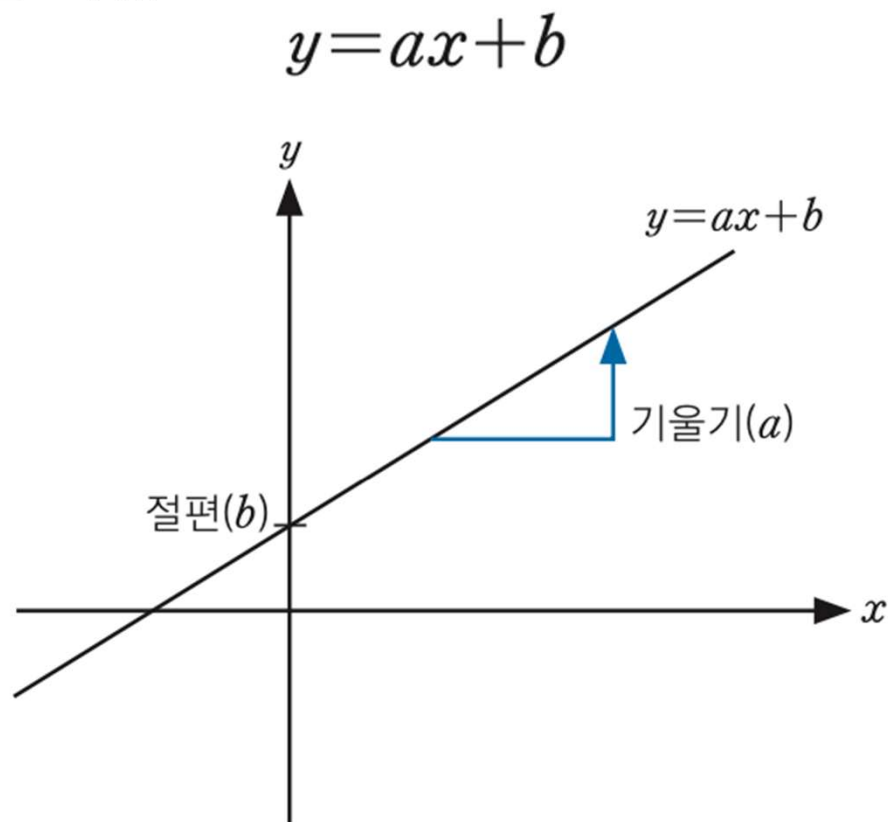


03 머신러닝의 기초를 다집니다

- 수치 예측

03-1 선형 회귀에 대해 알아보고 데이터를 준비합니다

1차 함수로 이해하는 선형 회귀



선형 회귀는 기울기와 절편을 찾아줍니다

학교에서 배울 때는 기울기와 절편보다 x , y 값에 관심을 기울입니다.

1차 함수 문제 기울기가 7이고 절편이 4인 1차 함수 $y=7x+4$ 가 있습니다. x 가 10이면 y 는 얼마인가요?

- ① 74
- ② 72
- ③ 71

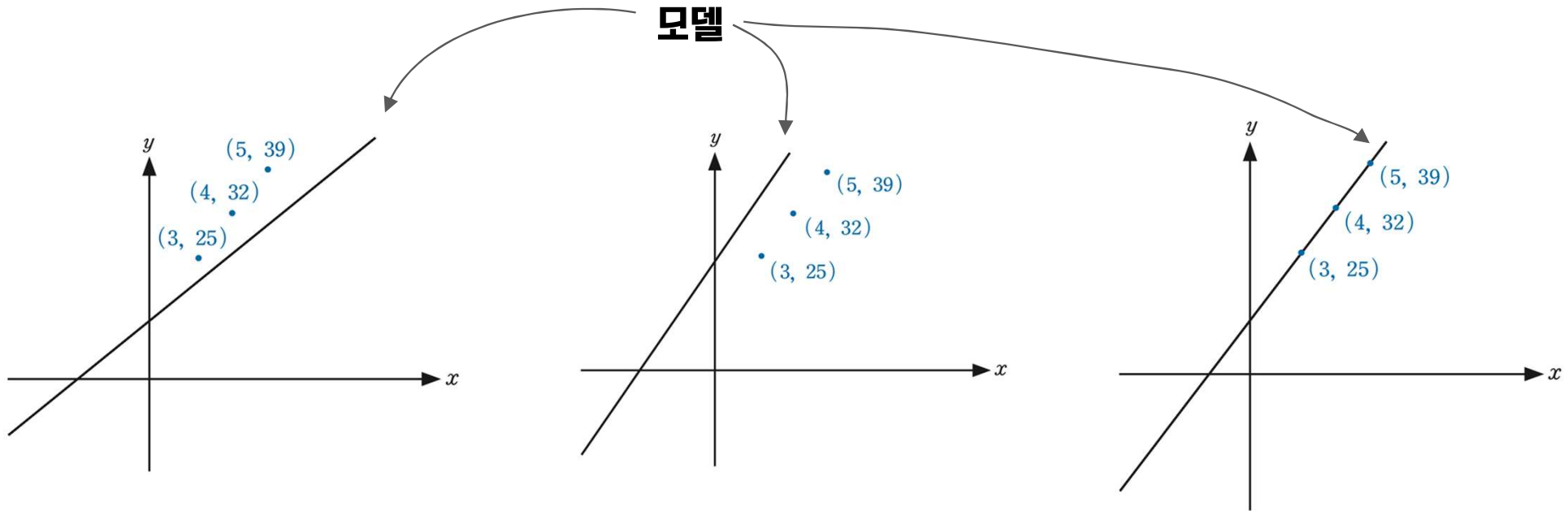
머신러닝은 x , y 가 주어질 때 기울기와 절편을 구합니다.

선형 회귀 문제 x 가 3일 때 y 는 25, x 가 4일 때 y 는 32, x 가 5일 때 y 는 39라면 기울기와 절편의 값으로 적절
한 것은 무엇인가요?

- ① 기울기는 6, 절편은 4
- ② 기울기는 7, 절편은 5
- ③ 기울기는 7, 절편은 4

The diagram shows the equation $1.5 \times x + 0.1 = y$. Arrows point from labels to parts of the equation: '가중치' (weight) points to 1.5, '입력' (input) points to x , '절편' (intercept) points to 0.1, and '타겟' (target) points to y .

그래프를 통해 선형 회귀의 문제 해결 과정을 이해합니다



최적의 모델
 $x = 6$ 일 때는? (46)

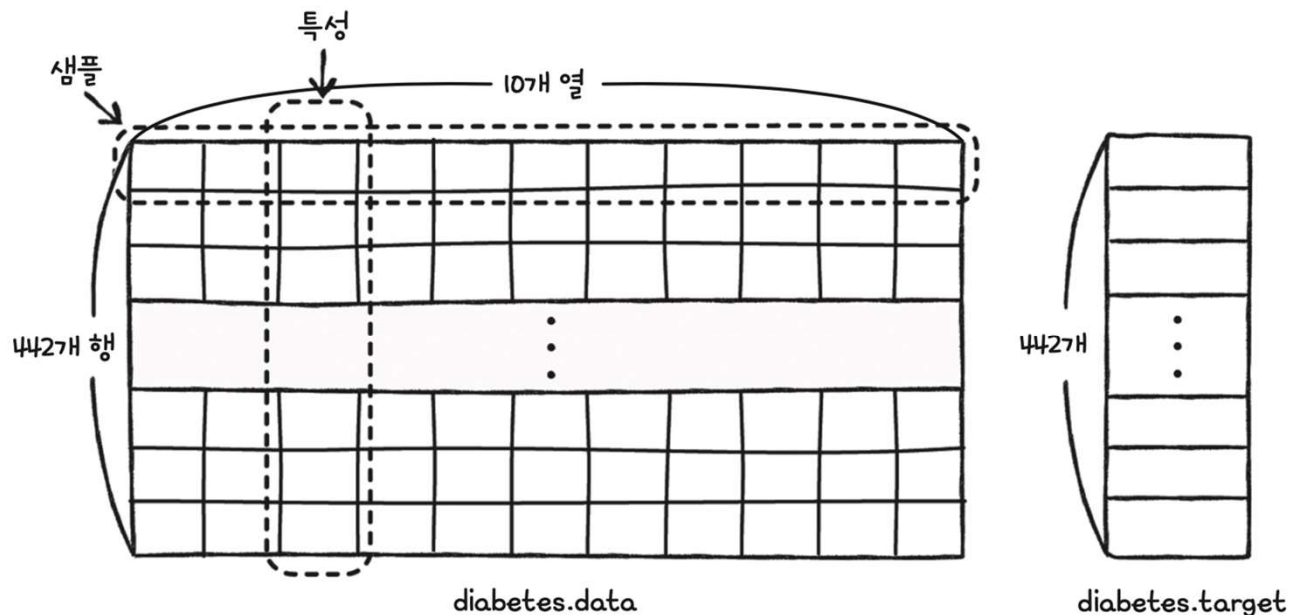
문제 해결을 위해 당뇨병 환자의 데이터 준비하기

```
from sklearn.datasets import load_diabetes
diabetes = load_diabetes( )
```

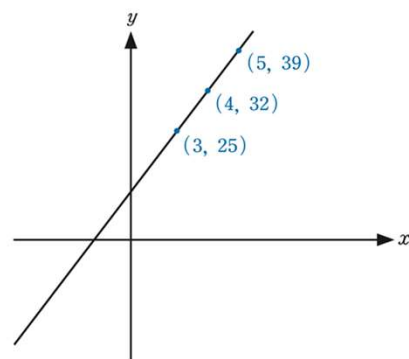
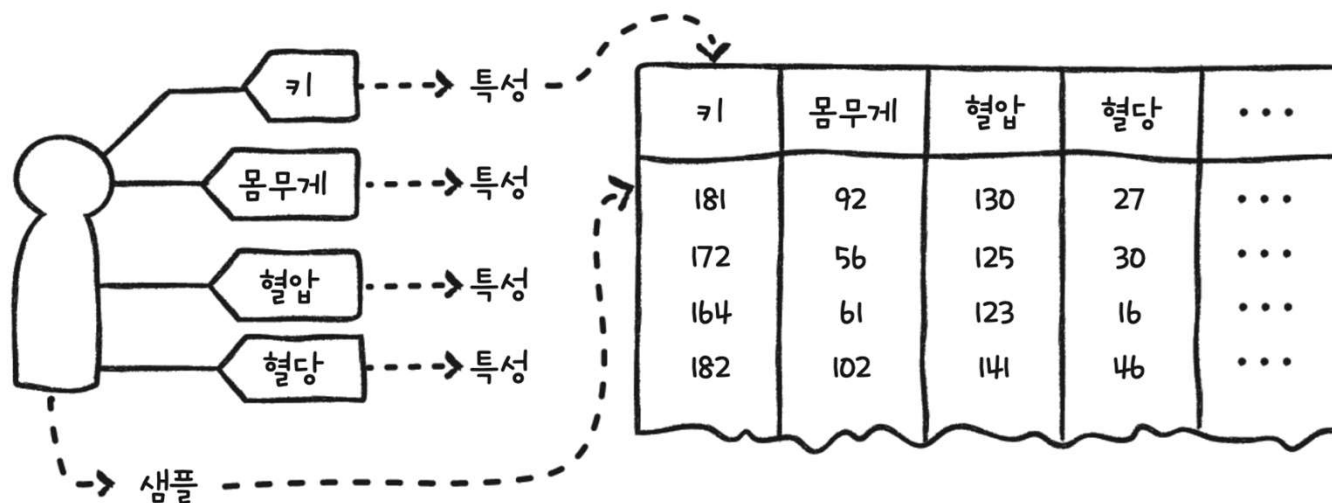
```
print(diabetes.data.shape, diabetes.target.shape)
(442, 10) (442,)
```

넘파이 배열

Bunch 클래스



샘플과 특성



여기에서는 특성 하나만 사용했습니다.

특성 2개를 사용하면? 3개를 사용하면?

입력 데이터와 타겟 데이터 자세히 보기

```
diabetes.data[0:3]
```

```
array([[ 0.03807591,  0.05068012,  0.06169621,  0.02187235, -0.0442235 ,  
        -0.03482076, -0.04340085, -0.00259226,  0.01990842, -0.01764613],  
       [-0.00188202, -0.04464164, -0.05147406, -0.02632783, -0.00844872,  
        -0.01916334,  0.07441156, -0.03949338, -0.06832974, -0.09220405],  
       [ 0.08529891,  0.05068012,  0.04445121, -0.00567061, -0.04559945,  
        -0.03419447, -0.03235593, -0.00259226,  0.00286377, -0.02593034]])
```

네 번째 특성의 값입니다.

첫 번째 샘플입니다.

```
diabetes.target[:3]
```

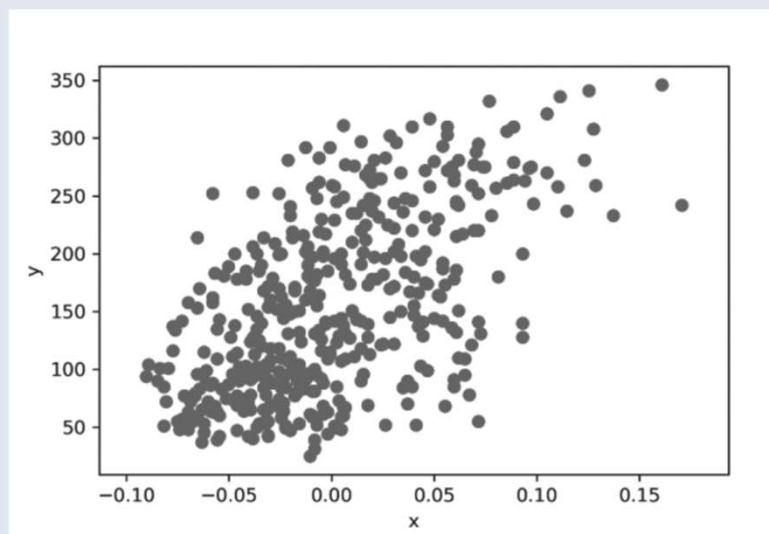
```
array([151., 75., 141.])
```

도메인 지식이 필요합니다

실전에서는 지도 학습 데이터를 만드는 데
많은 노력을 기울입니다

당뇨병 환자 데이터 시각화하기

```
import matplotlib.pyplot as plt
plt.scatter(diabetes.data[:, 2], diabetes.target)
plt.xlabel('x')
plt.ylabel('y')
plt.show( )
```



이후 코드를 간단하게 쓰기 위해

```
x = diabetes.data[:, 2]
y = diabetes.target
```