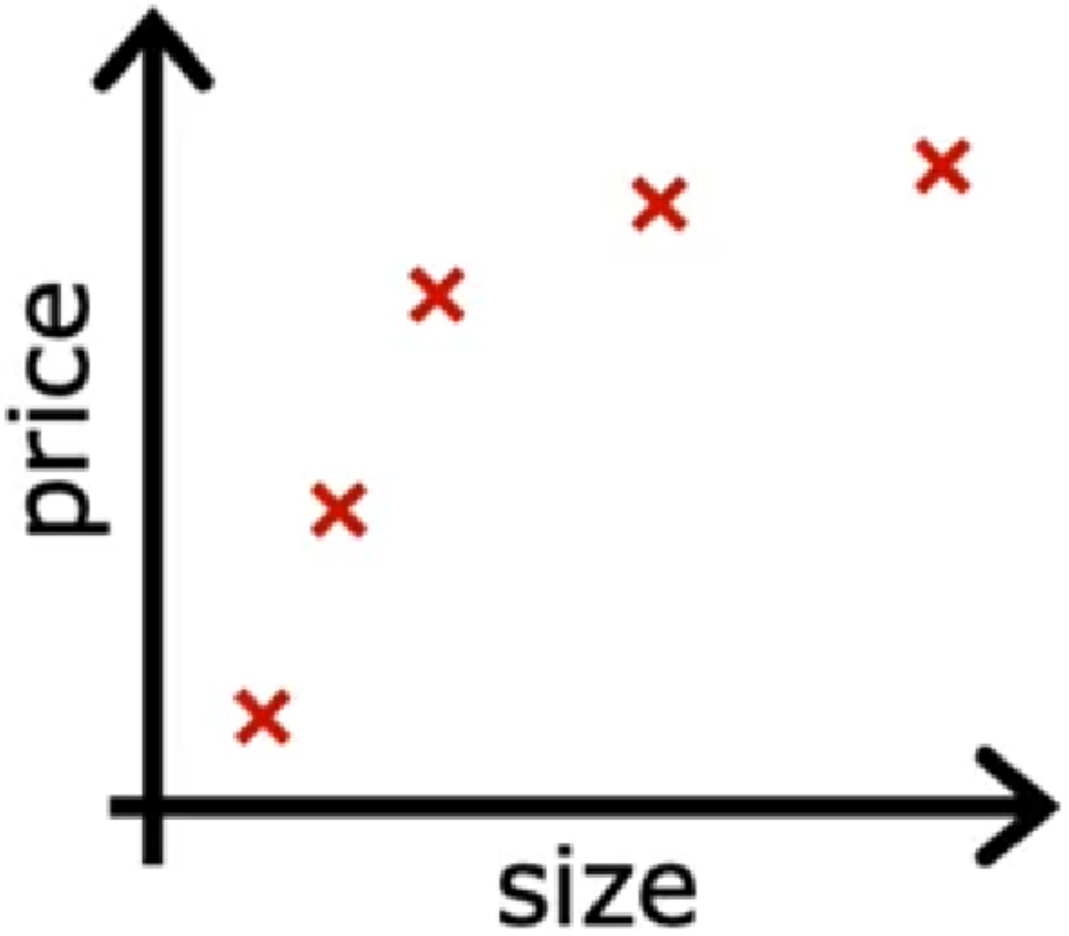
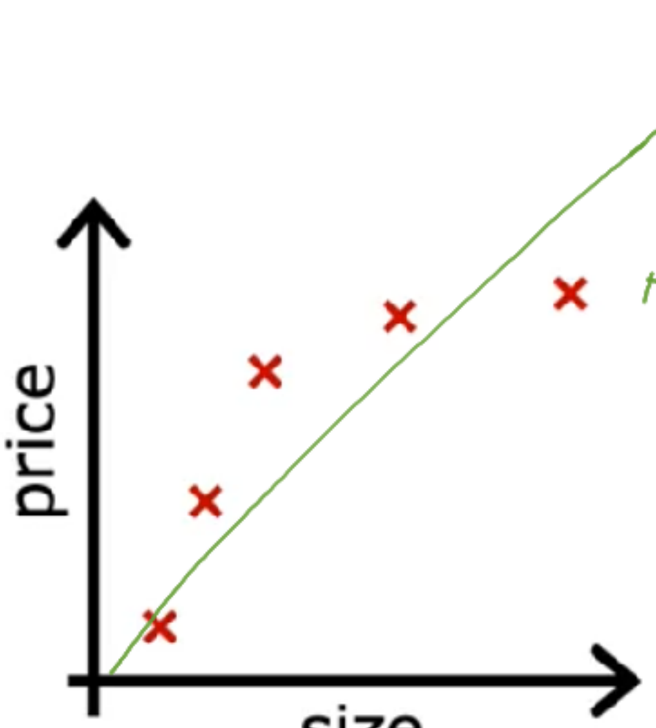
Regularization to Reducw Overfitting

over/under fitting

House Price

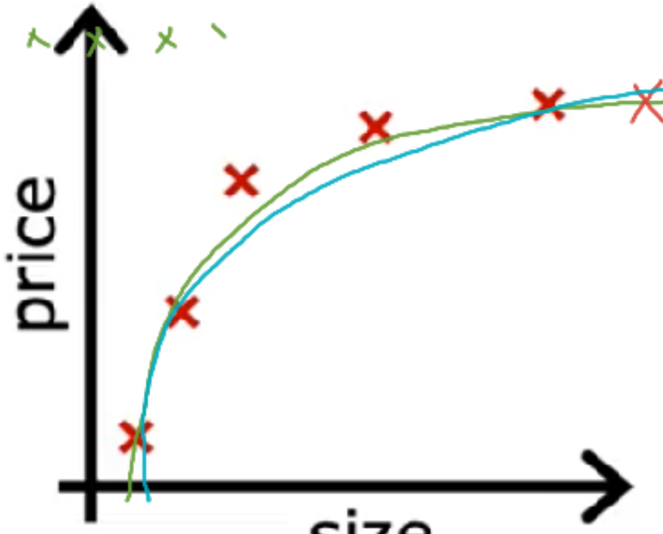


Overfitting > Regression



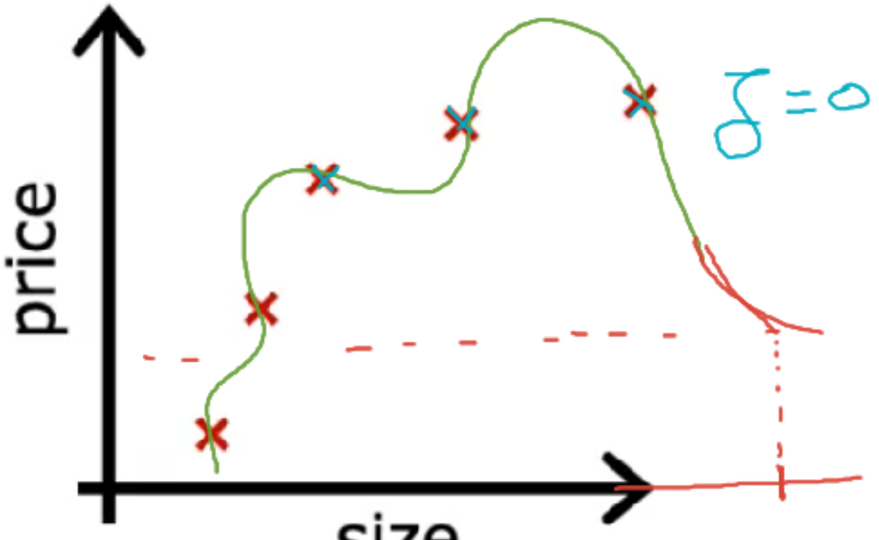
$w_1x + b$

underfitting



$w_1x + w_2x^2 + b$

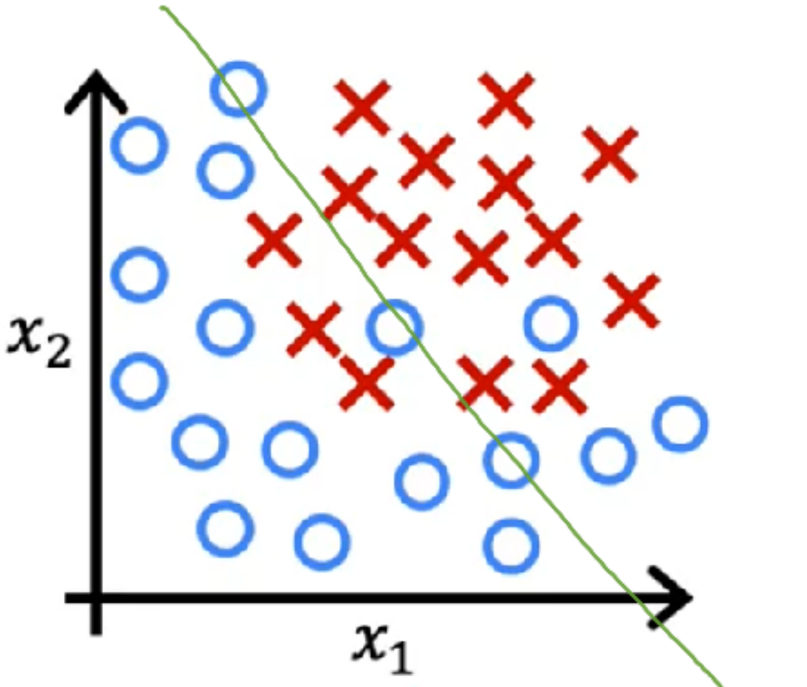
just right



$w_1x + w_2x^2 + w_3x^3 + w_4x^4 + b$

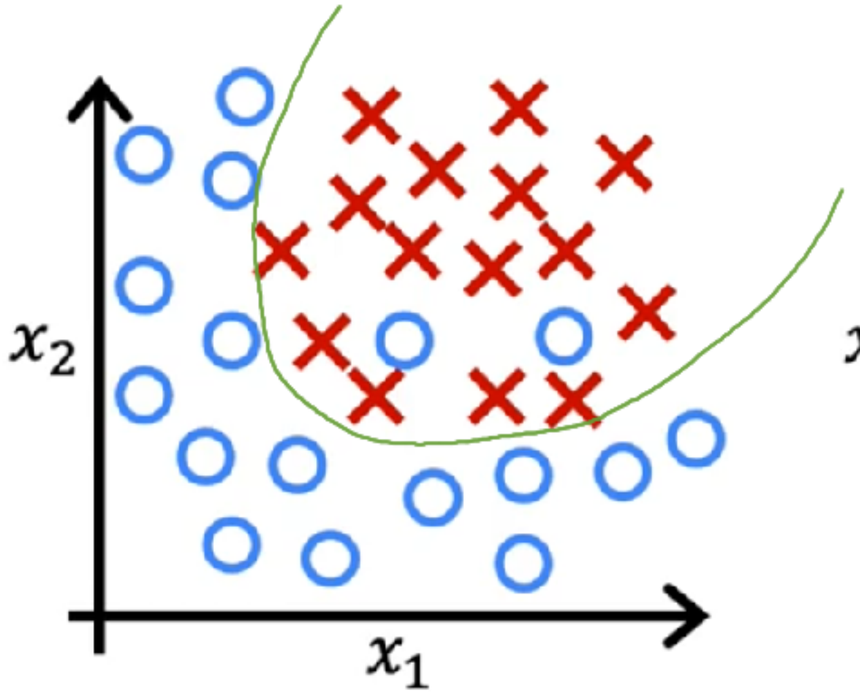
overfitting

Overfitting > Classification



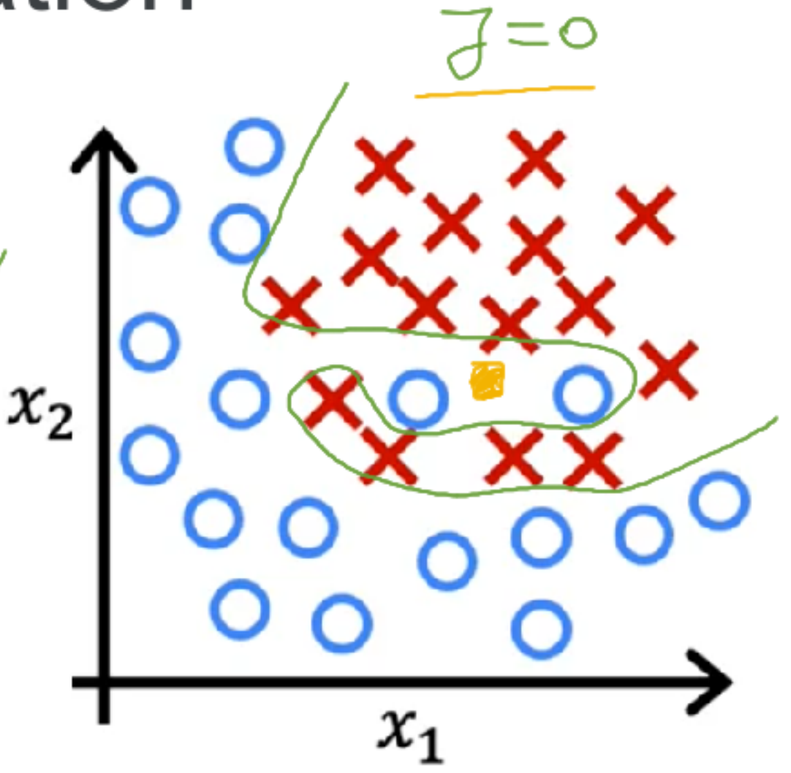
$$z = w_1 x_1 + w_2 x_2 + b$$

under



$$z = w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + b$$

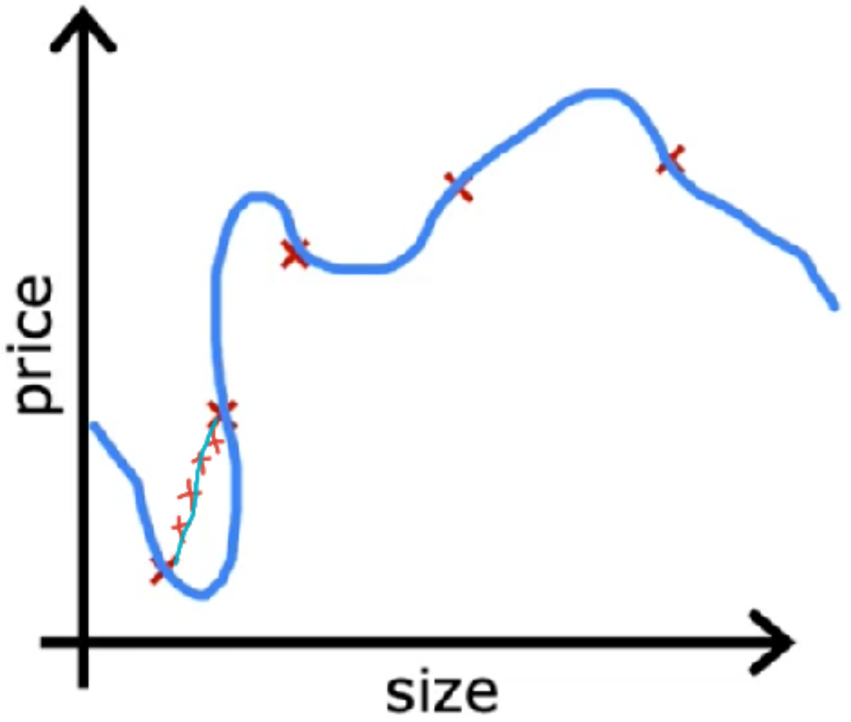
just right



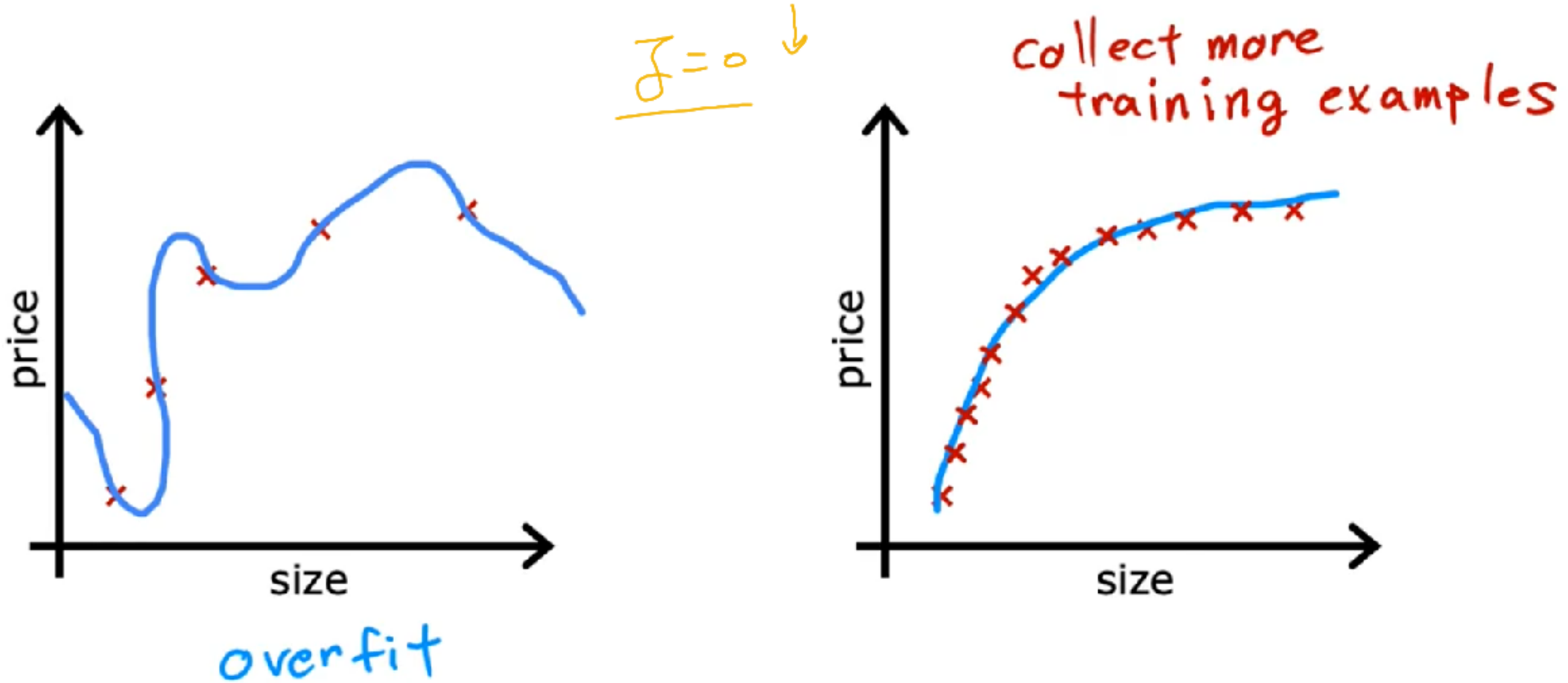
$$z = w_1 x_1 + w_2 x_2 + w_3 x_1^2 x_2 + w_4 x_1^2 x_2^2 + w_5 x_1^2 x_2^3 + w_6 x_1^3 x_2 + \cdots + b$$

overfit

Collect more training examples



Collect more training examples



Select features to include/exclude

size	bedrooms	floors	age	avg income	...	distance to coffee shop	price
x_1	x_2	x_3	x_4	x_5		x_{100}	y

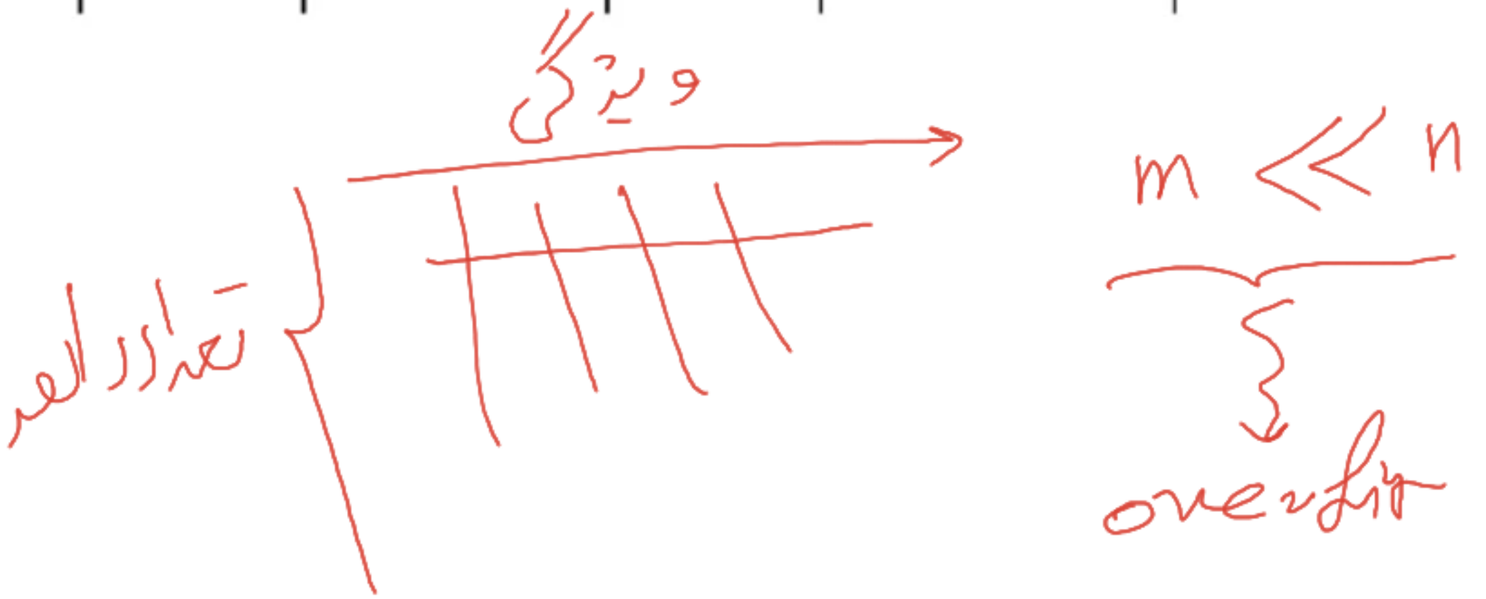
all features

+

insufficient data

↓

overfit



Select features to include/exclude

all features

+

insufficient data

↓

overfit

selected features

size
bedrooms
age

just right

feature selection

Select features to include/exclude

all features

\+

insufficient data

↓

overfit

selected features

size
bedrooms
age
just right

feature selection

course 2

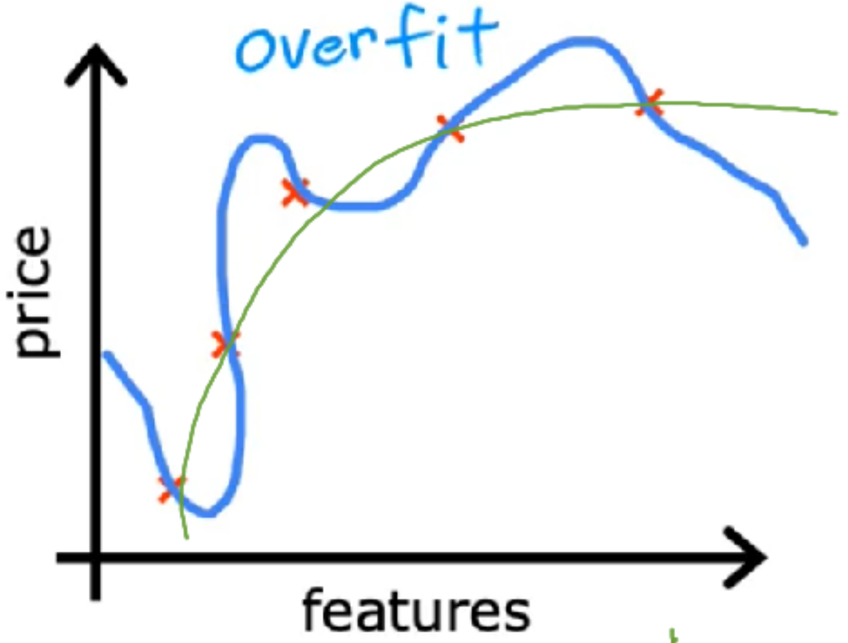
disadvantage

↓

useful features could be lost

Regularization

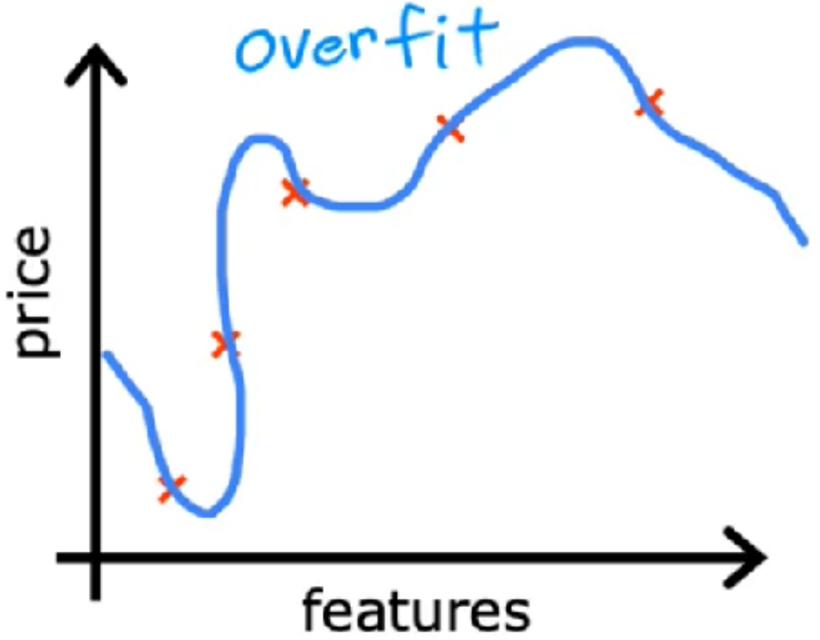
Reduce the size of parameters w_j



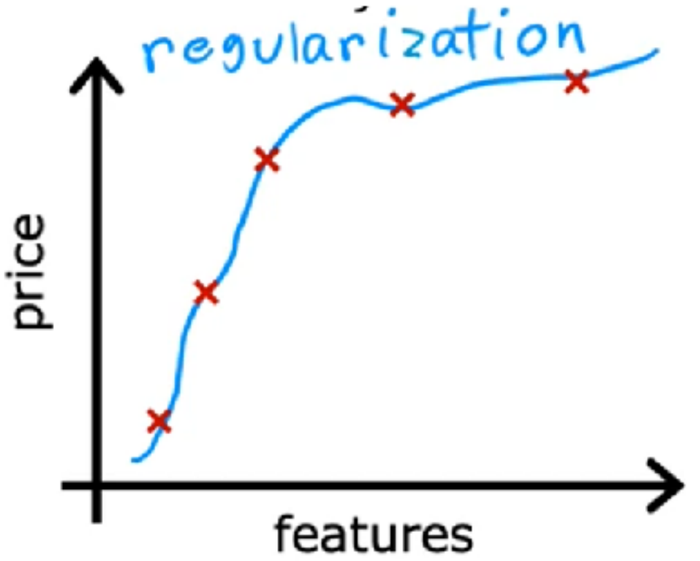
$$f(x) = 28x - 385x^2 + 39x^3 - 174x^4 + 100$$

Regularization

Reduce the size of parameters w_j



$$f(x) = 28x - 385x^2 + 39x^3 - 174x^4 + 100$$



$$f(x) = 13x - 0.23x^2 + 0.000014x^3 - 0.0001x^4 + 10$$

small values for w_j

Addressing overfitting

Options

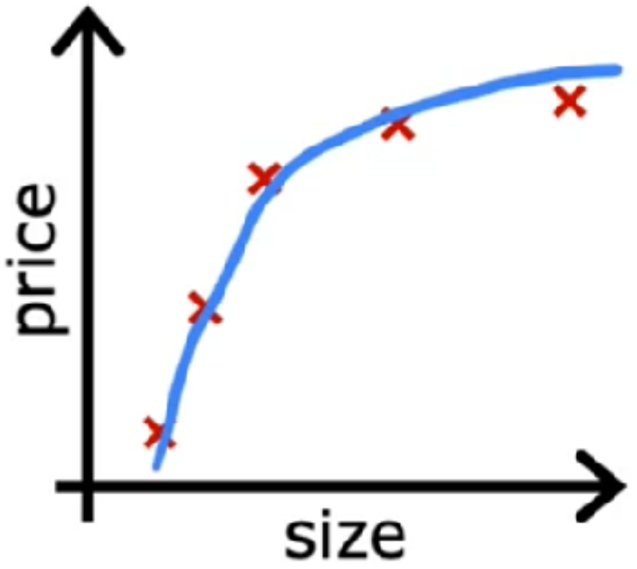
1. Collect more data

2. Select features
   - Feature selection in course 2

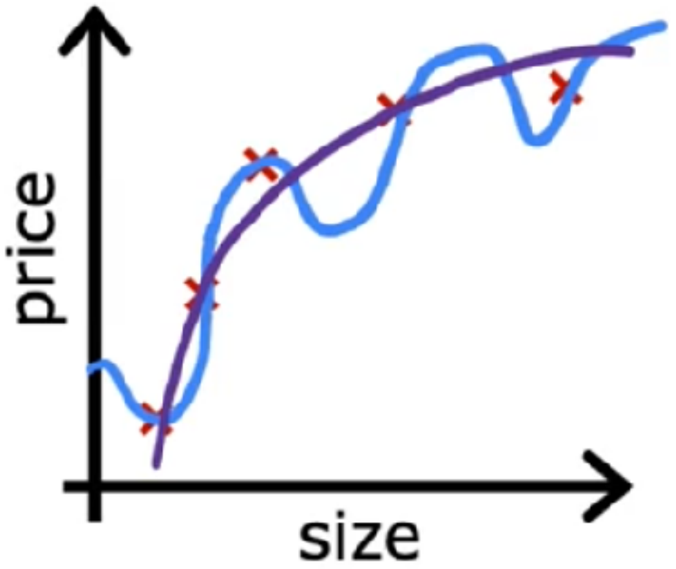
3. Reduce size of parameters ↓
   - "Regularization"

Cost Function with Regularization

Intuition



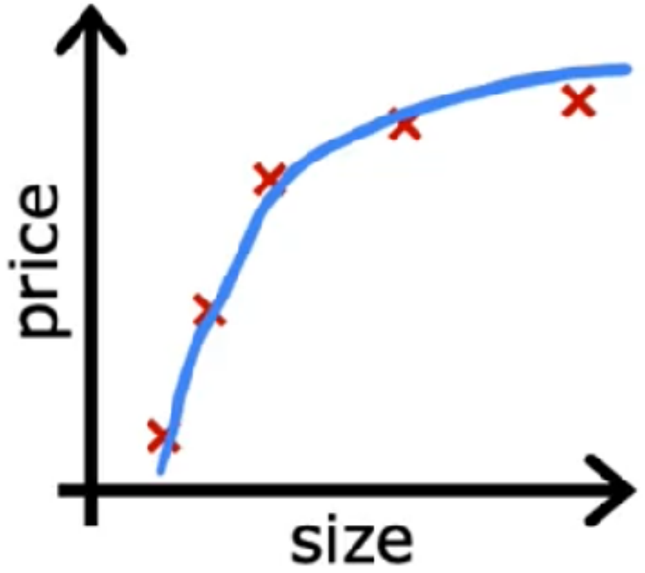
$w_1 x + w_2 x^2 + b$

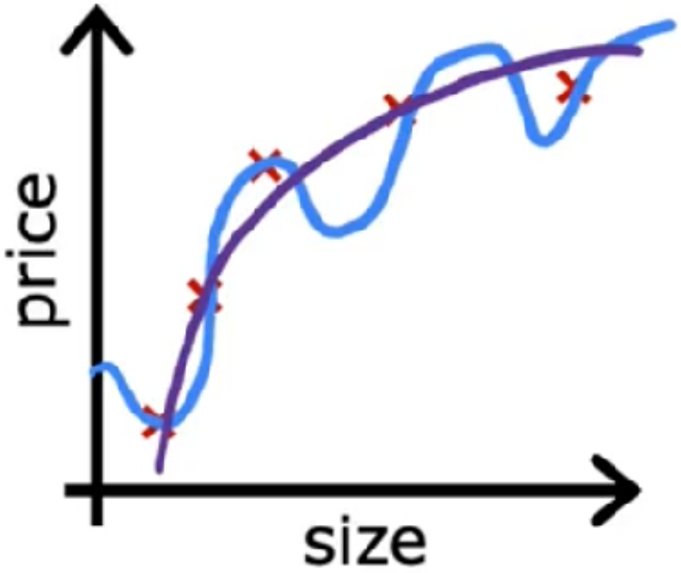


$w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + b$

make w_3, w_4 really small (≈ 0)

Intuition





$w_1x + w_2x^2 + b$

$w_1x + w_2x^2 + w_3x^3 + w_4x^4 + b$

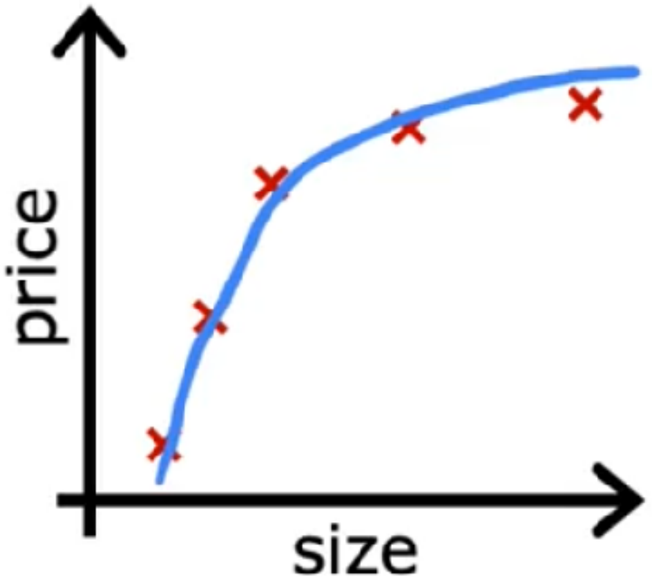
make w_3, w_4 really small (≈ 0)

$$\min_{\vec{w},b} \frac{1}{2m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2 + 1000\, w_3^2 + 1000\, w_4^2$$

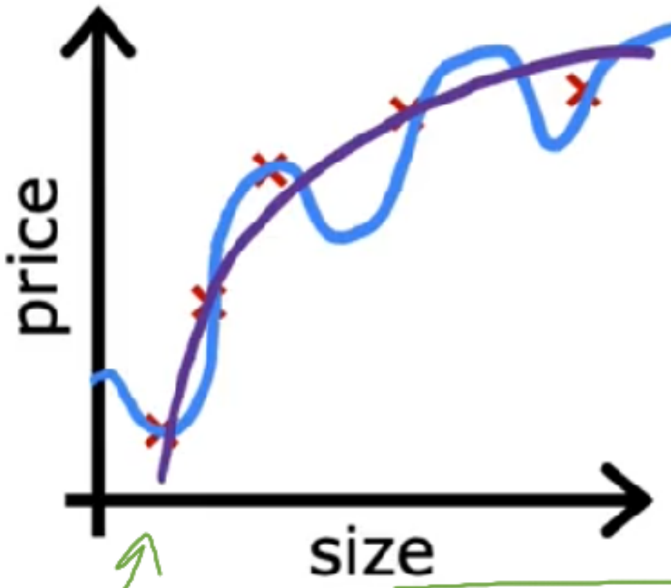
J

min J

Intuition



$$w_1 x + w_2 x^2 + b$$



$$w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + b$$

$w_3 \approx 0$ $w_4 \approx 0$

make w_3, w_4 really small (≈ 0)

$$\min_{\vec{w},b} \frac{1}{2m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right)^2 + 1000 w_3^2 + 1000 w_4^2$$

$w_3 \approx 0$ (0.001) $w_4 \approx 0$ (0.002)

Regularization

small values $w_1, w_2, \cdots, w_n, b$

simpler model

less likely to overfit

$w_3 \approx 0$

$w_4 \approx 0$

size	bedrooms	floors	age	avg income	...	distance to coffee shop	price
x_1	x_2	x_3	x_4	x_5		x_{100}	y

$n = 100$

n features

$w_1, w_1, w_2, \cdots, w_{100}, b$

w_3 , w_4

Regularization

small values $w_1, w_2, \cdots, w_n, b$

simpler model

less likely to overfit

$w_3 \approx 0$

$w_4 \approx 0$

size	bedrooms	floors	age	avg income	...	distance to coffee shop	price
x_1	x_2	x_3	x_4	x_5		x_{100}	y

$n = 100$

n features

$w_1, w_1, w_2, \cdots, w_{100}, b$

regularization term

$$J(\vec{w}, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} w_j^2$$

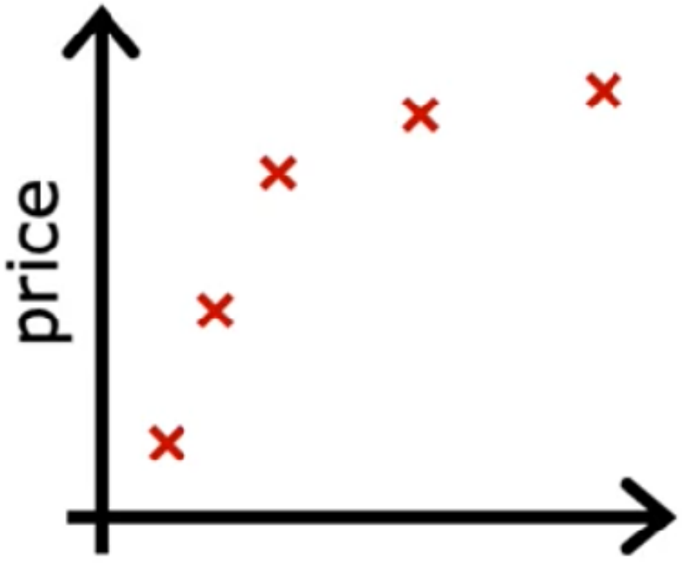
"lambda"

regularization parameter

$\lambda > 0$

Regularization

$$J(\vec{\mathrm{w}}, b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$



$$f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}) = w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + b$$

x	y
1	1.5
3	3
7	6

x_0	x	x^2	y
1			
1			
1			

$wx+b$

$w_1 x + w_2 x^2 + b$

$w_0 x_0 + w_1 x_1 + w_2 x_2 = \hat{y}$