

< Feature scaling >

- objective: gradient descent to run much faster

* example

$$\hat{\text{price}} = w_1 x_1 + w_2 x_2 + b$$

\downarrow \downarrow
 size of house number of bedrooms

ex)

| | |
|--|--|
| x_1 : size (feet ²) range: 300 ~ 2,000 <i>relatively large</i> | x_2 : number of bedrooms range: 0 ~ 5 |
|--|--|

ex)

Training example house: $x_1 = 2000$, $x_2 = 5$, price = \$500k
(y)

if

$$w_1 = 50, w_2 = 0.1, b = 50$$

$$\hat{\text{price}} = \underbrace{50 \times 2000}_{100,000k} + \underbrace{0.1 \times 5}_{0.5k} + \underbrace{50}_{50k}$$

$$= \$100,050.5k$$

↓
not good set of parameter choice
for w_1, w_2

if

$$w_1 = 0.1, w_2 = 50, b = 50$$

$$\hat{\text{price}} = 0.1 \times 2000k + 50 \times 5 + 50$$

$$= \$500k$$

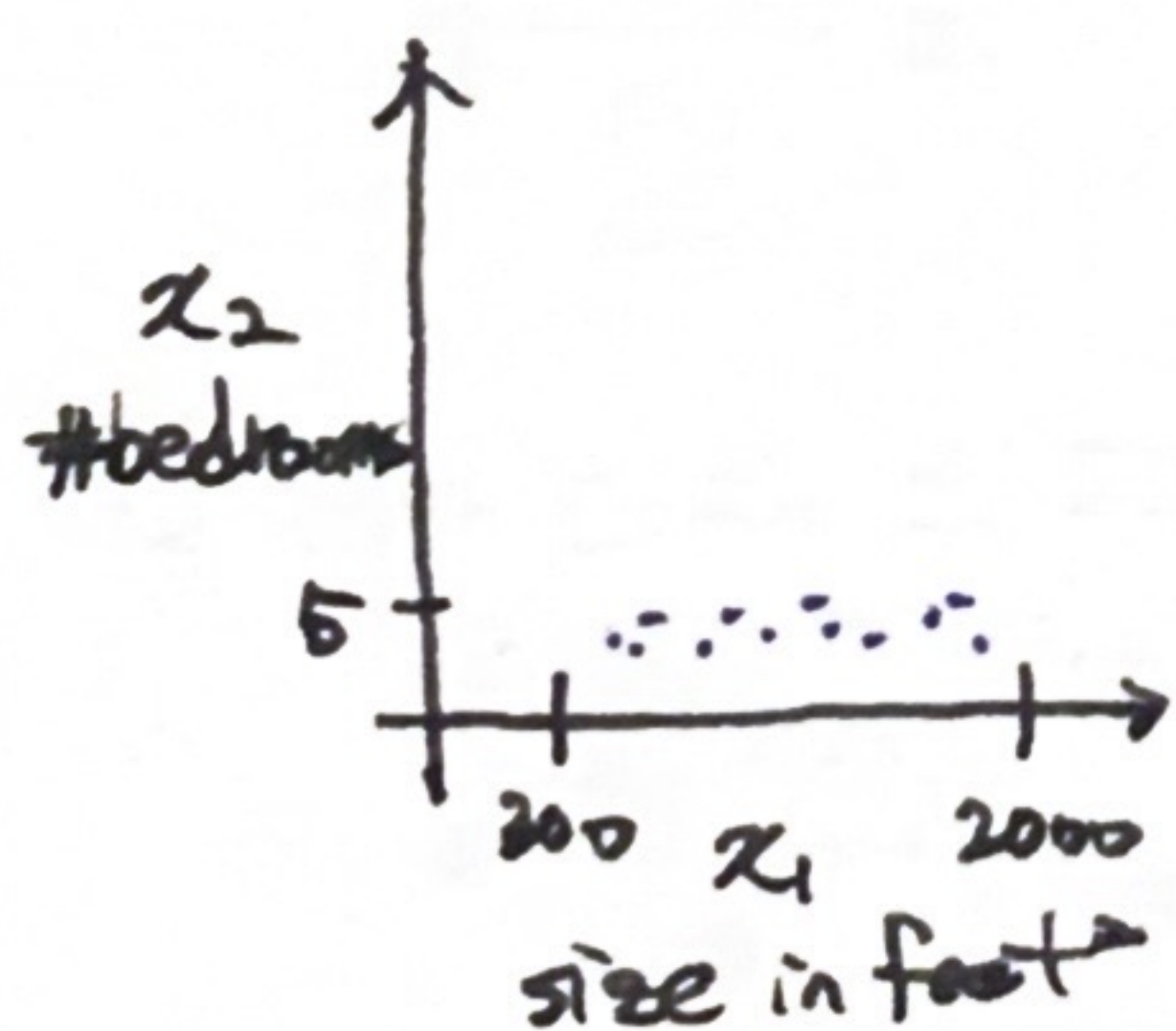
↓
reasonable estimation of w_1, w_2

* feature value의 range가 다른 feature value의 range scale보다 크다면

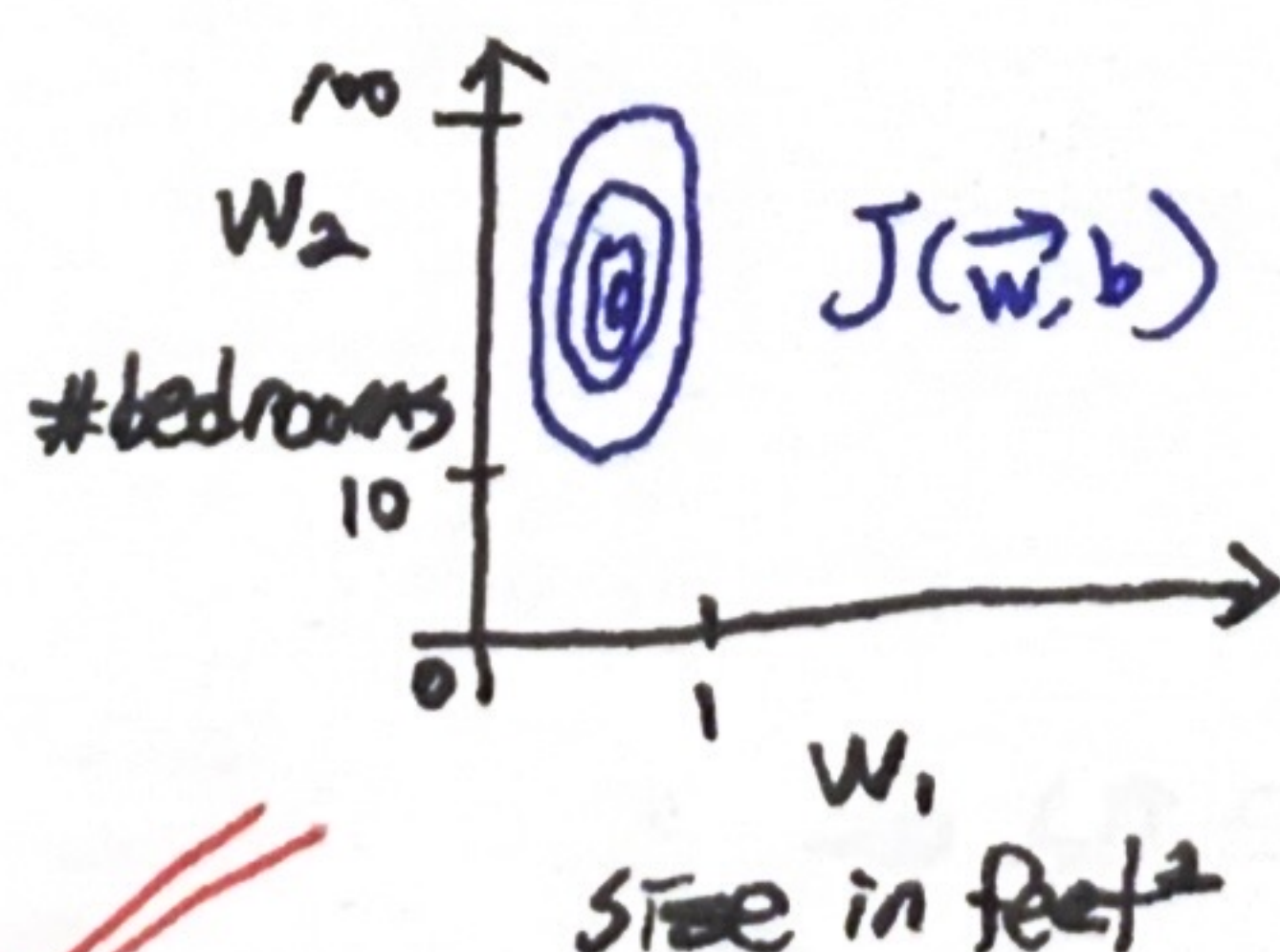
parameter value는 상대적으로 작은 값으로 설정하는 것이 일반적

< Feature size and Parameter size >

(Features) scatterplot



(Parameters) Contour Plot



⇒ "It takes too long time
for gradient descent to converge"

This is because...

small change to w_1 can have large impact on the estimated price and cost J

(because w_1 is multiplied by a large number (size in feet²)
300 ~ 2000)

it takes much larger change in w_2 in order to change the predictions much

because w_2 is multiplied by a small number (#bedrooms)
0 ~ 5