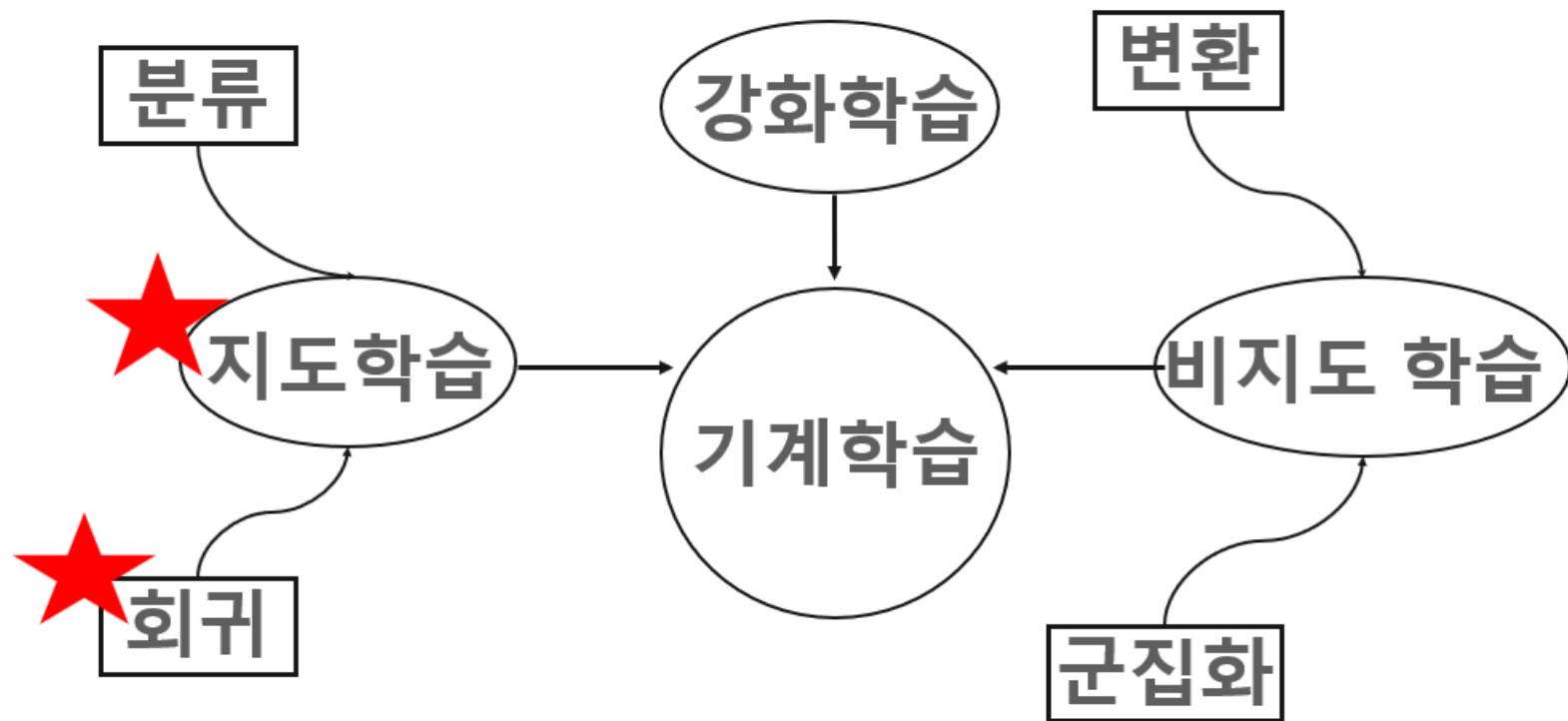


# 지도학습: 회귀분석

홍익 대학교  
Hyun-Sun Ryu

# 머신러닝의 종류



# 머신러닝의 종류

| 온도 | 판매량 |
|----|-----|
| 20 | 40  |
| 21 | 42  |
| 22 | 44  |

**양적  
데이터**  
(Numerical Data)



**회귀  
regression**

| 나이 | 생존 |
|----|----|
| 12 | O  |
| 35 | X  |
| 69 | O  |

**범주형  
데이터**  
(Categorical Data)



**분류** 이중분류  
다중분류  
**classification**

# 지도학습

- 과거의 데이터로부터 학습해서 결과를 예측하는 데 주로 사용
- 제시된 데이터를 보면 온도와 아이스크림 판매량은 원인과 결과로 맺어짐.
- 우리가 예측하고 싶은 데이터는 4월 6일의 아이스크림 판매량
- 과거의 데이터(4월 1일~4월 5일)를 활용하여 온도가 25도라고 예측된 4월 6일의 아이스크림 판매량을 예측

| 날짜          | 온도 | 아이스크림 판매량 |
|-------------|----|-----------|
| 2021. 4. 1. | 20 | 60        |
| 2021. 4. 2. | 21 | 63        |
| 2021. 4. 3. | 22 | 66        |
| 2021. 4. 4. | 23 | 69        |
| 2021. 4. 5. | 24 | 72        |
| 2021. 4. 6. | 25 |           |

과거의 데이터

알고싶은 데이터

# 지도학습

- 알고 싶은 데이터를 추측하기 위해서는 머신러닝의 지도학습이 이용
- 단 종속변수가 **양적 데이터**나 **범주형 데이터**의 형태를 가져야 함.
- 지도학습을 위해서는 충분히 많은 데이터가 필요
- 데이터는 원인이 되는 **독립변수**와 결과가 되는 **종속변수**의 구조로 이루어져 함.
- 머신러닝 지도학습을 통해 데이터를 기반으로 모델을 생성
- 예시에 따르면 판매량=(온도) $\times$ 3이라는 모델을 생성. 제시된 예시는 수가 적지만 생필품 예측, 농산물 예측 등 큰 데이터를 기반으로 했을때 사람이 계산을 통해 예측하는 것은 불가능
- 따라서 컴퓨터를 활용해 이를 예측하는 것이 매우 중요함

# Simple Practice

- Data 파일: 판매량 데이터.xlsx

The screenshot shows a data analysis tool interface. On the left is a vertical toolbar with icons for various functions. The main area displays a 'File' dialog box with a red box highlighting the '...' button next to the file name '판매량 데이터.xlsx'. Below the dialog, a 'File' icon is shown in a red box. To the right, a 'File' window displays the loaded data. The 'Info' section shows: 5 instance(s), 3 feature(s) (no missing values), Data has no target variable, and 1 meta attribute(s). The 'Columns (Double click to edit)' section contains a table with the following data:

|   | Name | Type     | Role    | Values |
|---|------|----------|---------|--------|
| 1 | 날짜   | datetime | feature |        |
| 2 | 온도   | numeric  | feature |        |
| 3 | 판매량  | numeric  | feature |        |
| 4 | 요일   | text     | meta    |        |

At the bottom of the 'File' window, there are buttons for 'Browse documentation datasets', 'Reset', and 'Apply', along with a status bar showing a question mark, a file icon, and the number 5.

# Simple Practice

- 데이터 Column별 type과 role을 설정
- type은 날짜의 경우 datetime, 온도는 numeric, 판매량은 numeric, 요일은 text
- 온도에 따른 판매량을 알기위함 이므로 판매량은 target, 온도는 feature로 설정

The screenshot shows a window titled 'File' with a dropdown menu set to '판매량 데이터.xlsx'. Below the file selection, there is an 'Info' section stating: '5 instance(s)', '3 feature(s) (no missing values)', 'Data has no target variable.', and '1 meta attribute(s)'. The main section is titled 'Columns (Double click to edit)' and contains a table with the following data:

|   | Name | Type     | Role    | Values |
|---|------|----------|---------|--------|
| 1 | 날짜   | datetime | feature |        |
| 2 | 온도   | numeric  | feature |        |
| 3 | 판매량  | numeric  | target  |        |
| 4 | 요일   | text     | meta    |        |

At the bottom of the window, there is a 'Browse documentation datasets' button, a 'Reset' button, and an 'Apply' button. The status bar at the very bottom shows icons for help, save, and a count of 5.

# Simple Practice

- 날짜의 경우 target에 영향을 주는 요인은 아니지만 날짜별 파악을 하면 좋기 때문에 참고할 사항인 meta로 설정
- 요일의 경우 target에 영향을 주지 않을 뿐더러 굳이 필요없는 내용이므로 skip

The screenshot shows a data loading window with the following details:

- File:** 판매량 데이터.xlsx
- Info:** 5 instance(s), 3 feature(s) (no missing values), Data has no target variable, 1 meta attribute(s)
- Columns (Double click to edit):**

|   | Name | Type     | Role    | Values |
|---|------|----------|---------|--------|
| 1 | 날짜   | datetime | meta    |        |
| 2 | 온도   | numeric  | feature |        |
| 3 | 판매량  | numeric  | target  |        |
| 4 | 요일   | text     | skip    |        |

At the bottom, there are buttons for "Browse documentation datasets", "Reset", and "Apply". A status bar at the very bottom shows icons for help, data, and a count of 5.



# Simple Practice

- Data table 추가
- 판매량은 회색 음영, meta데이터는 노란색 음영, 요일은 빠진것을 확인

The screenshot shows the Orange3 data mining software interface. A workflow is visible with a 'File' widget connected to a 'Data Table' widget. The 'Data Table' widget's configuration window is open, displaying the following information:

**Data Table**

Info  
5 instances (no missing data)  
1 feature  
Numeric outcome  
1 meta attribute

Variables  
☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

Selection  
☒ Select full rows

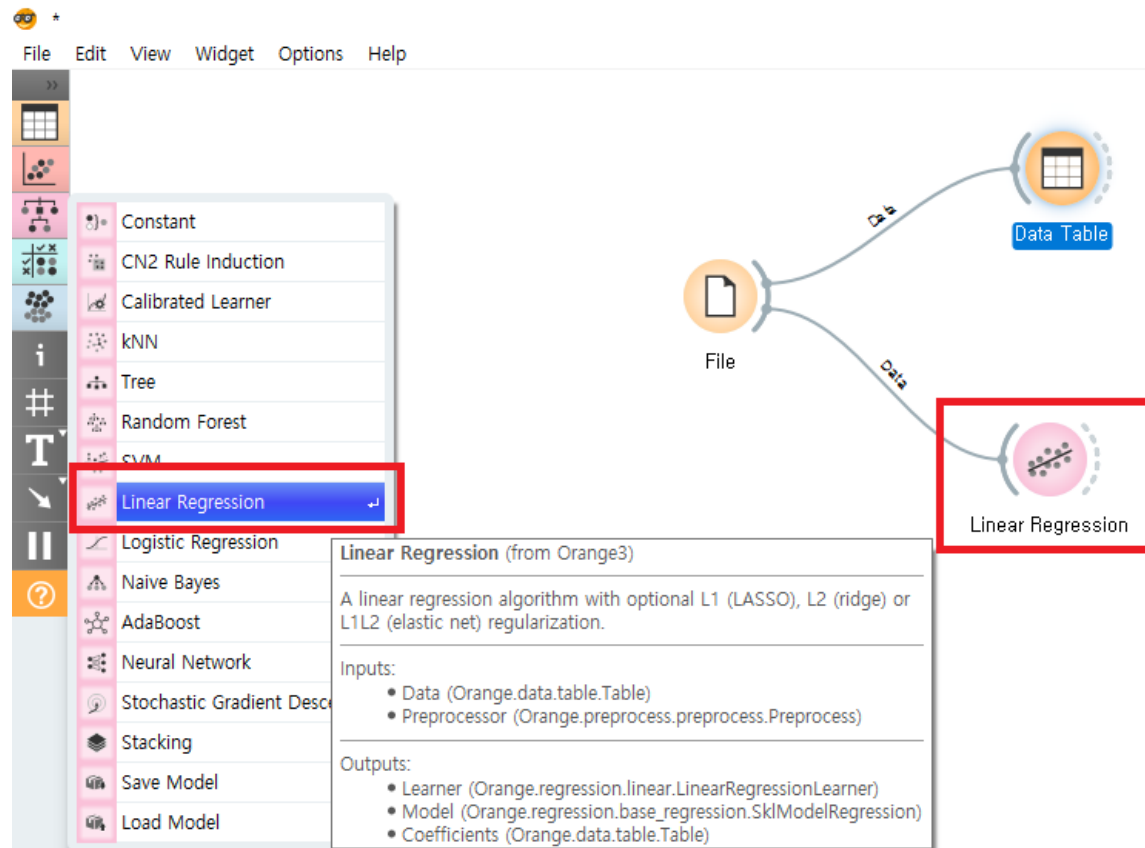
Restore Original Order

☒ Send Automatically

The data table contains 5 instances. The first three columns are '판매량' (Sales Volume), '날짜' (Date), and '온도' (Temperature). The '판매량' column is highlighted in gray, and the '날짜' and '온도' columns are highlighted in yellow. The data is as follows:

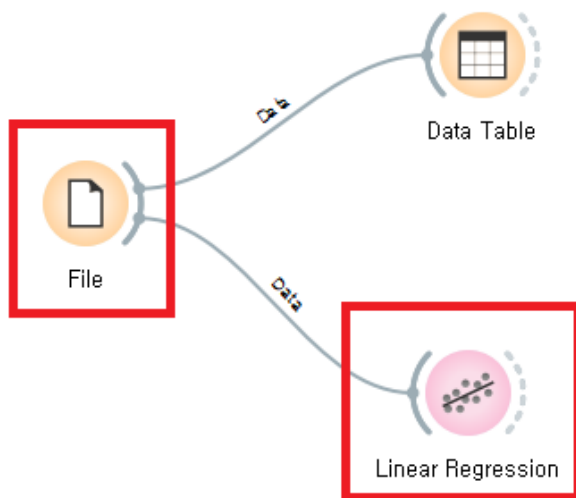
|   | 판매량 | 날짜             | 온도 |
|---|-----|----------------|----|
| 1 | 60  | 2021-04-01 ... | 20 |
| 2 | 63  | 2021-04-02 ... | 21 |
| 3 | 66  | 2021-04-03 ... | 22 |
| 4 | 69  | 2021-04-04 ... | 23 |
| 5 | 72  | 2021-04-05 ... | 24 |

# Simple Practice



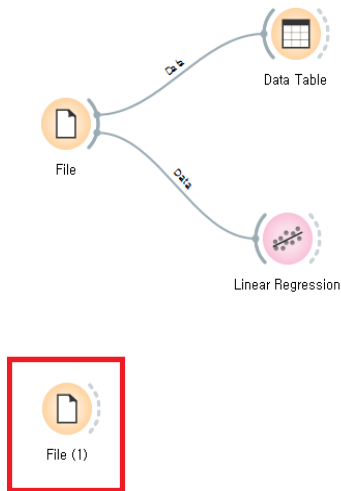
# Simple Practice

- 기존 데이터와 Linear Regression 위젯을 연결한다는 것은 Linear Regression이라는 학습 방법으로 파일의 데이터를 학습하여 모델을 만드는 것
- 기존 데이터에 따르면  $y=3x$ 라는 모델을 만들어내는 것



# Simple Practice

- 예측하고 싶은 데이터를 추가하도록 합니다. 파일 위젯을 불러와 데이터를 업로드
- 예측하고 싶은 데이터의 type과 role을 설정합니다. 이 데이터에는 판매량(target)이 없는 것 확인




The screenshot shows the 'File (1)' widget interface. The 'File' dropdown is set to '예측 데이터.xlsx'. The 'Info' section indicates 7 instances, 2 features, and 1 meta attribute. The 'Columns' table is as follows:

| Name | Type     | Role    | Values |
|------|----------|---------|--------|
| 날짜   | datetime | meta    |        |
| 온도   | numeric  | feature |        |
| 요일   | text     | skip    |        |

At the bottom, there is a 'Browse documentation datasets' button, a 'Reset' button, and an 'Apply' button.

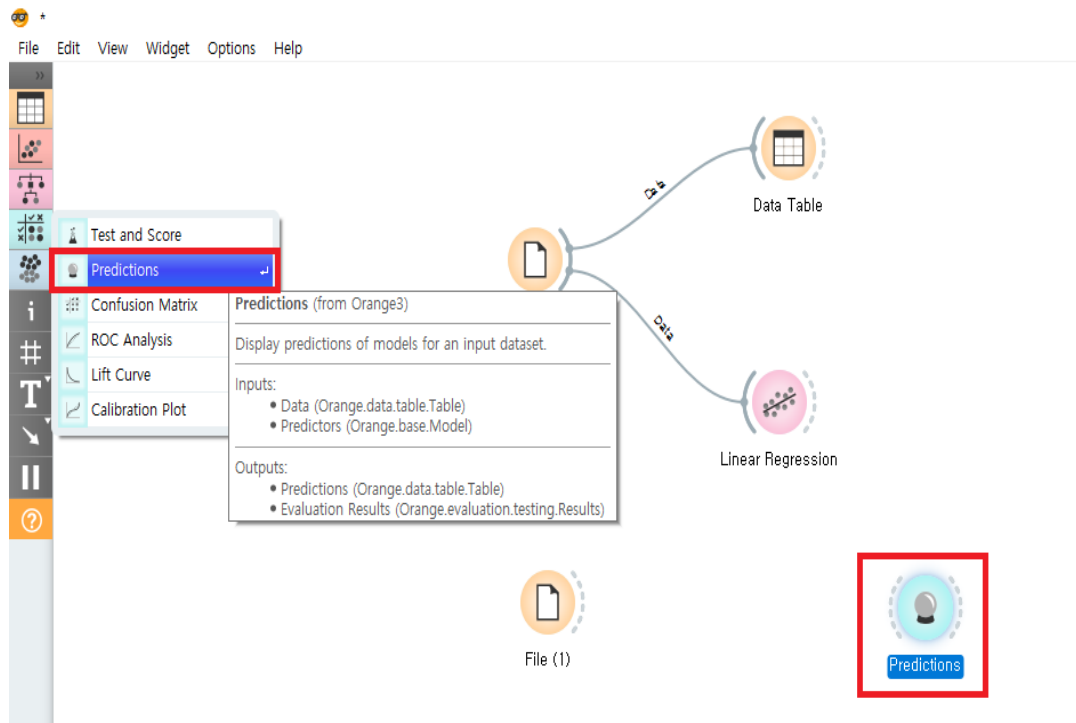
# Simple Practice

| 위젯   | 설명                              | 입력   | 출력   |
|--|---------------------------------|------|--|
| <br>Predictions | 입력 데이터 세트에 대한 모델의 예측값을<br>표시한다. | Data | Predictors, Predictions,<br>Evaluation results |

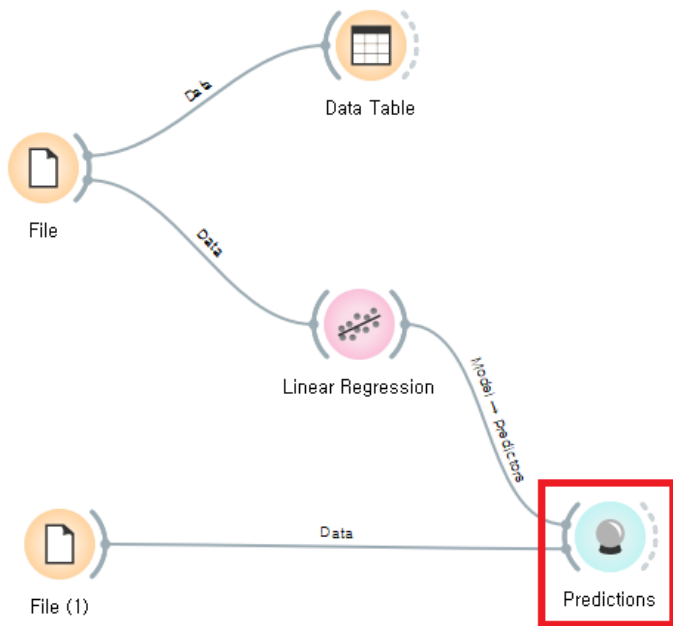
- Predictions 위젯은 데이터 세트와 하나 이상의 예측 변수(알고리즘이 아닌 예측 모델)를 수신하여 데이터와 예측을 출력
- Predictions 위젯은 예측 모델의 확률과 최종 결정을 보여줌.

# Simple Practice

- Linear Regression을 활용해 모델을 생성했고 이에 따라 예측하고 싶은 새로운 데이터가 있음.
- 이를 예측하기 위해서는 Predictions 위젯이 필요.
- 왼쪽 Evaluate 메뉴에서 Predictions 위젯을 클릭하거나 드래그&드랍



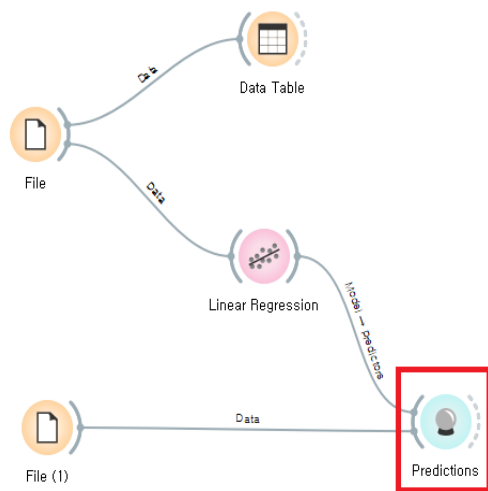
# Simple Practice



- Predictions 위젯은 두 가지가 필요
- 첫째, Linear Regression 모델이 필요
- 둘째, 예측하고 싶은 데이터가 필요
- 셋째, 이 두가지를 연결

# Simple Practice

- 예측한 내용을 확인하기 위해 Predictions 위젯을 더블클릭



Predictions

Show probabilities for

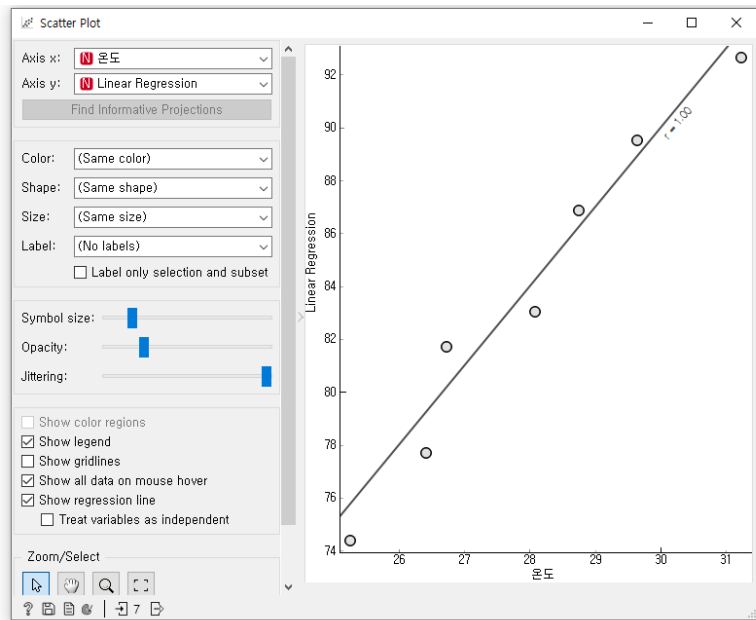
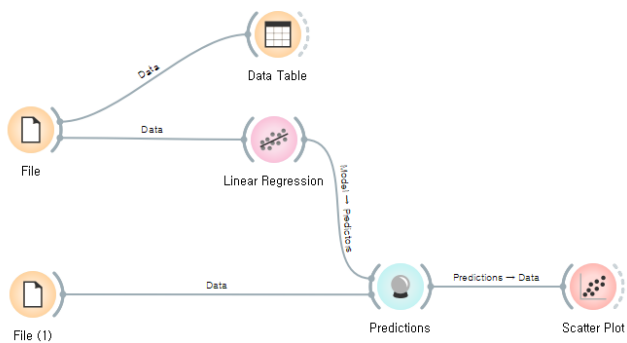
|   | Linear Regression | 날짜             | 온도 |
|---|-------------------|----------------|----|
| 1 | 75                | 2021-05-01 ... | 25 |
| 2 | 78                | 2021-05-02 ... | 26 |
| 3 | 81                | 2021-05-03 ... | 27 |
| 4 | 84                | 2021-05-04 ... | 28 |
| 5 | 87                | 2021-05-05 ... | 29 |
| 6 | 90                | 2021-05-06 ... | 30 |
| 7 | 93                | 2021-05-07 ... | 31 |

Restore Original Order



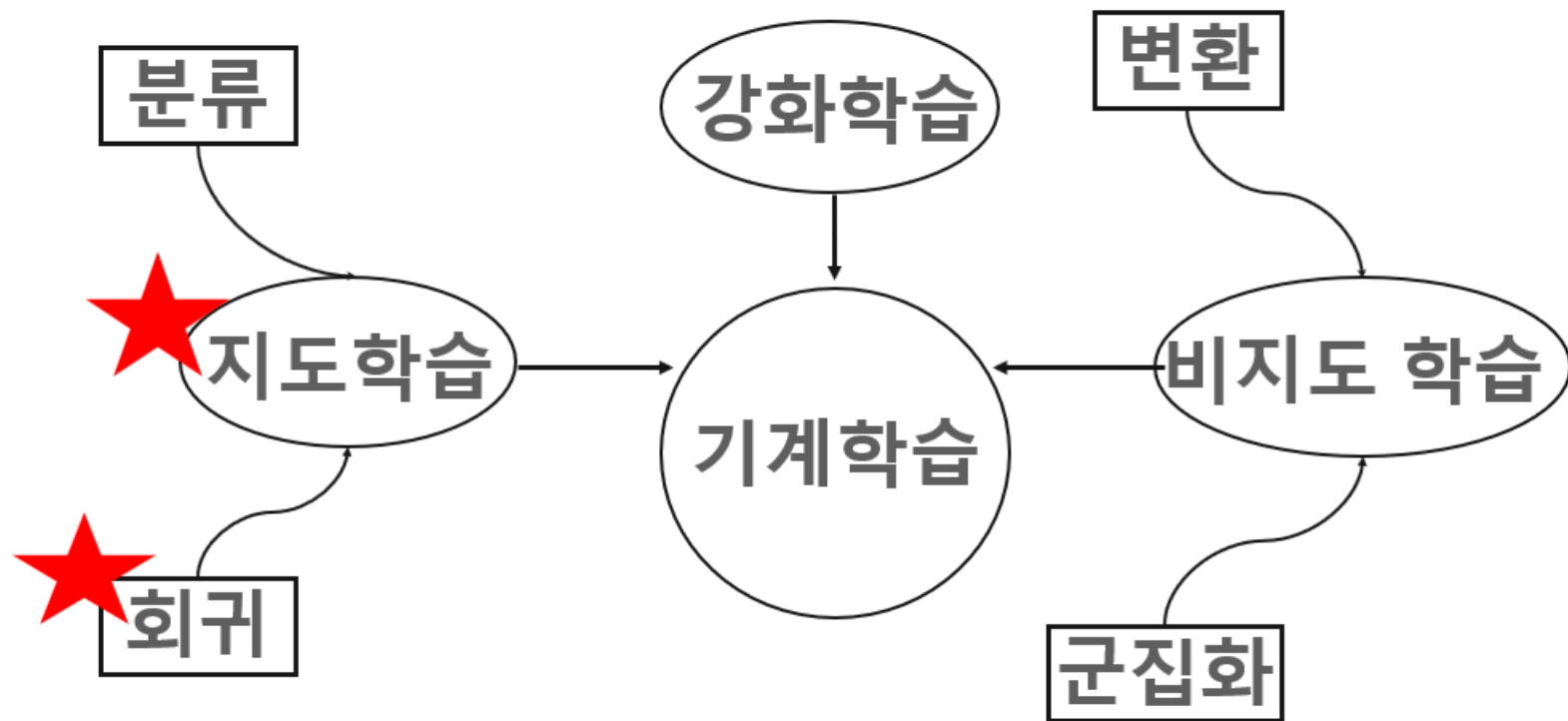
# Simple Practice

- 결과를 시각적으로 확인하기 위해서 Scatter Plot위젯을 추가
- x축에는 온도를 설정하고 y축에는 Linear Regression값을 설정하여 분포를 확인



# 회귀분석(Regression)

# 머신러닝의 종류



# 회귀분석

- 회귀(Regression)란 말은 다시 돌아간다는 뜻으로, 평균으로 돌아간다는 뜻
- 통계학에서 사용하는 자료 분석 방법 중 하나
- 주어진 자료들이 특정한 경향성을 띠고 있다는 아이디어로부터 출발
- 여러 자료들 간의 관계성을 수학적으로 추정하고 이를 설명
- 어떤 변수들이 한 변수의 원인이 되는지 분석하는 방법
- 원인으로 작용하는 변수를 독립변수(X)라고 하며, 결과로 나타나는 변수를 종속변수(Y)
- 일반적으로 '통계의 꽃'이라고 불림.

# 머신러닝의 종류

| 온도 | 판매량 |
|----|-----|
| 20 | 40  |
| 21 | 42  |
| 22 | 44  |

**양적  
데이터**  
(Numerical Data)



**회귀  
regression**

| 나이 | 생존 |
|----|----|
| 12 | O  |
| 35 | X  |
| 69 | O  |

**범주형  
데이터**  
(Categorical Data)



**분류  
classification**

# 회귀분석



독립변수의 수

- 단순회귀분석(1개)
- 다중회귀분석(2개 이상)



독립변수와  
종속변수의 관계

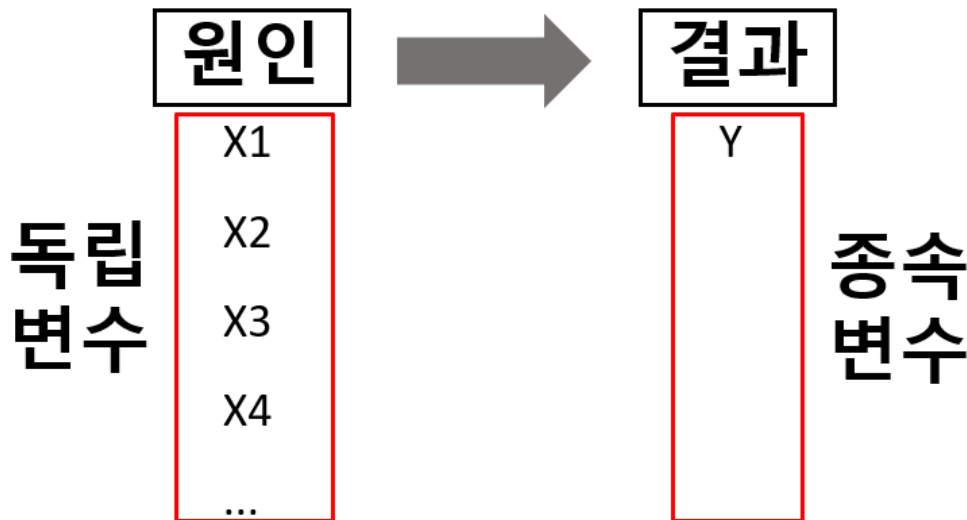
- 선형회귀분석(선형)
- 비선형회귀분석(비선형)

독립변수의 척도

- 일반회귀분석
- 더미변수를 이용한 회귀분석

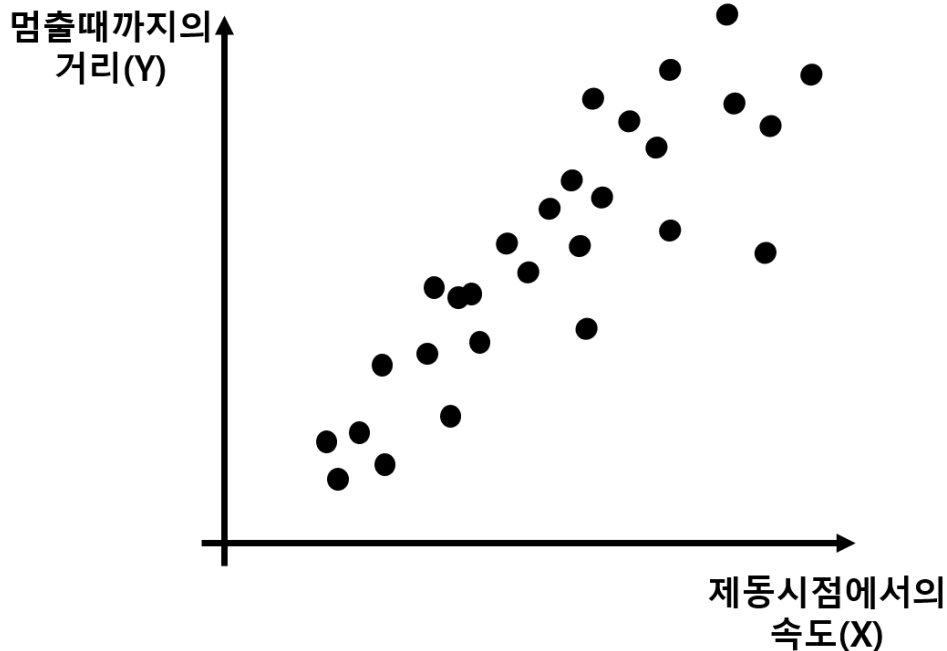
# 단순회귀분석, 다중회귀분석

- **단순회귀분석**은 원인이 되는 **변수가 1개**일 때 임. 즉 X1만 원인의 결과로 Y가 나타남.
- **다중회귀분석**은 원인이 되는 **변수가 2개 이상**일 때 임. 즉 X1, X2, X3 등 다양한 원인의 결과로 Y가 나타나는 것임



# 단순, 선형회귀분석

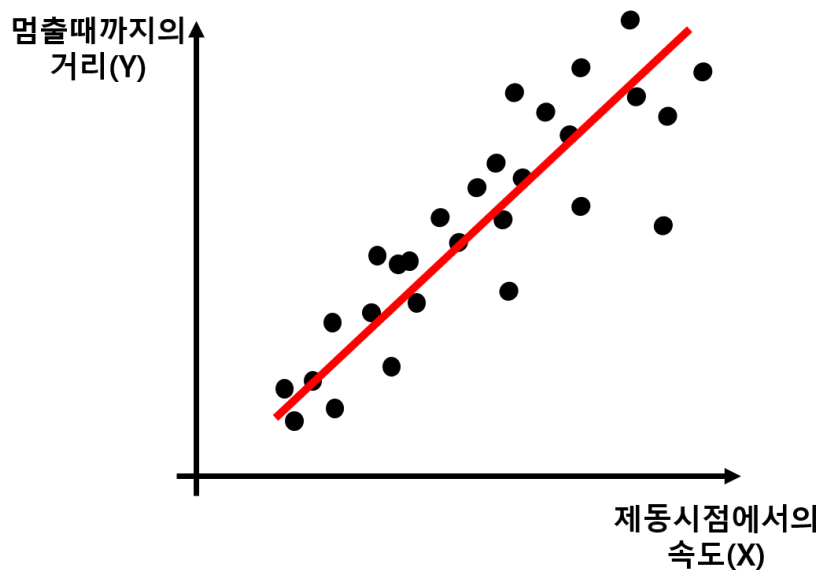
- X축은 제동시점에서의 속도
- Y축은 제동시점부터 자동차가 완전히 멈출 때 까지의 거리
- 다수의 실험을 통한 결과값을 점으로 찍어 산점도(Scatter plot)로 나타냄.





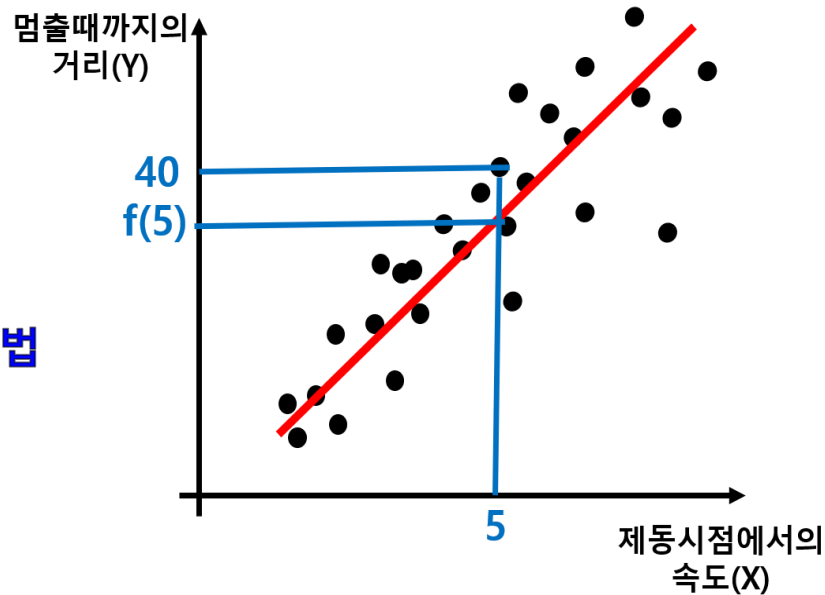
# 단순, 선형회귀분석

- 모든 점을 동시에 지나가는 직선은 존재하지 않음.
- 하지만 **변수 간의 관계를 잘 나타낼 수 있는 직선은 존재**
- 실제 값과 예측한 값의 차이가 작은 직선이 분석을 잘 한 직선
- 이 직선을 찾기 위해서 우리는 **최소제곱법**을 활용

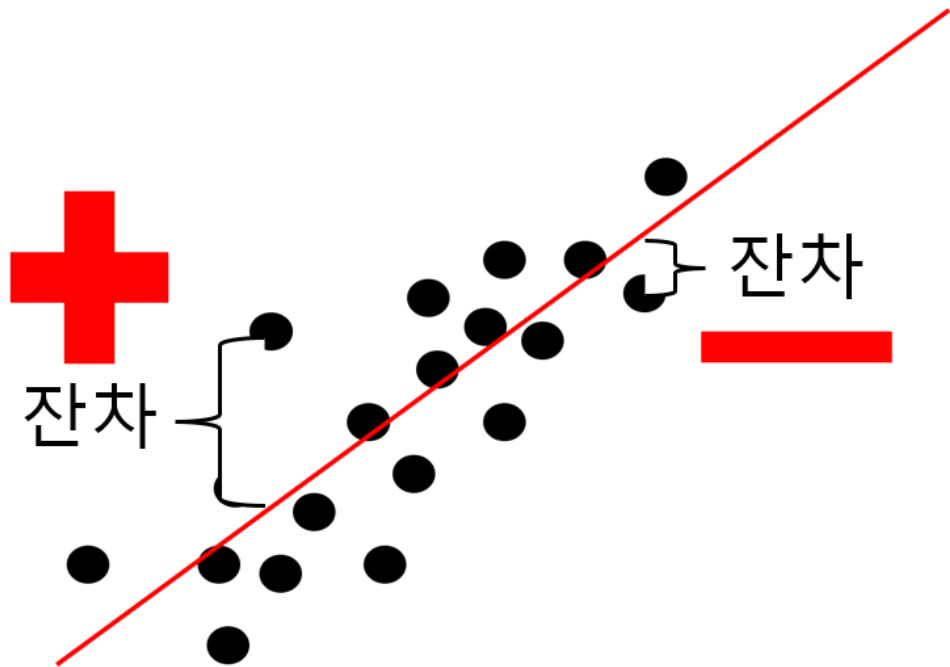


# 최소제곱법(Method of least squares)

- 임의의 점을 예시로 들어보자 : 직선식을  $f(x)=ax+b$ 라고 가정
- 파란색 선을 따라가보면 실제값은 40이며 예측값은  $f(5)$
- 값의 차이에 절댓값을 씌움 (예:  $|40-f(5)|$  )
- 절댓값을 씌운 값을 '**잔차(residual)**'
- 모든 데이터의 잔차를 구해서 그 값이 가장 작은 직선이 정확하게 예측한 직선
- **잔차의 합이 최소인 직선을 구함 → 최소제곱법**



# 최소제곱법(Method of least squares)



# 최소제곱법(Method of least squares)

- 단순히 절댓값만 씌워져 있는 경우 (편)미분을 적용하기 어려움
- 따라서 잔차의 합이 최소인 것을 구하지 않고 잔차의 제곱의 합이 최소인 직선을 구하는 것이 편리
- 이를 **최소제곱법**
- $(|40-f(x_1)|)^2 + (|40-f(x_2)|)^2 + \dots + \dots$
- 단순히 절댓값만 씌워져 있는 경우 (편)미분을 적용하기 어려움
- 따라서 잔차의 합이 최소인 것을 구하지 않고 잔차의 제곱의 합이 최소인 직선을 구하는 것이 편리
- 이를 **최소제곱법**
- $(|40-f(x_1)|)^2 + (|40-f(x_2)|)^2 + \dots + \dots$

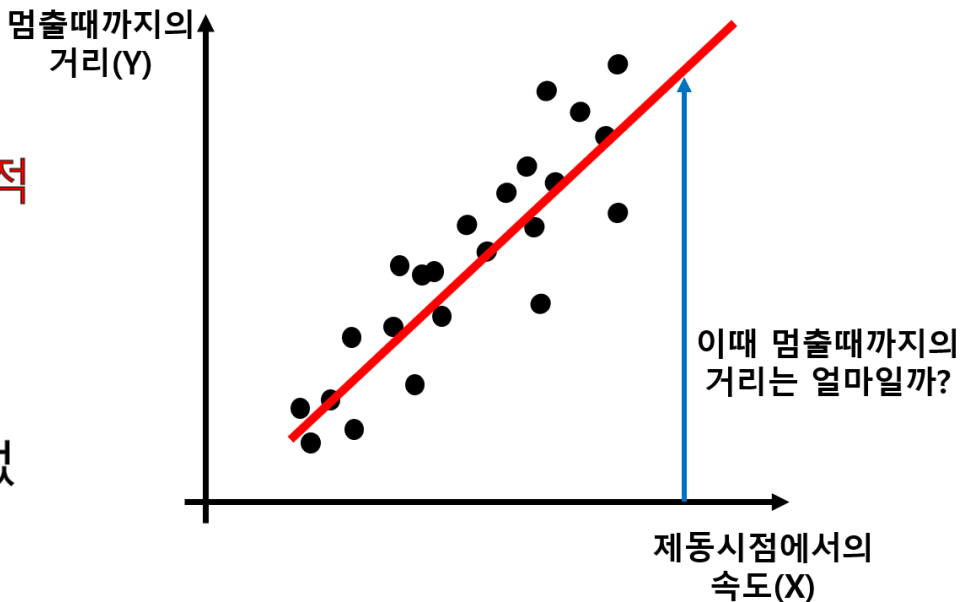
# 단순, 선형회귀분석

- 회귀식  $y=ax+b$  의 장점

- 1) 변수간의 인과관계를 정확하게 수학적  
으로 표현

- 2) 예측 가능성.

산점도에 나타나 있지 않은 값은 알 수 없  
으나 대략적으로 추측 가능

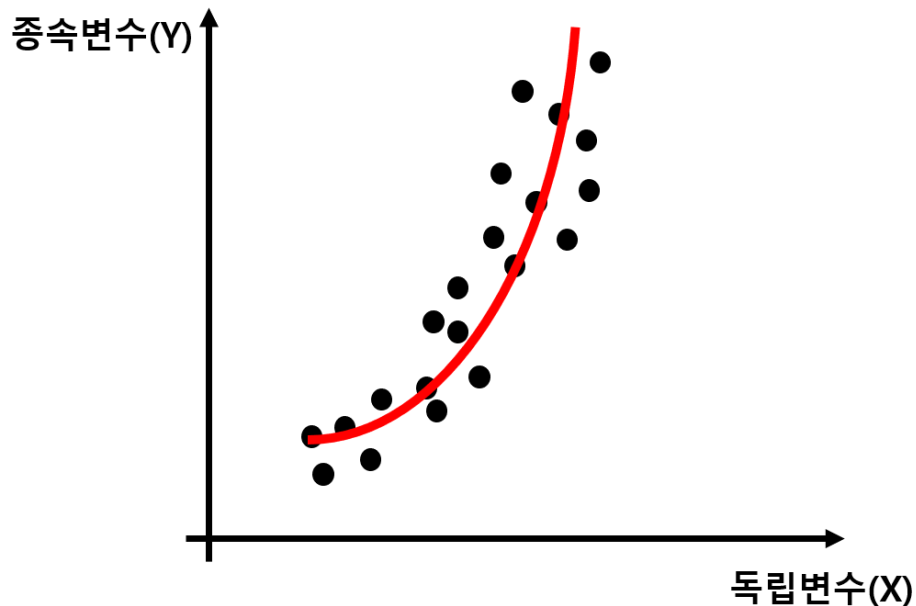


# 단순, 선형회귀분석

- 오렌지3에서 (독립)변수 → **Features**
- 오렌지3에서 (종속)변수 → **Target**
- 주어진 데이터의 feature값으로 target을 예측
- 이를 위해서는 직선형태의 회귀선을 구하는 것이 목적이며 선형회귀식은  $y=ax+b$  의 형태. a는 기울기, b는 절편
- 회귀식을 구하기 위해 사용되는 방법으로 **최소제곱법**
- 최소제곱법으로 구해진 직선이 우리가 원하는 최종의 회귀분석의 식임.

# 비선형회귀분석

- **비선형 회귀분석**이란 직선이 아닌 그래프로 두 변수 간의 관계를 분석하는 것임.
- 선형회귀분석에 비해 예측이 어려움



# 다중회귀분석

- MEDV는 해당 지역의 주택 가격의 중앙값을 나타냄. 값이 클수록 비싼 값의 주택이 많이 있음을 추정

|    | MEDV | CRIM    | ZN   | INDUS | CHAS | NOX    | RM    | AGE   | DIS    | RAD | TAX | PTRATIO | B      | LSTAT |
|----|------|---------|------|-------|------|--------|-------|-------|--------|-----|-----|---------|--------|-------|
| 1  | 24.0 | 0.00632 | 18.0 | 2.31  | 0    | 0.5380 | 6.575 | 65.2  | 4.0900 | 1   | 296 | 15.3    | 396.90 | 4.98  |
| 2  | 21.6 | 0.02731 | 0.0  | 7.07  | 0    | 0.4690 | 6.421 | 78.9  | 4.9671 | 2   | 242 | 17.8    | 396.90 | 9.14  |
| 3  | 34.7 | 0.02729 | 0.0  | 7.07  | 0    | 0.4690 | 7.185 | 61.1  | 4.9671 | 2   | 242 | 17.8    | 392.83 | 4.03  |
| 4  | 33.4 | 0.03237 | 0.0  | 2.18  | 0    | 0.4580 | 6.998 | 45.8  | 6.0622 | 3   | 222 | 18.7    | 394.63 | 2.94  |
| 5  | 36.2 | 0.06905 | 0.0  | 2.18  | 0    | 0.4580 | 7.147 | 54.2  | 6.0622 | 3   | 222 | 18.7    | 396.90 | 5.33  |
| 6  | 28.7 | 0.02985 | 0.0  | 2.18  | 0    | 0.4580 | 6.430 | 58.7  | 6.0622 | 3   | 222 | 18.7    | 394.12 | 5.21  |
| 7  | 22.9 | 0.08829 | 12.5 | 7.87  | 0    | 0.5240 | 6.012 | 66.6  | 5.5605 | 5   | 311 | 15.2    | 395.60 | 12.43 |
| 8  | 27.1 | 0.14455 | 12.5 | 7.87  | 0    | 0.5240 | 6.172 | 96.1  | 5.9505 | 5   | 311 | 15.2    | 396.90 | 19.15 |
| 9  | 16.5 | 0.21124 | 12.5 | 7.87  | 0    | 0.5240 | 5.631 | 100.0 | 6.0821 | 5   | 311 | 15.2    | 386.63 | 29.93 |
| 10 | 18.9 | 0.17004 | 12.5 | 7.87  | 0    | 0.5240 | 6.004 | 85.9  | 6.5921 | 5   | 311 | 15.2    | 386.71 | 17.10 |
| 11 | 15.0 | 0.22489 | 12.5 | 7.87  | 0    | 0.5240 | 6.377 | 94.3  | 6.3467 | 5   | 311 | 15.2    | 392.52 | 20.45 |
| 12 | 18.9 | 0.11747 | 12.5 | 7.87  | 0    | 0.5240 | 6.009 | 82.9  | 6.2267 | 5   | 311 | 15.2    | 396.90 | 13.27 |
| 13 | 21.7 | 0.09378 | 12.5 | 7.87  | 0    | 0.5240 | 5.889 | 39.0  | 5.4509 | 5   | 311 | 15.2    | 390.50 | 15.71 |
| 14 | 20.4 | 0.62976 | 0.0  | 8.14  | 0    | 0.5380 | 5.949 | 61.8  | 4.7075 | 4   | 307 | 21.0    | 396.90 | 8.26  |
| 15 | 18.2 | 0.63796 | 0.0  | 8.14  | 0    | 0.5380 | 6.096 | 84.5  | 4.4619 | 4   | 307 | 21.0    | 380.02 | 10.26 |
| 16 | 19.9 | 0.62739 | 0.0  | 8.14  | 0    | 0.5380 | 5.834 | 56.5  | 4.4986 | 4   | 307 | 21.0    | 395.62 | 8.47  |
| 17 | 23.1 | 1.05393 | 0.0  | 8.14  | 0    | 0.5380 | 5.935 | 29.3  | 4.4986 | 4   | 307 | 21.0    | 386.85 | 6.58  |
| 18 | 17.5 | 0.78420 | 0.0  | 8.14  | 0    | 0.5380 | 5.990 | 81.7  | 4.2579 | 4   | 307 | 21.0    | 386.75 | 14.67 |
| 19 | 20.2 | 0.80271 | 0.0  | 8.14  | 0    | 0.5380 | 5.456 | 36.6  | 3.7965 | 4   | 307 | 21.0    | 288.99 | 11.69 |
| 20 | 18.2 | 0.72580 | 0.0  | 8.14  | 0    | 0.5380 | 5.727 | 69.5  | 3.7965 | 4   | 307 | 21.0    | 390.95 | 11.28 |
| 21 | 13.6 | 1.25179 | 0.0  | 8.14  | 0    | 0.5380 | 5.570 | 98.1  | 3.7979 | 4   | 307 | 21.0    | 376.57 | 21.02 |
| 22 | 19.6 | 0.85204 | 0.0  | 8.14  | 0    | 0.5380 | 5.965 | 89.2  | 4.0123 | 4   | 307 | 21.0    | 392.53 | 13.83 |
| 23 | 15.2 | 1.23247 | 0.0  | 8.14  | 0    | 0.5380 | 6.142 | 91.7  | 3.9769 | 4   | 307 | 21.0    | 396.90 | 18.72 |
| 24 | 14.5 | 0.98843 | 0.0  | 8.14  | 0    | 0.5380 | 5.813 | 100.0 | 4.0952 | 4   | 307 | 21.0    | 394.54 | 19.88 |
| 25 | 15.6 | 0.75026 | 0.0  | 8.14  | 0    | 0.5380 | 5.924 | 94.1  | 4.3996 | 4   | 307 | 21.0    | 394.33 | 16.30 |
| 26 | 13.9 | 0.84054 | 0.0  | 8.14  | 0    | 0.5380 | 5.599 | 85.7  | 4.4546 | 4   | 307 | 21.0    | 303.42 | 16.51 |
| 27 | 16.6 | 0.67191 | 0.0  | 8.14  | 0    | 0.5380 | 5.813 | 90.3  | 4.6820 | 4   | 307 | 21.0    | 376.88 | 14.81 |
| 28 | 14.8 | 0.95577 | 0.0  | 8.14  | 0    | 0.5380 | 6.047 | 88.8  | 4.4534 | 4   | 307 | 21.0    | 306.38 | 17.28 |
| 29 | 18.4 | 0.77299 | 0.0  | 8.14  | 0    | 0.5380 | 6.495 | 94.4  | 4.4547 | 4   | 307 | 21.0    | 387.94 | 12.80 |
| 30 | 21.0 | 1.00245 | 0.0  | 8.14  | 0    | 0.5380 | 6.674 | 87.3  | 4.2390 | 4   | 307 | 21.0    | 380.23 | 11.98 |
| 31 | 12.7 | 1.13081 | 0.0  | 8.14  | 0    | 0.5380 | 5.713 | 94.1  | 4.2330 | 4   | 307 | 21.0    | 360.17 | 22.60 |
| 32 | 14.5 | 1.35472 | 0.0  | 8.14  | 0    | 0.5380 | 6.072 | 100.0 | 4.1750 | 4   | 307 | 21.0    | 376.73 | 13.04 |
| 33 | 13.2 | 1.38799 | 0.0  | 8.14  | 0    | 0.5380 | 5.950 | 82.0  | 3.9900 | 4   | 307 | 21.0    | 232.60 | 27.71 |
| 34 | 13.1 | 1.15172 | 0.0  | 8.14  | 0    | 0.5380 | 5.701 | 95.0  | 3.7872 | 4   | 307 | 21.0    | 358.77 | 18.35 |
| 35 | 13.5 | 1.61282 | 0.0  | 8.14  | 0    | 0.5380 | 6.096 | 96.9  | 3.7598 | 4   | 307 | 21.0    | 248.31 | 20.34 |
| 36 | 18.9 | 0.06417 | 0.0  | 5.96  | 0    | 0.4990 | 5.933 | 68.2  | 3.3603 | 5   | 279 | 19.2    | 396.90 | 9.68  |
| 37 | 20.0 | 0.09744 | 0.0  | 5.96  | 0    | 0.4990 | 5.841 | 61.4  | 3.3779 | 5   | 279 | 19.2    | 377.56 | 11.41 |
| 38 | 21.0 | 0.08014 | 0.0  | 5.96  | 0    | 0.4990 | 5.850 | 41.5  | 3.9342 | 5   | 279 | 19.2    | 396.90 | 8.77  |
| 39 | 24.7 | 0.17505 | 0.0  | 5.96  | 0    | 0.4990 | 5.966 | 30.2  | 3.8473 | 5   | 279 | 19.2    | 393.43 | 10.13 |
| 40 | 30.8 | 0.02763 | 75.0 | 2.95  | 0    | 0.4280 | 6.595 | 21.8  | 5.4011 | 3   | 252 | 18.3    | 395.63 | 4.32  |



# 다중회귀분석

- 다중회귀분석은 종속변수에 영향을 주는 독립변수가 여러 개 작용하는 것을 뜻함.
- 예를 들면,  $y = ax_1 + bx_2 + cx_3 + \dots + \dots$  와 같이 여러 개의 x변수가 y의 값에 영향을 주는 것
- 다음장에 제시될 표는 보스톤 지역의 집값을 보여주는 데이터
- 각각의 행 하나는 그 지역 하나를 의미
- 각각의 열은 지역의 특성
- 다양한 열들이 영향을 주며 MEDV, 즉 집값을 결정하게 됨

# 보스턴시 주택가격 데이터

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>



# 보스턴시 주택가격 데이터

|               |  |
|---------------|--|
| [01] CRIM     | 자치시(town) 별 1인당 범죄율                            |
| [02] ZN       | 25,000 평방피트를 초과하는 거주지역의 비율                     |
| [03] INDUS    | 비소매상업지역이 점유하고 있는 토지의 비율                        |
| [04] CHAS     | 찰스강에 대한 더미변수(강의 경계에 위치한 경우는 1, 아니면 0)          |
| [05] NOX      | 10ppm 당 농축 일산화질소                               |
| [06] RM       | 주택 1가구당 평균 방의 개수                               |
| [07] AGE      | 1940년 이전에 건축된 소유주택의 비율                         |
| [08] DIS      | 5개의 보스턴 직업센터까지의 접근성 지수                         |
| [09] RAD      | 방사형 도로까지의 접근성 지수                               |
| [10] TAX      | 10,000 달러 당 재산세율                               |
| [11] PTRATIO  | 자치시(town)별 학생/교사 비율                            |
| [12] B        | $1000(Bk - 0.63)^2$ , 여기서 Bk는 자치시별 흑인의 비율을 말함. |
| [13] LSTAT    | 모집단의 하위계층의 비율(%)                               |
| [14] CAT.MDEV | MDEV가 \$30,000을 넘는지에 대한 변수 (넘는 경우 1, 아닌 경우 0)  |
| [15] MEDV     | 본인 소유의 주택가격(중앙값) (단위: \$1,000)                 |

질문 있나요?

hsryu13@hongik.ac.kr

