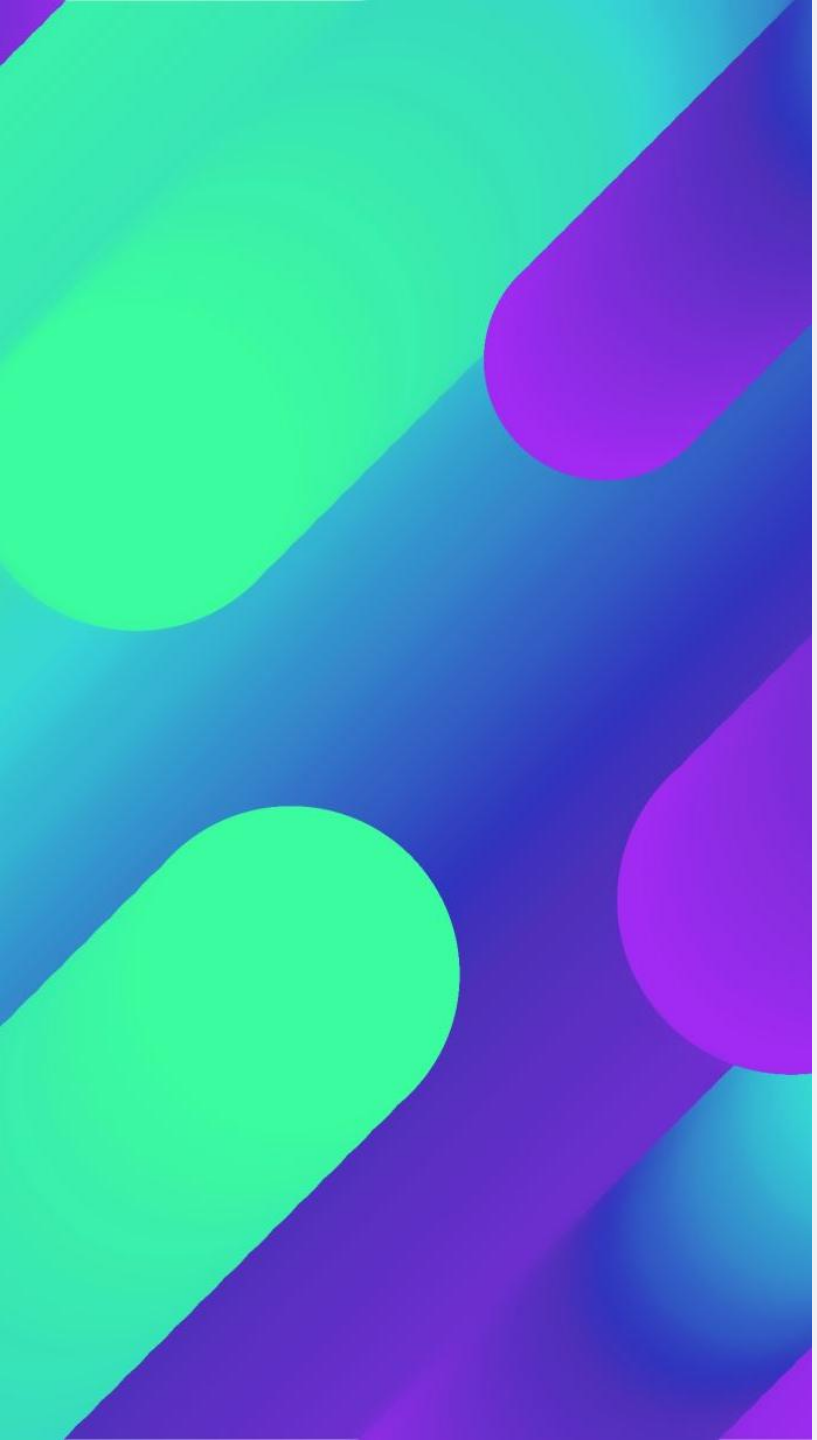
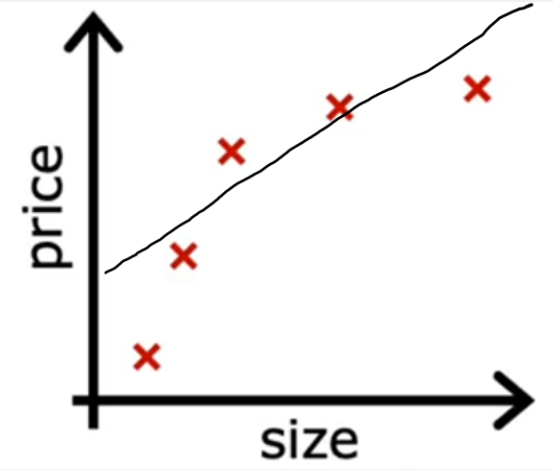


# Regularization

Overfitting

# Regression

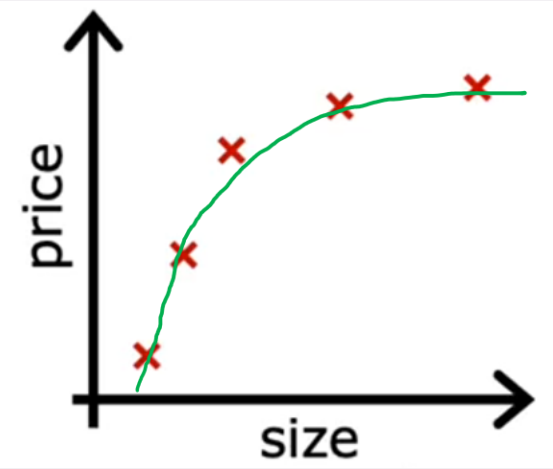



$f(x) = w_1 x + b$

×

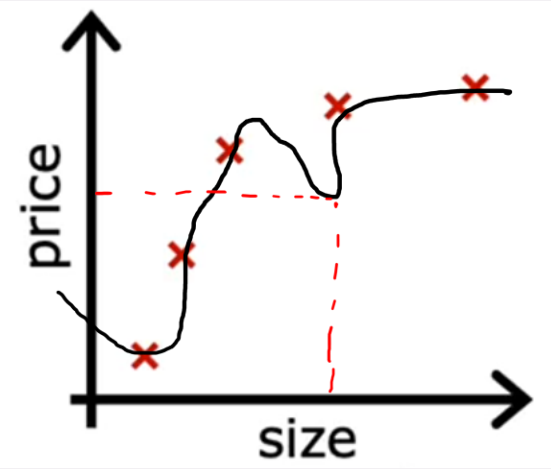
High Bias

Under fitting




$f(x) = w_1 x + w_2 x^2 + b$

✓




$f(x) = w_1 x + w_2 x^2 + \ldots$

$\ldots + w_3 x^3 + w_4 x^4$

$\ldots + b$

×

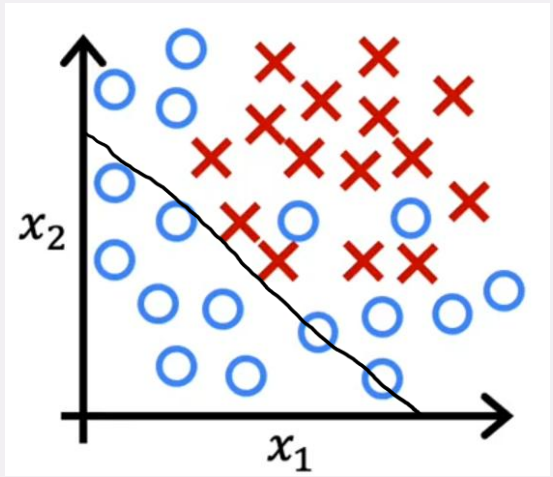
High Variance

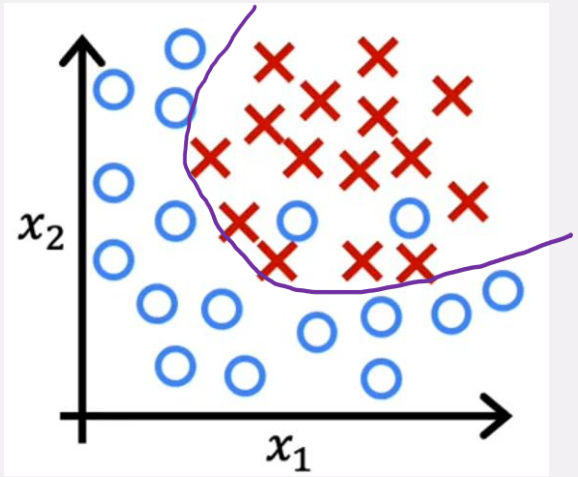
Over fitting

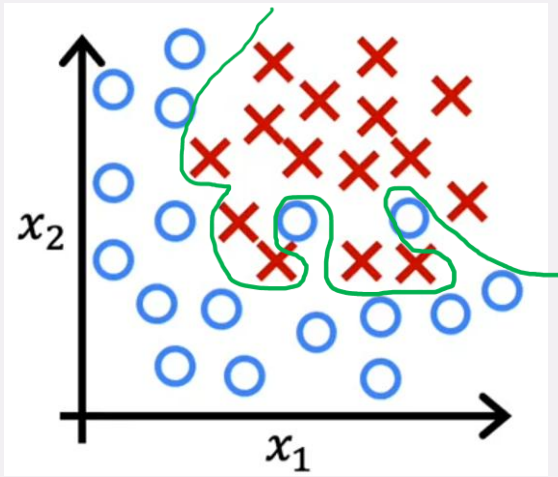
# Classification

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

$$Z = w_1 x_1 + w_2 x_2 + w_3 x_1^2 x_2 + w_4 x_1^2 x_2^2 + \dots + b$$




High Bias

$f(\vec{x}) = g(z)$

Under fitting

$z = w_1 x_1 + w_2 x_2 + b$

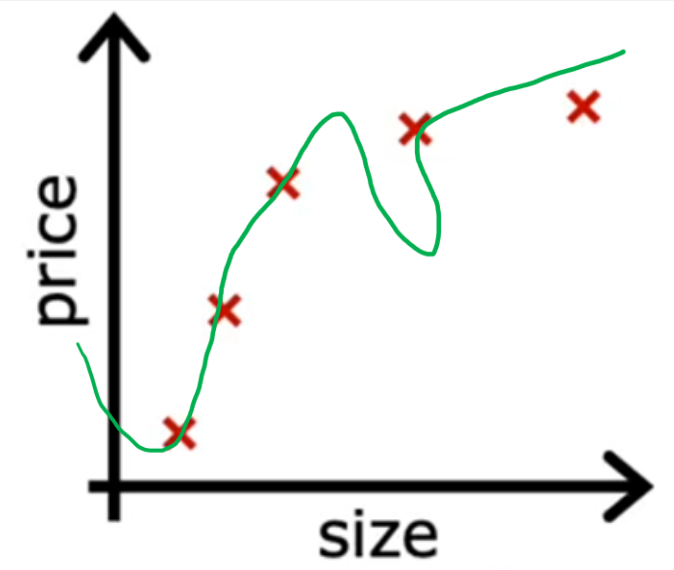
High Variance

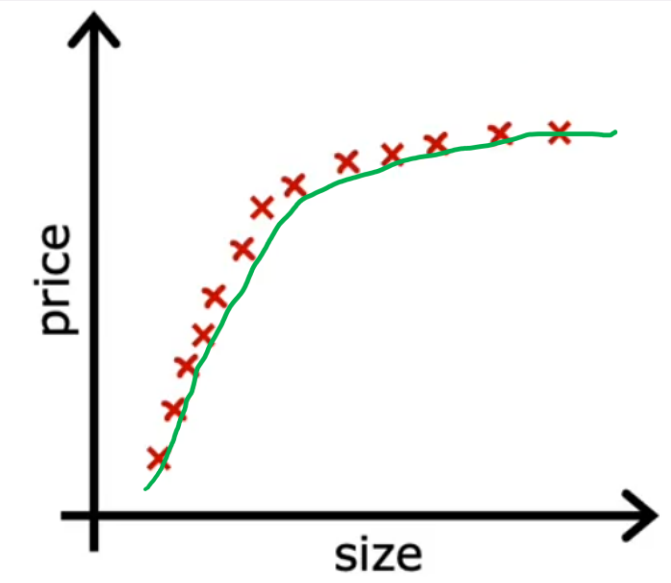
Overfitting

# Addressing Overfitting condition

- Collect more training examples (data)
- Feature Selection (Reduce number of features)
- Regularization

# Collect more training examples

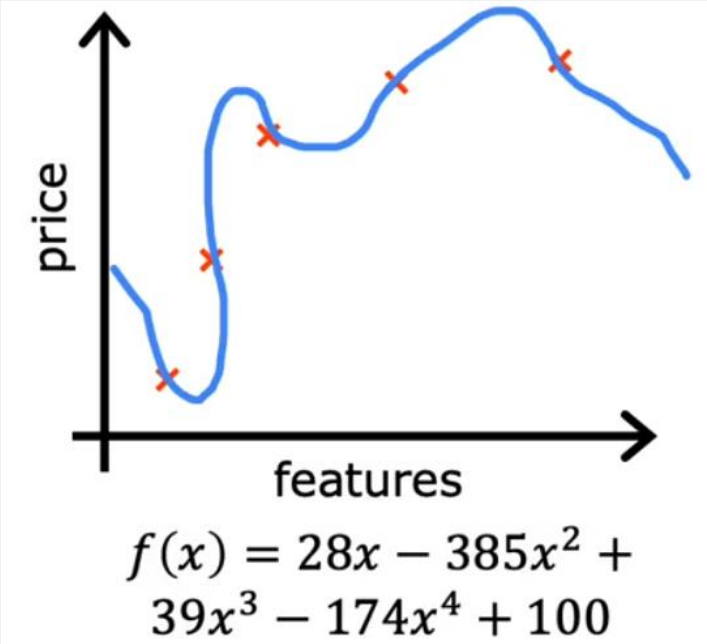


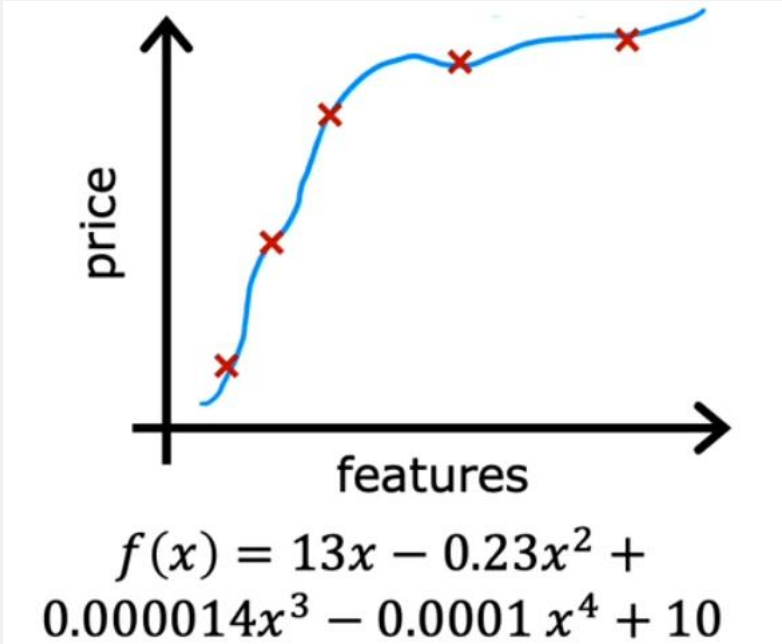



# Feature Selection

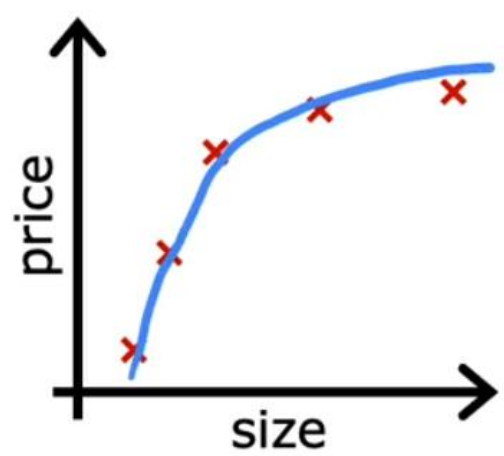
| size | bedrooms | floors | age | avg income | ... | distance to coffee shop | price |
|---|---|---|---|---|---|---|---|
| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | ... | $x_{5-0}$ | $y$ |
| | | | | | | | |

# Regularization

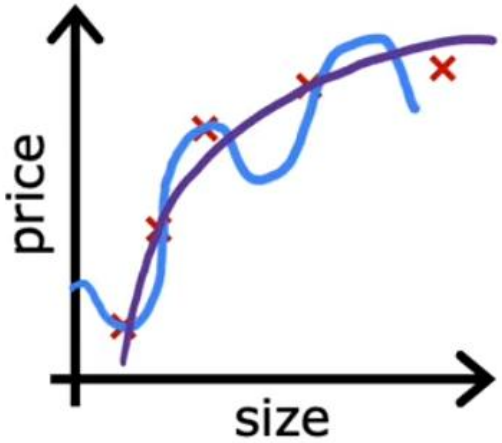


# Regularization (intuition)




$w_1x + w_2x^2 + b$




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

$\approx 0 \quad \approx 0$

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

$w_3 \downarrow 0.001 \qquad w_4 \downarrow 0.002$

# Regularization

| size | bedrooms | floors | age | avg income | ... | distance to coffee shop | price |
|---|---|---|---|---|---|---|---|
| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | | $x_{50}$ | $y$ |

$w_1, w_2, w_3, \cdots w_{50}, b$

Regularization Parameter

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

$$+ \frac{\lambda}{2m} b^2$$

optional

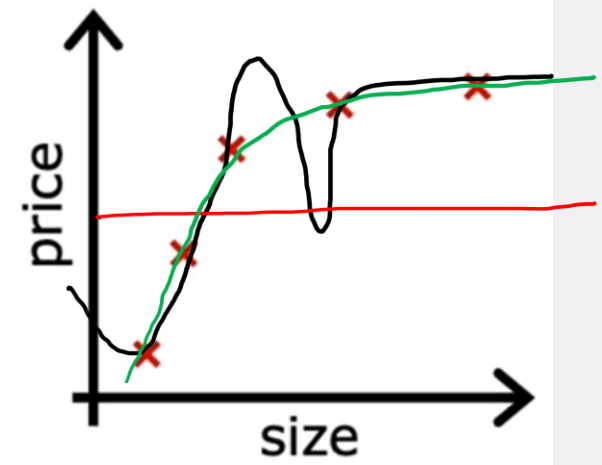
$f(\vec{x}) = \vec{w}\vec{x} + b$

$\lambda \to 0$ High Variance

# Regularization

$\lambda \to \infty \to w \to 0 \to f(\vec{x}) \approx b$

High Bias




$\lambda \to 10^3$ Just Right

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