복습

```
fish_length = bream_length + smelt_length  
fish_weight = bream_weight + smelt_weight  
  
fish_data = list(zip(fish_length, fish_weight))  
  
fish_data = list(zip(fish_length, fish_weight))  
  
fish_target = [1] * len(bream_length) + [0] * len(smelt_length)
```



df.head()

	fish_length	fish_weight	fish
0	25.4	242.0	1
1	26.3	290.0	1
2	26.5	340.0	1
3	29.0	363.0	1
4	29.0	430.0	1

# | 지난시간 복습

df.head()

	fish_length	fish_weight
0	25.4	242.0
1	26.3	290.0
2	26.5	340.0
3	29.0	363.0
4	29.0	430.0

독립변수



fish

1
1
1
1
1

종속변수

# 데이터 준비하기

df.head()

	fish_length	fish_weight
0	25.4	242.0
1	26.3	290.0
2	26.5	340.0
3	29.0	363.0
4	29.0	430.0

독립변수

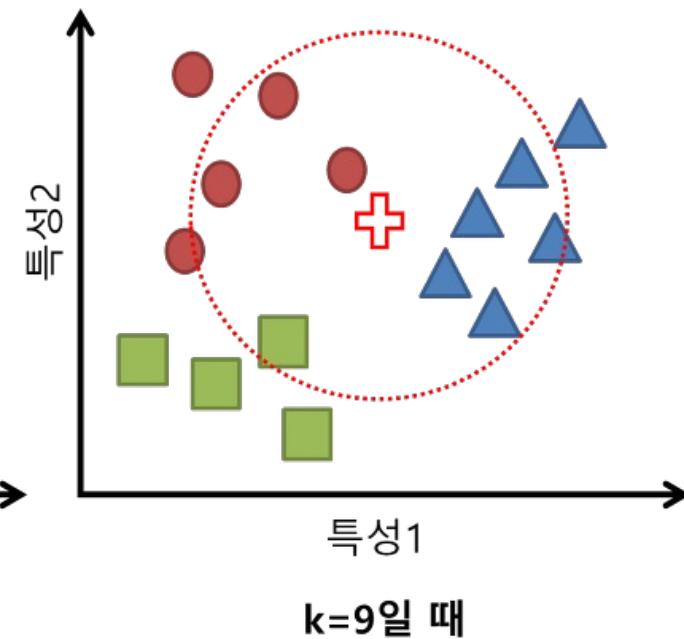
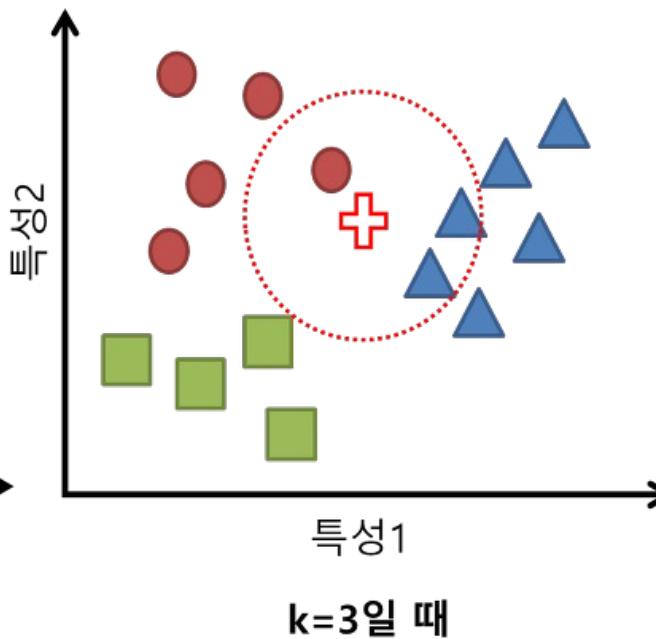
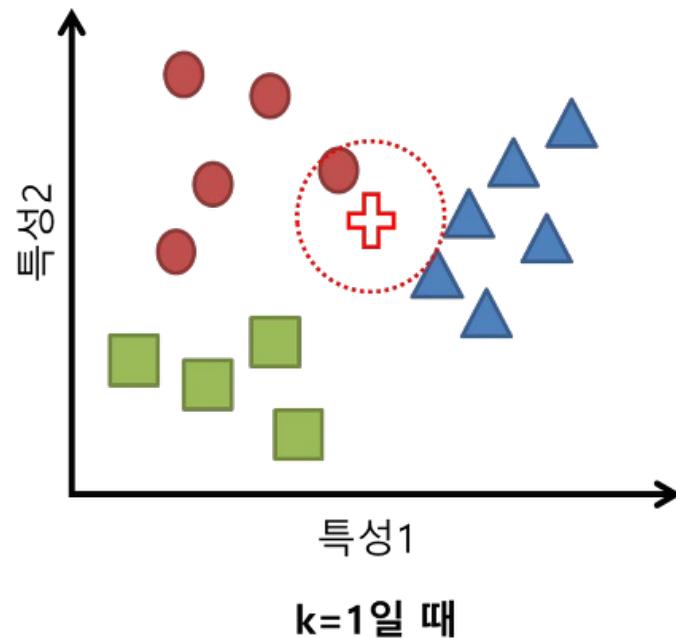


fish

1
1
1
1
1

종속변수

# 최근접이웃



# | 지난시간 복습

```
from sklearn.neighbors import KNeighborsClassifier  
  
kn_clf = KNeighborsClassifier()  
kn_clf.fit(fish_data, fish_target)  
KNeighborsClassifier()
```

독립변수 ←  
종속변수 ←

df.head()

	fish_length	fish_weight
0	25.4	242.0
1	26.3	290.0
2	26.5	340.0
3	29.0	363.0
4	29.0	430.0

fish

1
1
1
1
1

# 1. 지도학습의 이해 : 분류 vs 회귀

# | 지도학습의 이해 : 분류 vs 회귀

df.head()

	fish_length	fish_weight
0	25.4	242.0
1	26.3	290.0
2	26.5	340.0
3	29.0	363.0
4	29.0	430.0

독립변수



fish

1
1
1
1
1

종속변수

# | 지도학습의 이해 : 분류 vs 회귀

df.head()

fish\_length

0 25.4

1 26.3

2 26.5

3 29.0

4 29.0



fish\_weight

242.0

290.0

340.0

363.0

430.0

독립변수

종속변수

# | 지도학습의 이해 : 분류 vs 회귀

물고기의 종류를 예측하는 모델

vs

물고기의 무게를 예측하는 모델

# | 지도학습의 이해 : 분류 vs 회귀

물고기의 종류 = 1 or 2

vs

물고기의 무게 =  $\min < x \text{ and } x < \max$

# | 지도학습의 이해 : 분류 vs 회귀

물고기의 종류 = 범주형 데이터 like 객관식

vs

물고기의 무게 = 수치형 데이터 like 주관식

# | 지도학습의 이해 : 분류 vs 회귀

“가지고 있는 데이터에 **독립변수**와 **종속변수**가 있고,  
**종속변수**가 **숫자**일 때 **회귀**를 이용하면 됩니다.”

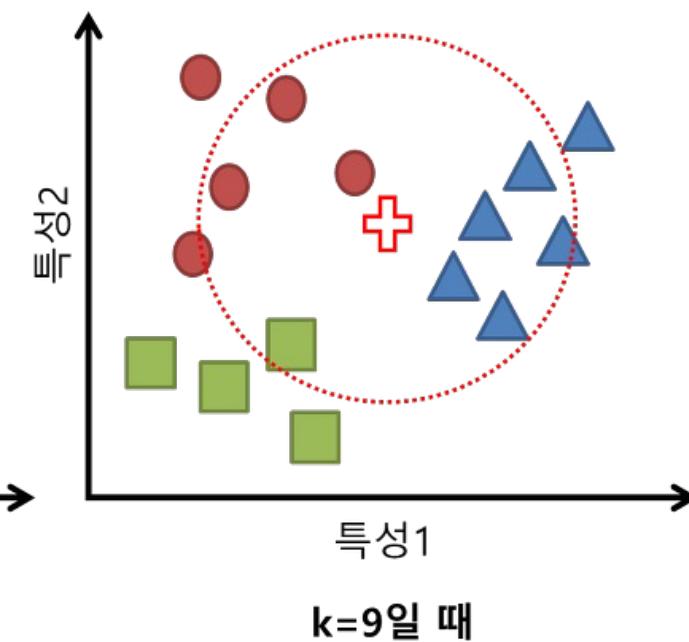
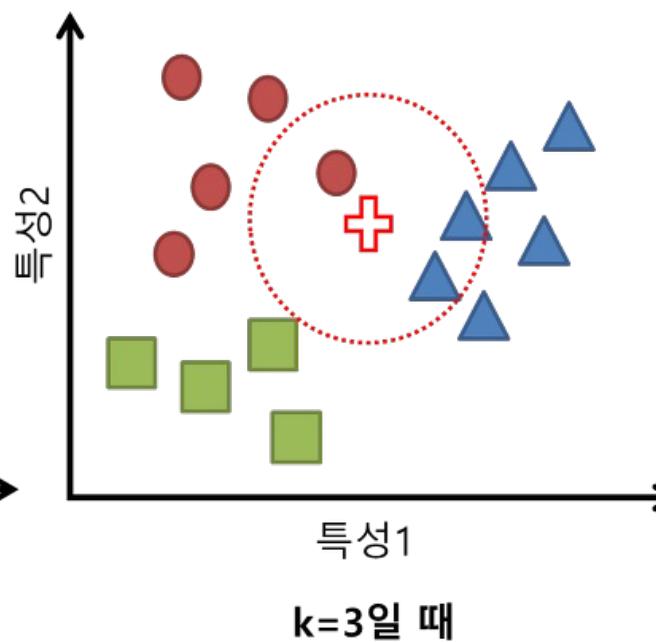
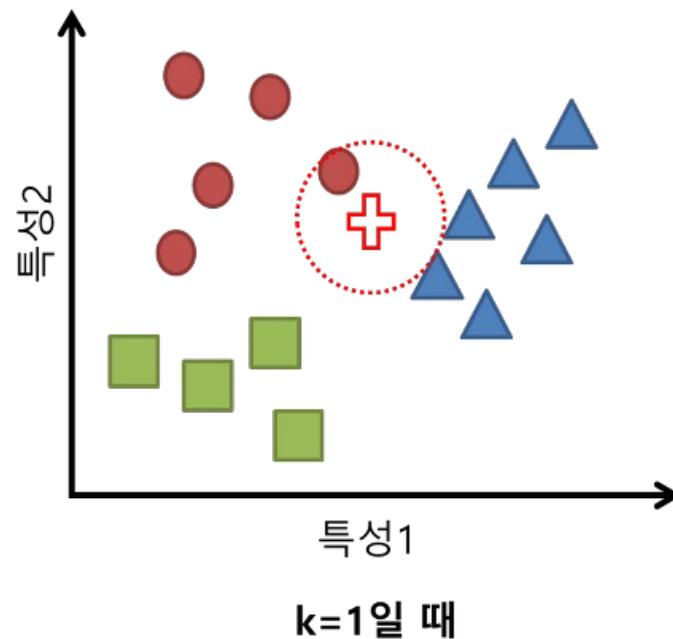
---

“가지고 있는 데이터에 **독립변수**와 **종속변수**가 있고,  
**종속변수**가 **이름**일 때 **분류**를 이용하면 됩니다.”

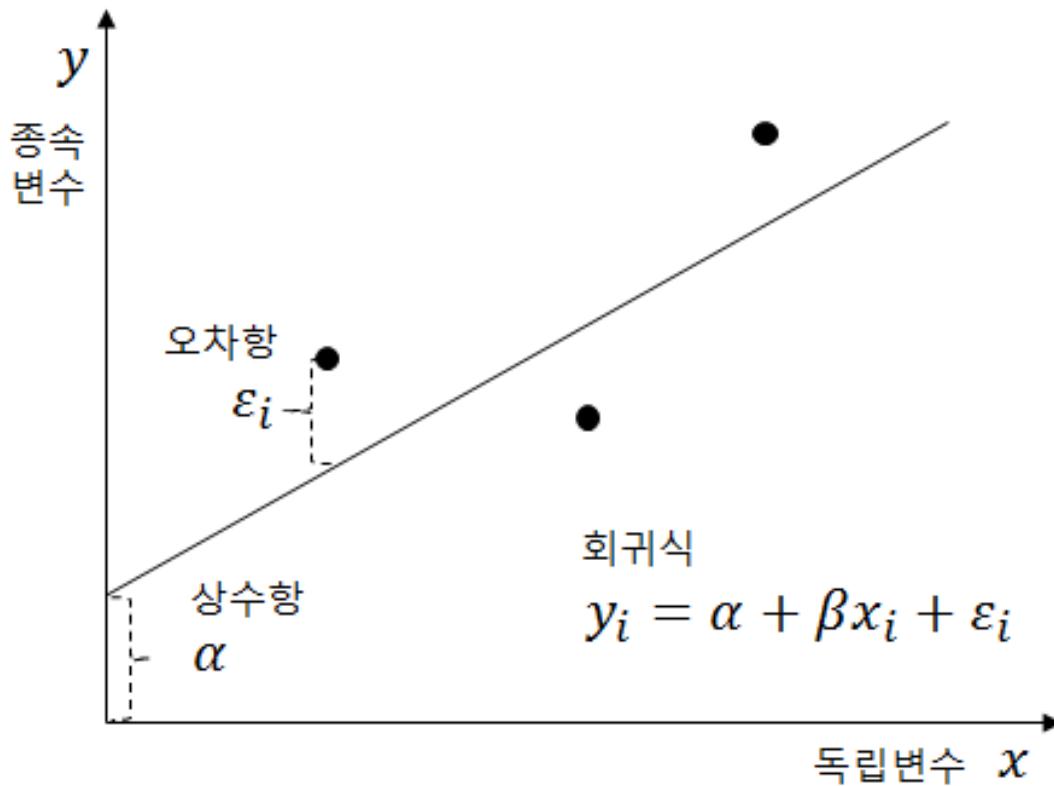
?

## 2. 사례 기반 회귀 vs 모델 기반 회귀

# | 사례 기반의 회귀



# | 모델 기반의 회귀

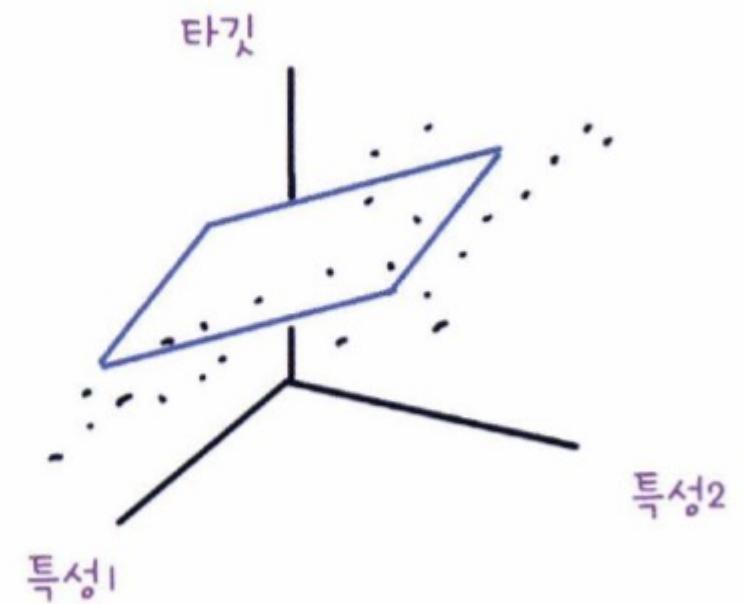
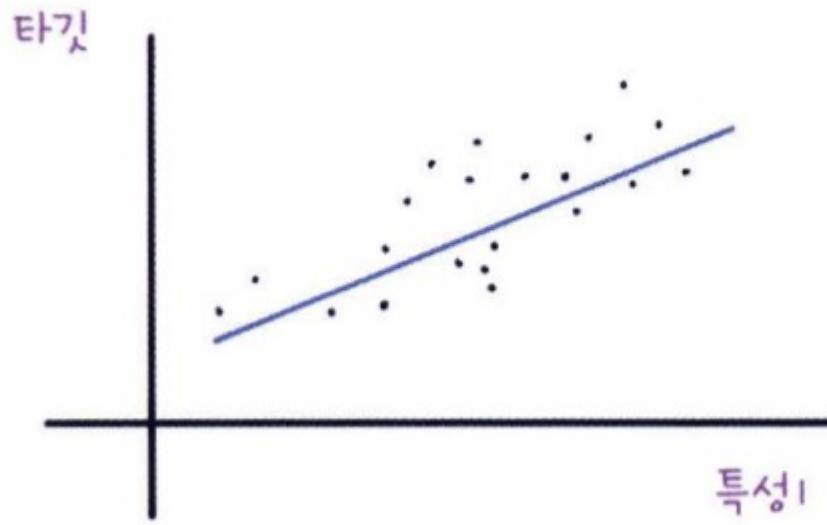


# | 사례 기반의 회귀 vs 모델 기반의 회귀

최근접이웃 vs 선형회귀

### 3. 특성 공학과 규제

# | 특성 공학과 규제



# | 지도학습의 이해 : 분류 vs 회귀

	<b>fish_length</b>	<b>fish_weight</b>
<b>0</b>	25.4	242.0
<b>1</b>	26.3	290.0
<b>2</b>	26.5	340.0
<b>3</b>	29.0	363.0
<b>4</b>	29.0	430.0

독립변수



	<b>np_fish_new_feature</b>
	6146.8
	7627.0
	9010.0
	10527.0
	12470.0

새로운 독립변수

# | 지도학습의 이해 : 분류 vs 회귀

	fish_length	fish_weight	np_fish_new_feature
0	25.4	242.0	6146.8
1	26.3	290.0	7627.0
2	26.5	340.0	9010.0
3	29.0	363.0	10527.0
4	29.0	430.0	12470.0

독립변수



종속변수

<코드 미리보기>

## | 코드 미리보기

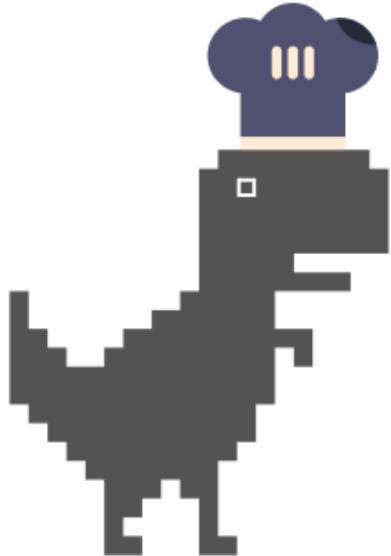
```
from sklearn.linear_model import LinearRegression  
  
# Modeling  
lr = LinearRegression()  
lr.fit(train_input_reshaped_polynomial, train_target)
```

# 코드 미리보기

```
print()
print(f'The Linear Model is\n y = {lr.coef_[0]} * x^2 + {lr.coef_[1]} * x + {lr.intercept_}')
print()
print(f'Prediction is\n{lr.predict([[4900, 70]])[0]}')
print()
print(f'Train Score is\n{lr.score(train_input_reshaped_polynomial, train_target)}')
print()
print(f'Test Score is\n{lr.score(test_input_reshaped_polynomial, test_target)}')
```

<실습>

# | 상황 살펴보기



티라노 닭강정은

인기 있는 신생 프렌차이즈 업체이다.

그러던 어느날,

티라노 사장님은

심각한 고민에 빠졌다.

그것은 바로

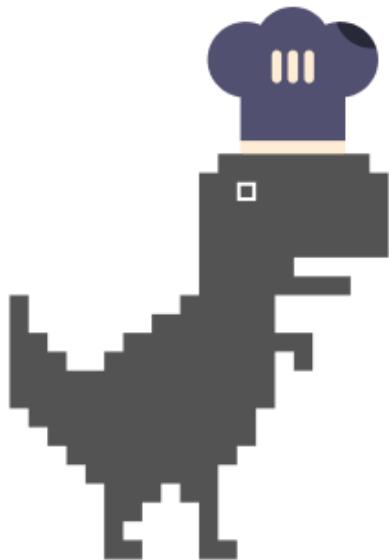
점포의 갯수는 점점 늘어나는데,

그에 따른 수요예측이 힘들었던 것이다.

그렇게 티라노 사장은

고려대학교 경영대학 경영데이터분석학회 bada를 찾게 되는데...

# | 상황 살펴보기



종속변수

점포의 갯수

독립변수

수요량



?

세션 끝!