Learning Curves

Bias and Variance

Advice for Applying Machine Learning

Introduction

- Learning curves is often a very useful thing to plot.
  - you wanted to sanity check that your algorithm is working correctly, or
  - if you want to improve the performance of the algorithm.
- Tool to diagnose if a physical learning algorithm may be suffering from
  - bias,
  - variance problem or
  - a bit of both.

Learning curves

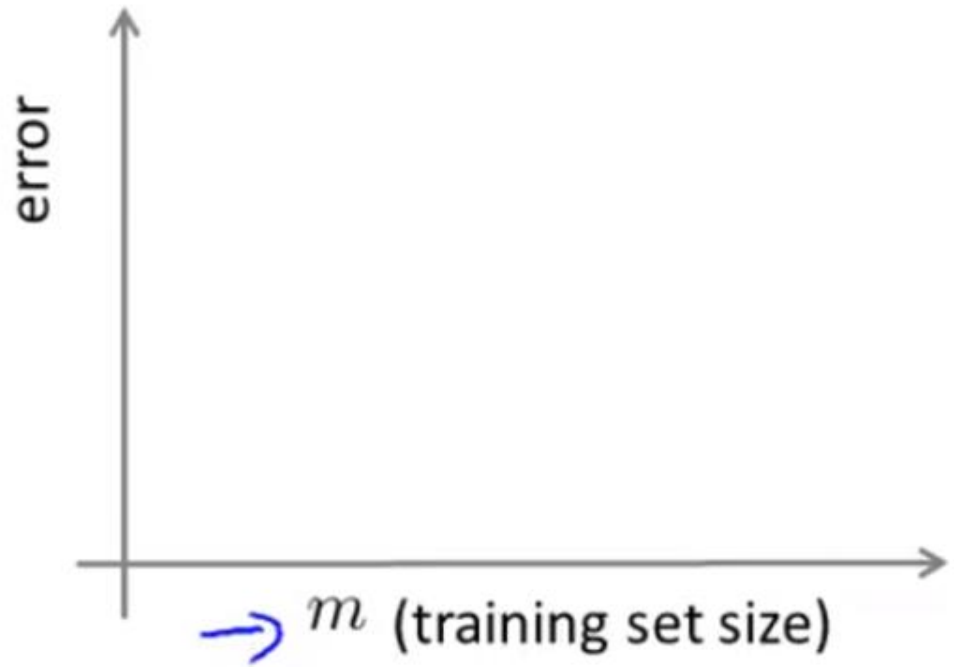
$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

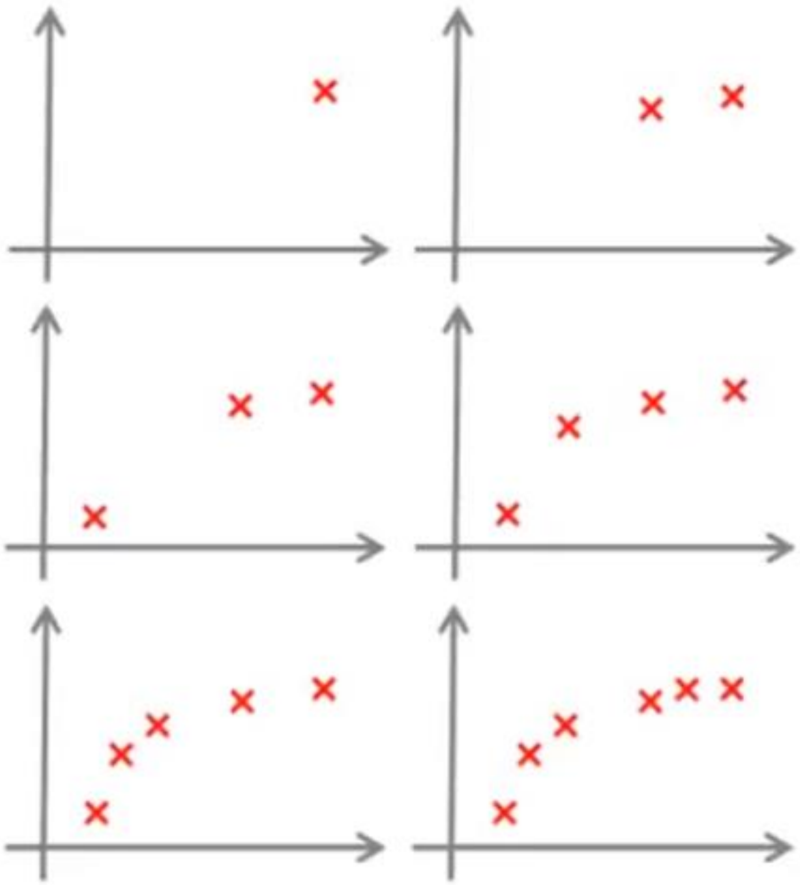
Learning curves

$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



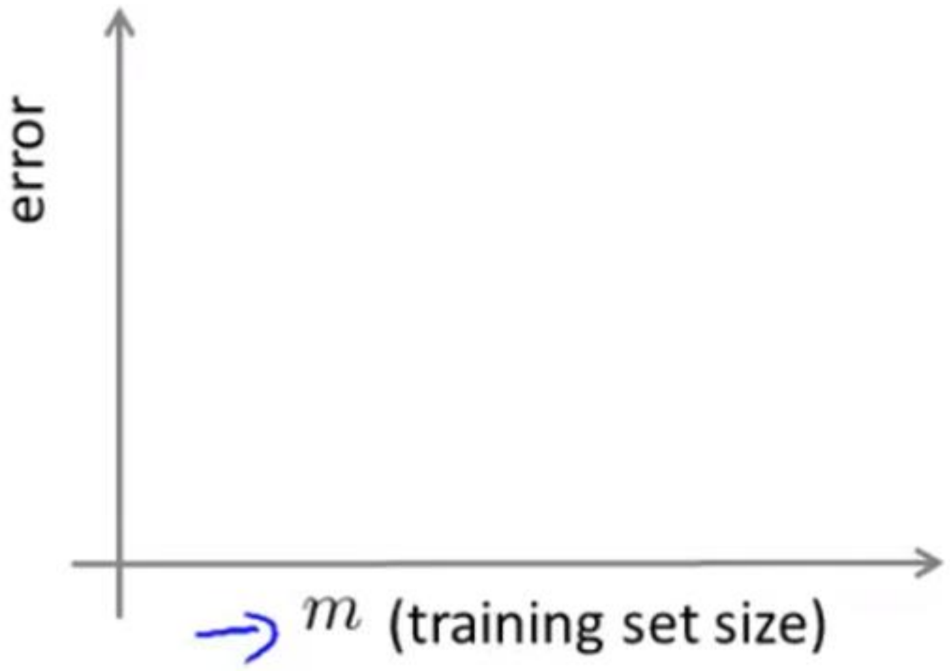
$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



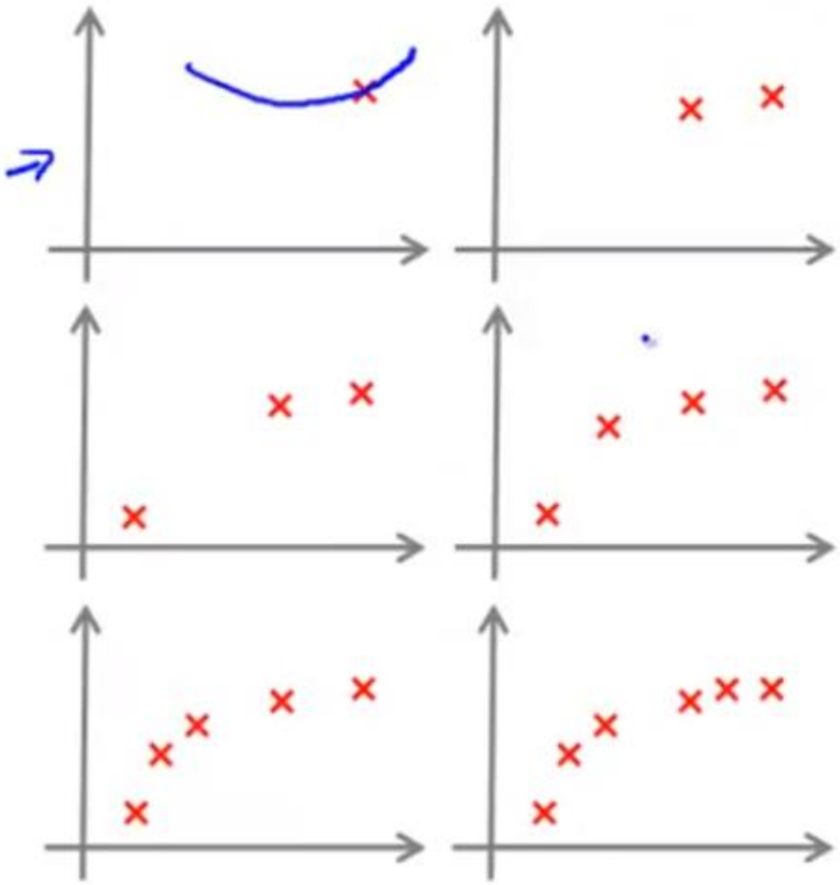
Learning curves

$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



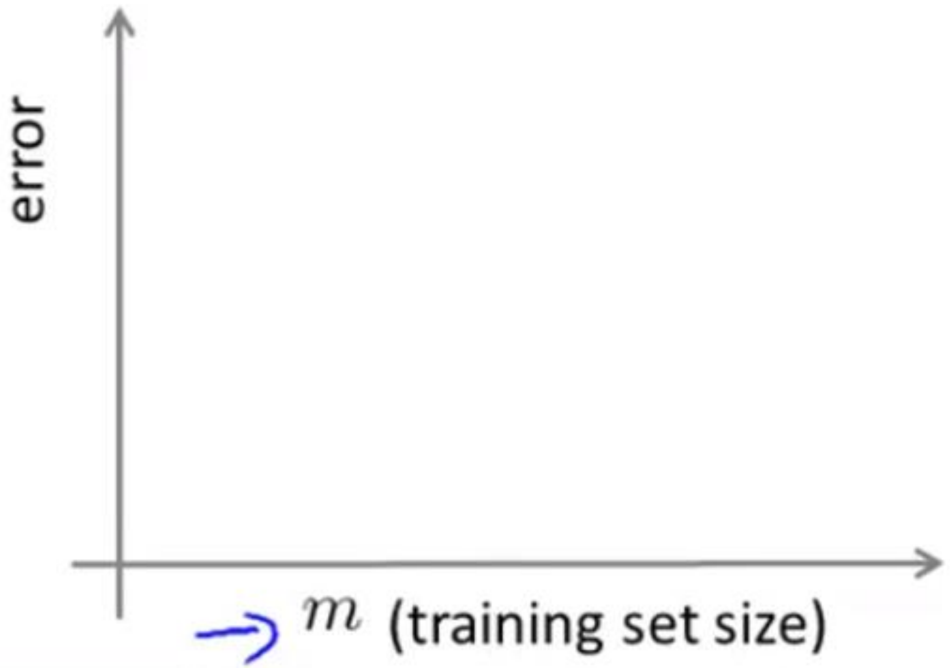
$$h_\theta(x) = \underline{\theta_0 + \theta_1 x + \theta_2 x^2}$$



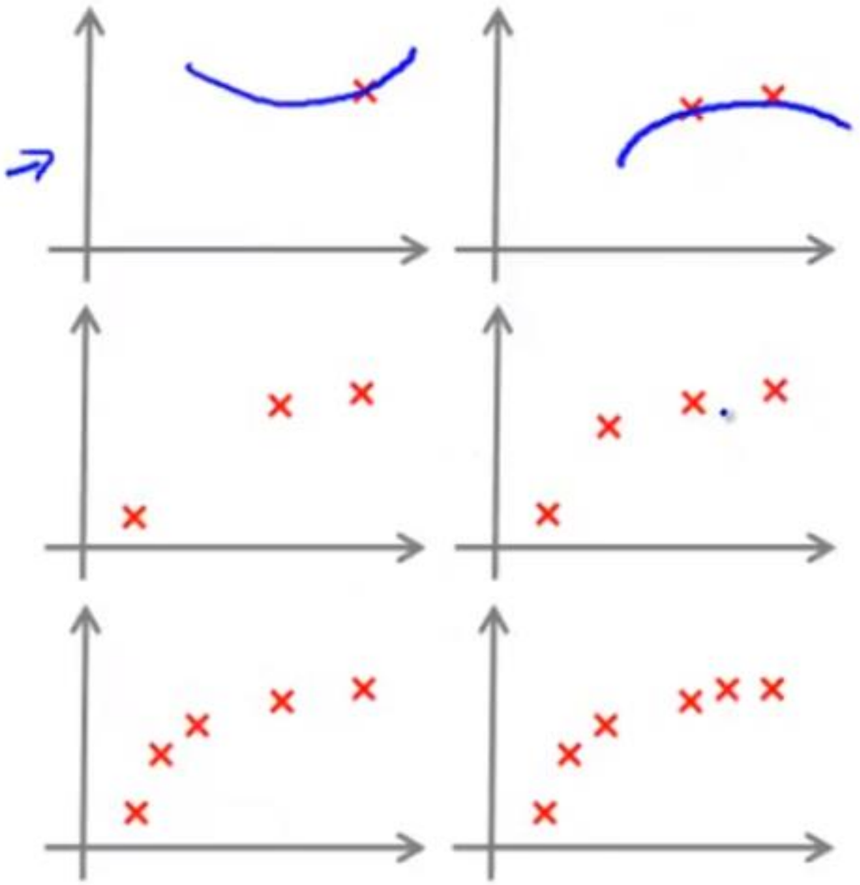
Learning curves

$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

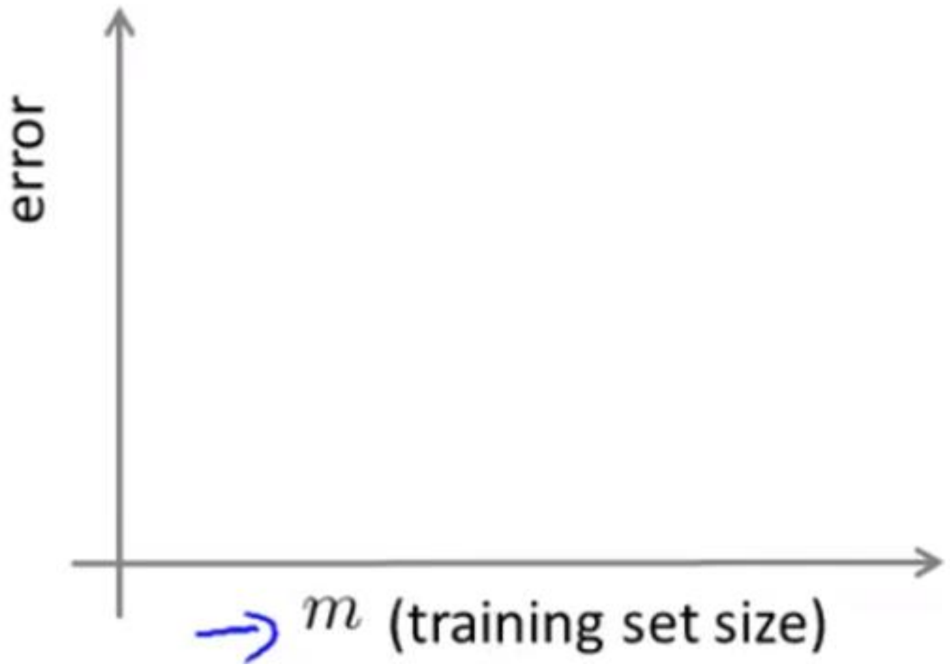


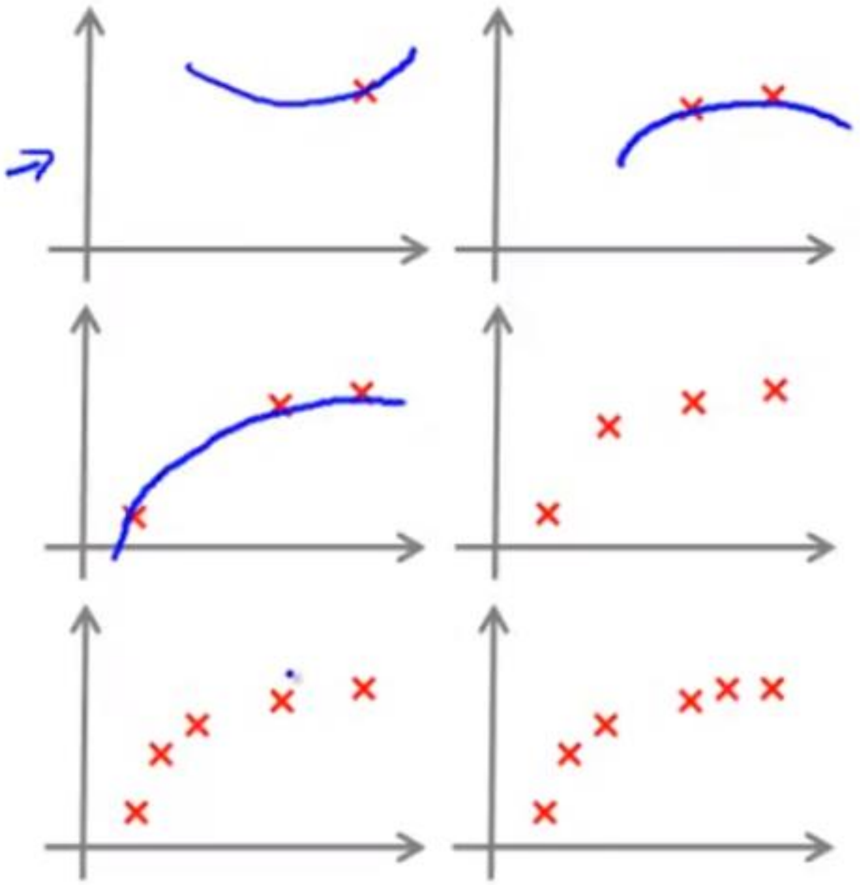
Learning curves

$$h_\theta(x) = \underline{\theta_0 + \theta_1 x + \theta_2 x^2}$$

$$J_{train}(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



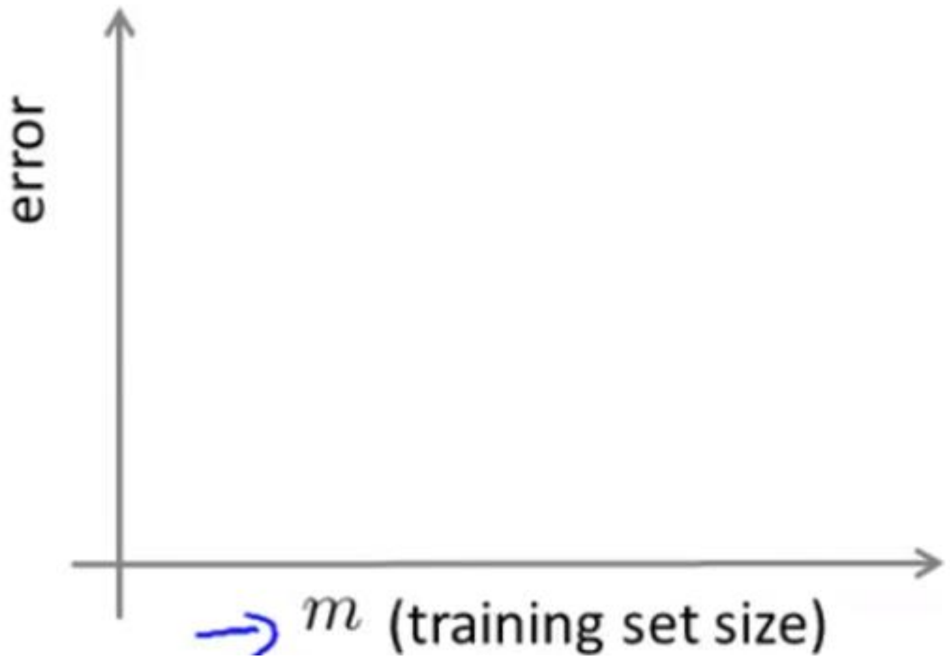


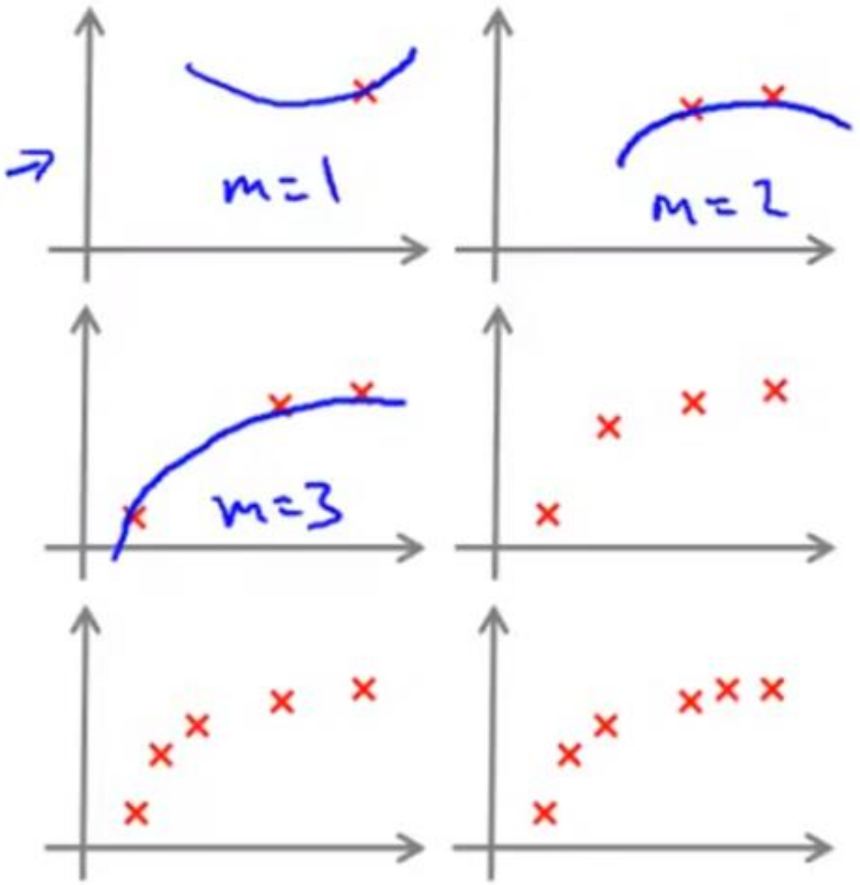
Learning curves

$$h_\theta(x) = \underline{\theta_0 + \theta_1 x + \theta_2 x^2}$$

$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



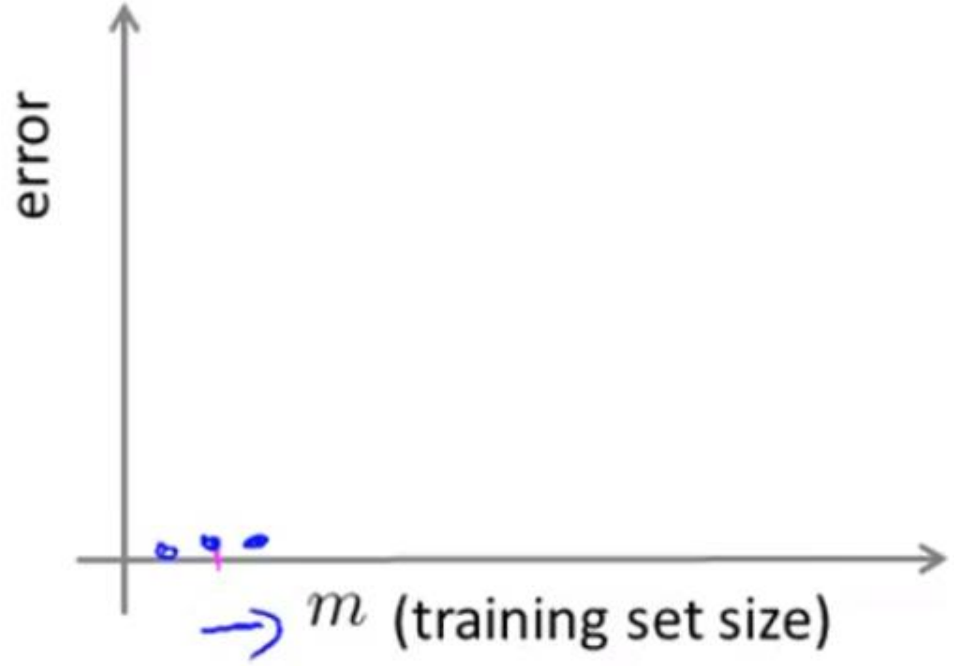


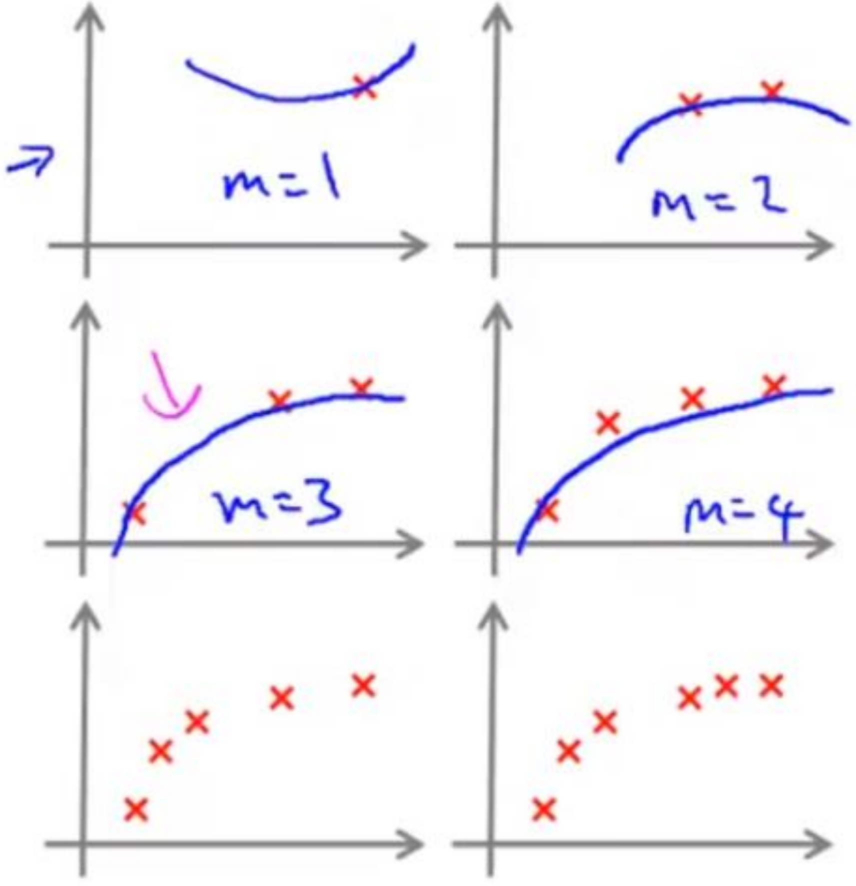
Learning curves

$$h_\theta(x) = \underline{\theta_0 + \theta_1 x + \theta_2 x^2}$$

$$\rightarrow \underline{J_{train}(\theta)} = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \leftarrow$$

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



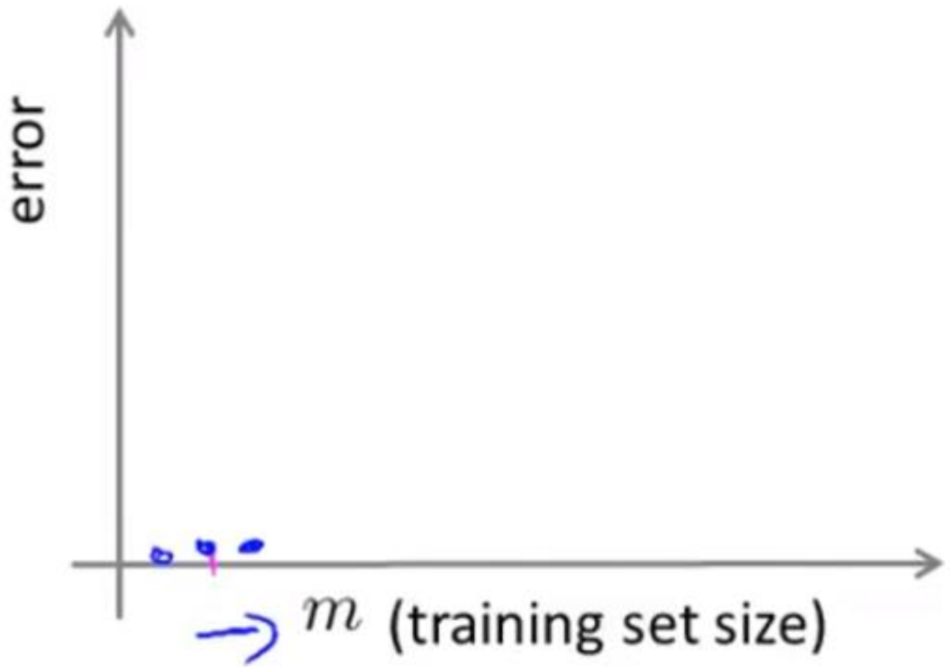


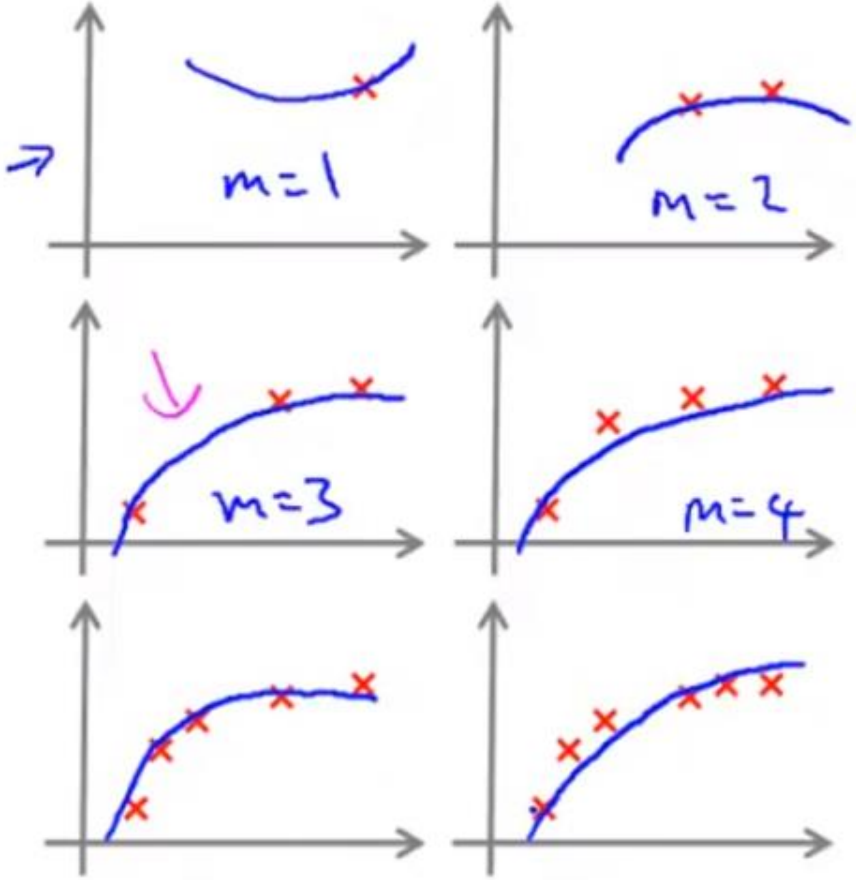
Learning curves

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



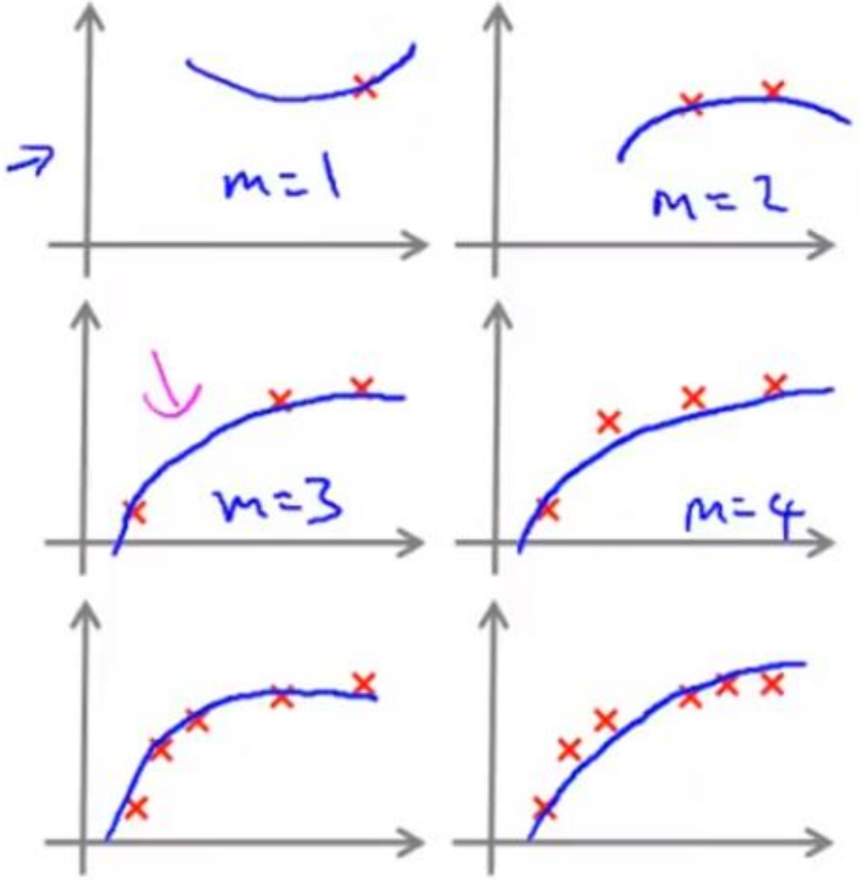


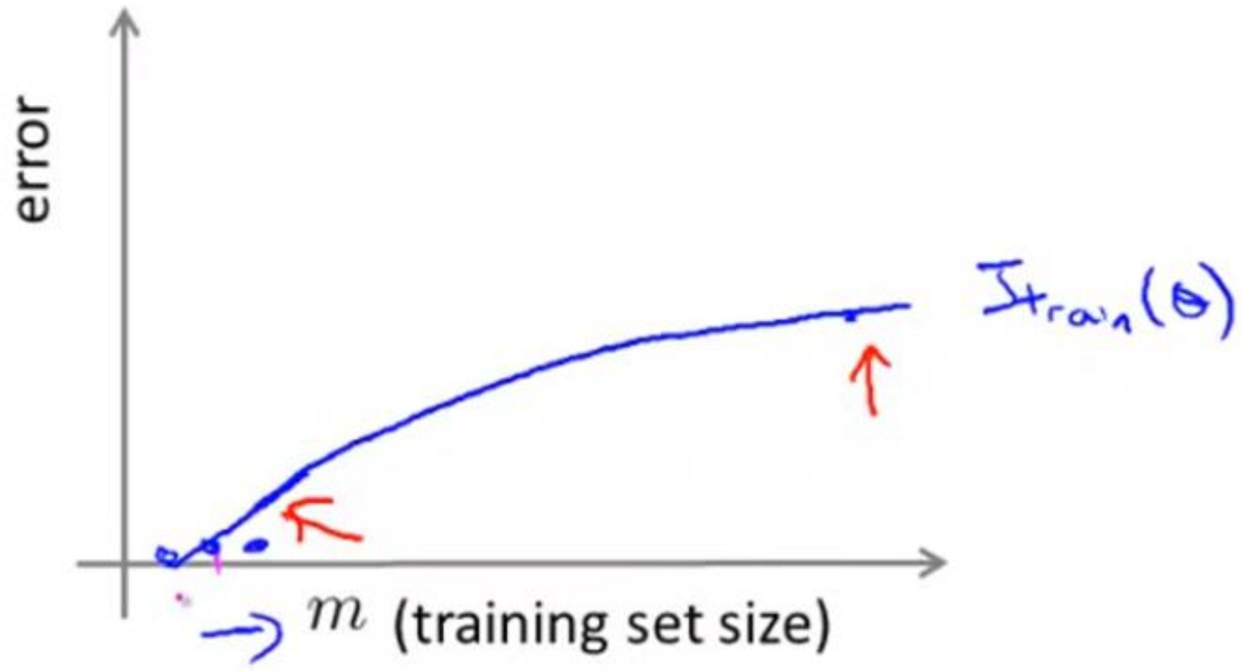
Learning curves

$$h_\theta(x) = \underline{\theta_0 + \theta_1 x + \theta_2 x^2}$$

$$\underline{J_{train}(\theta)} = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



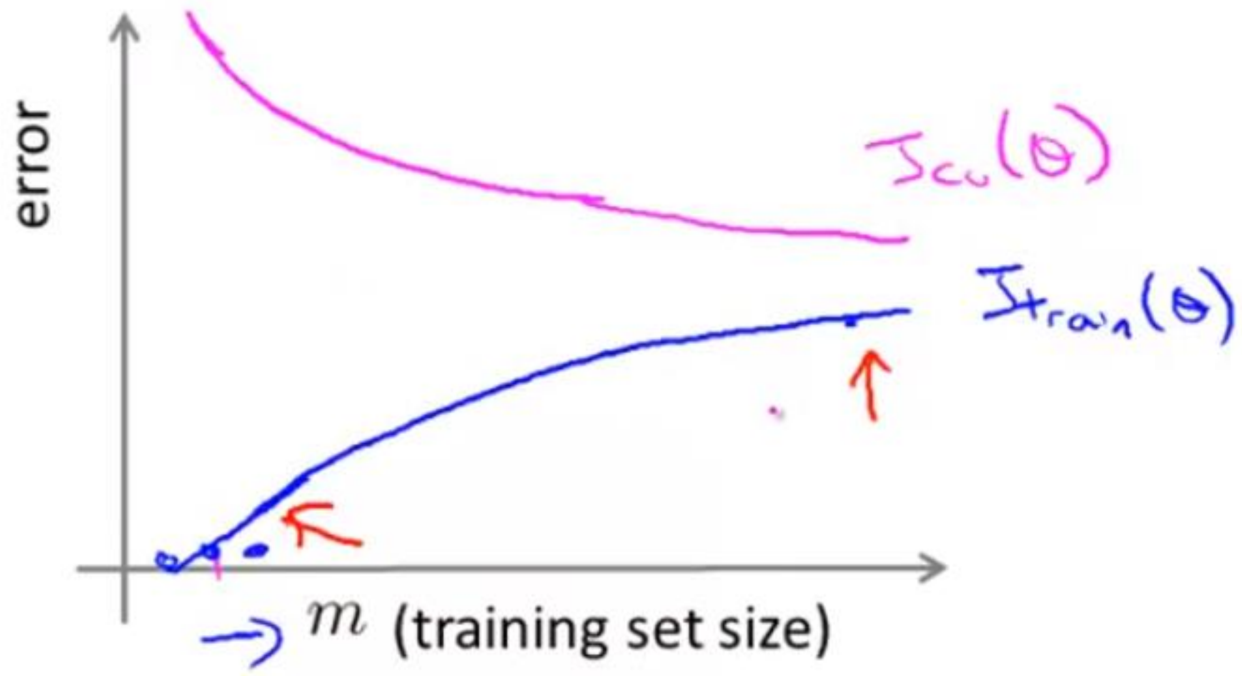


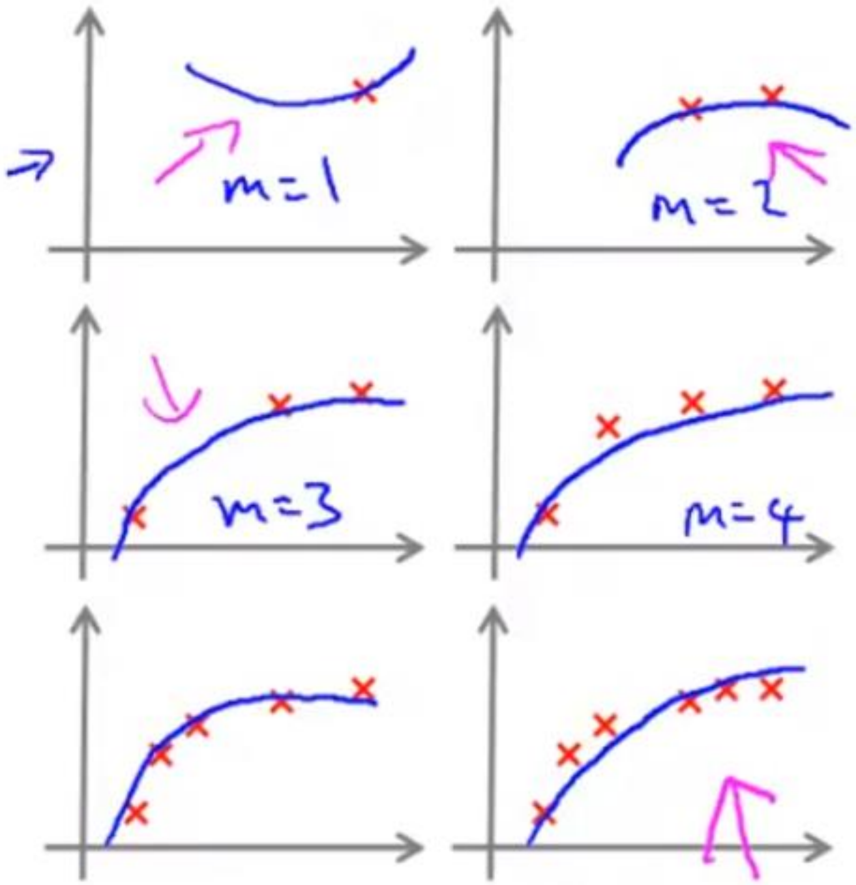
Learning curves

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

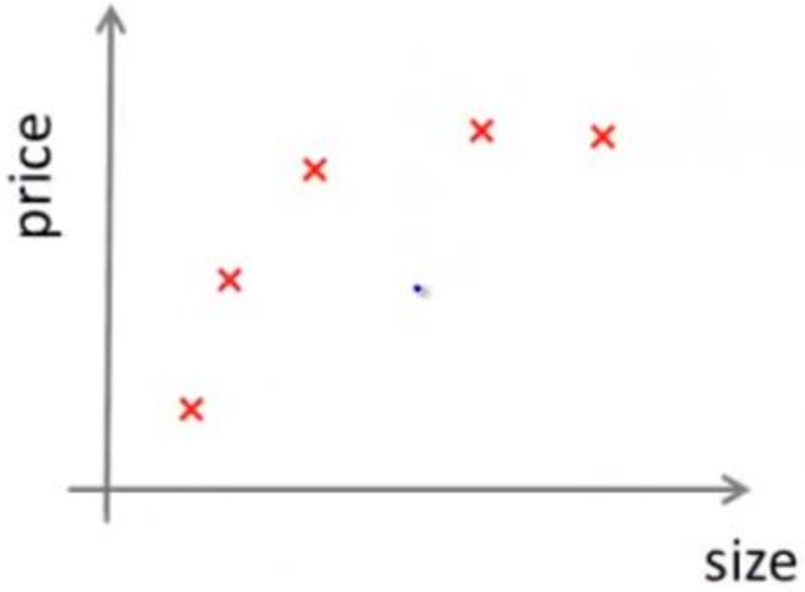
$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



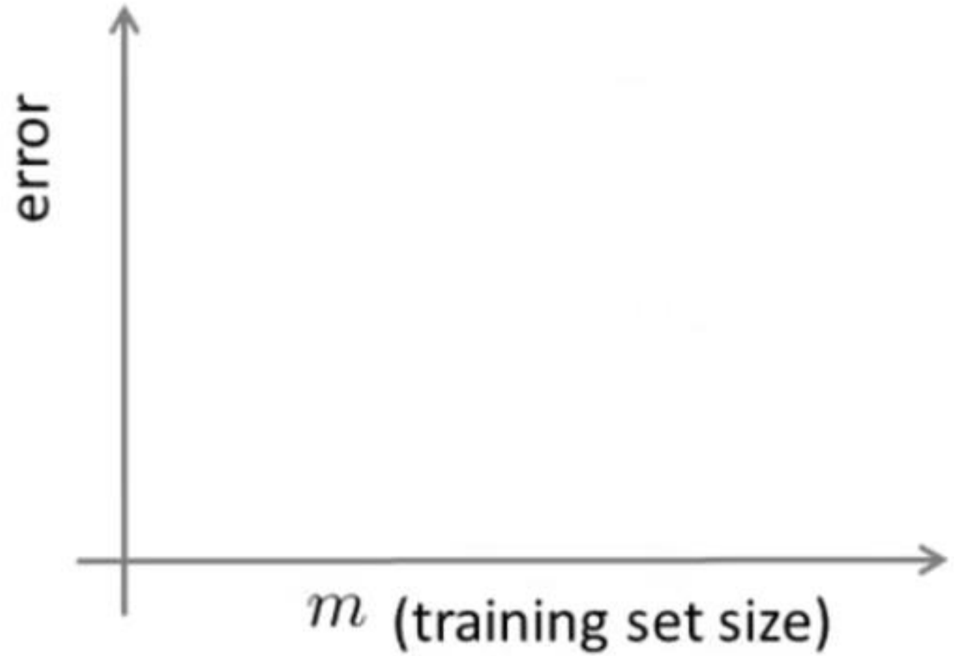


High bias

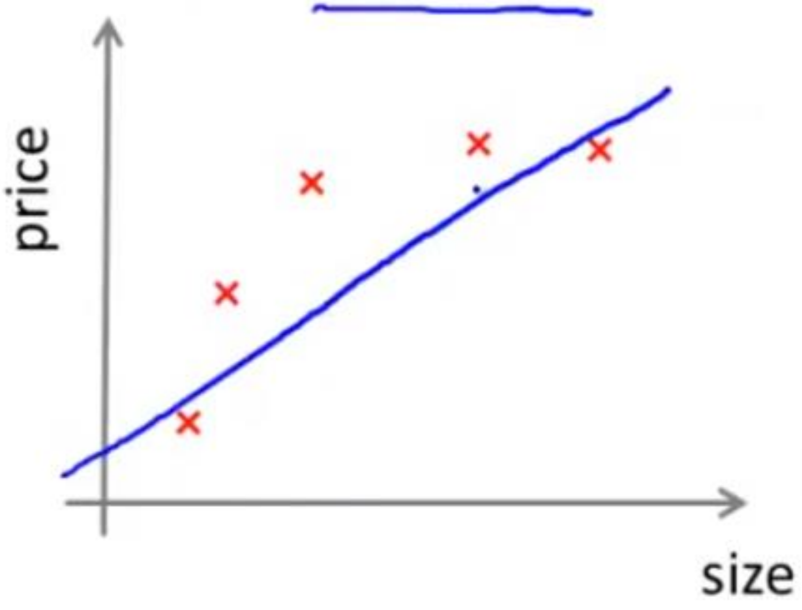
$h_\theta(x) = \theta_0 + \theta_1 x$



High bias

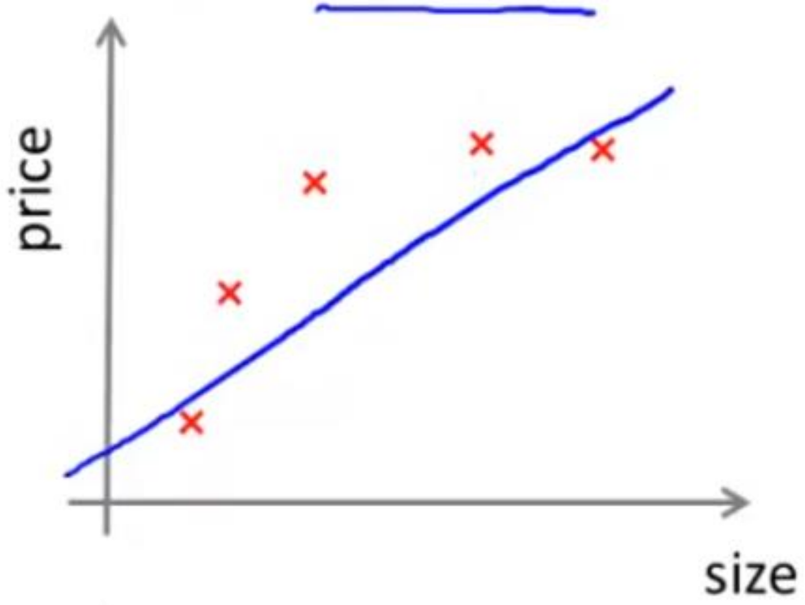


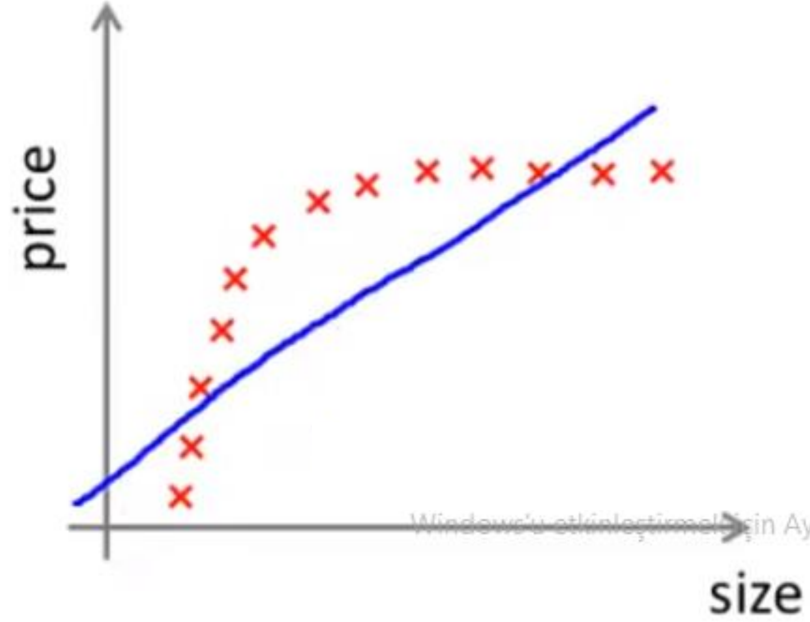
$$h_\theta(x) = \underbrace{\theta_0 + \theta_1 x}$$



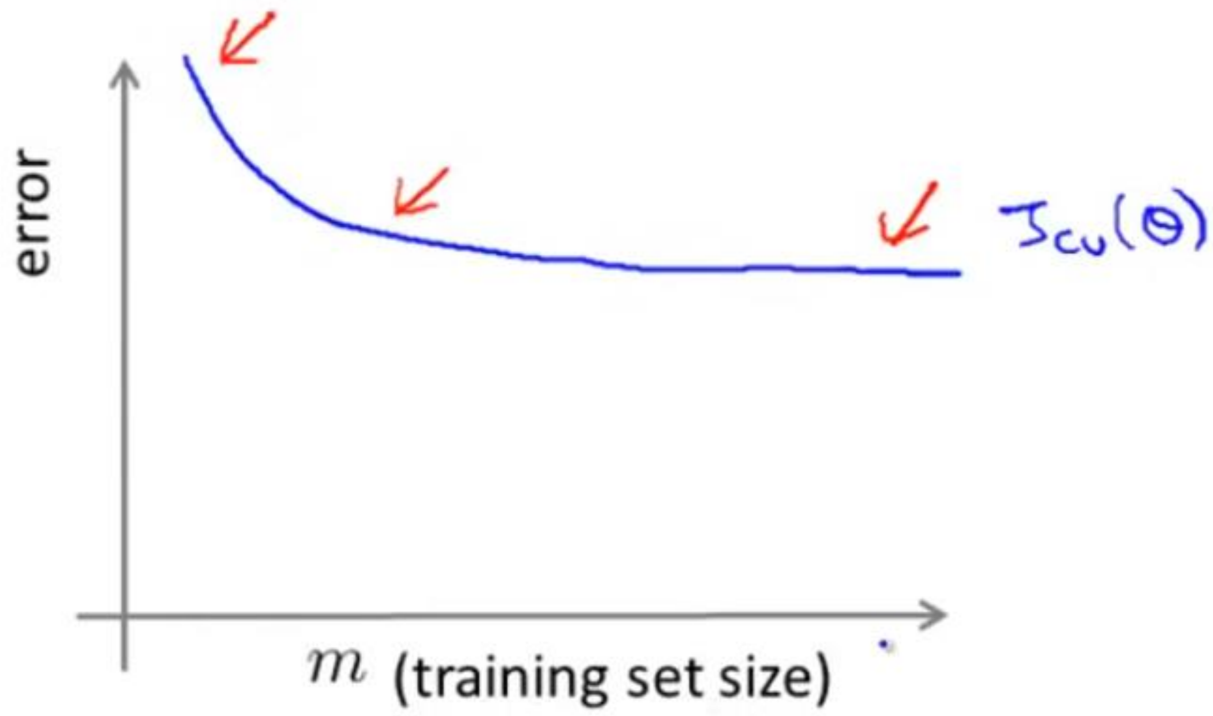
High bias

$$h_\theta(x) = \theta_0 + \theta_1 x$$

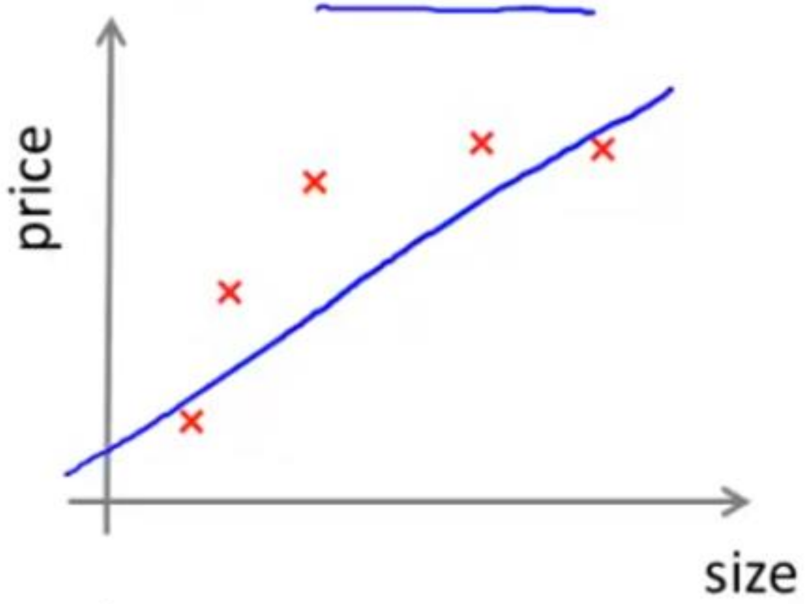


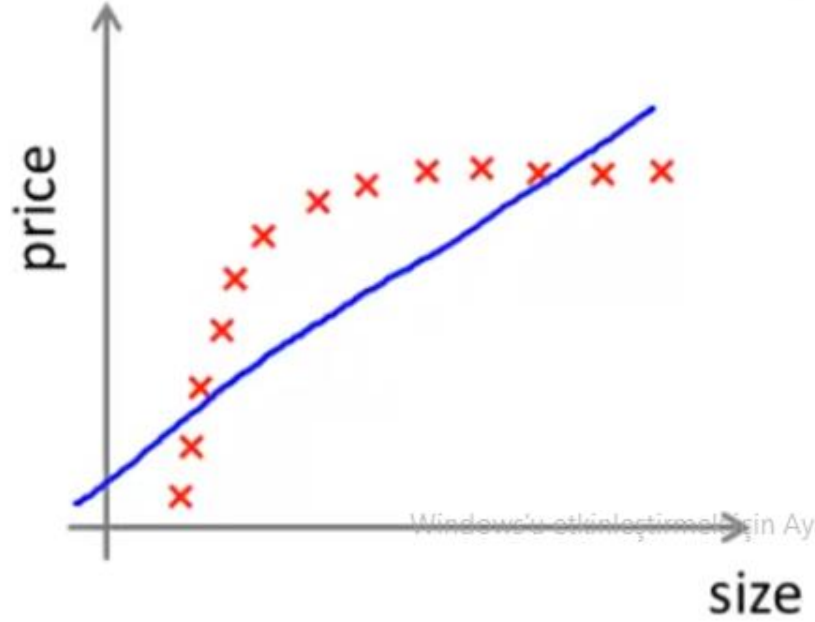


High bias

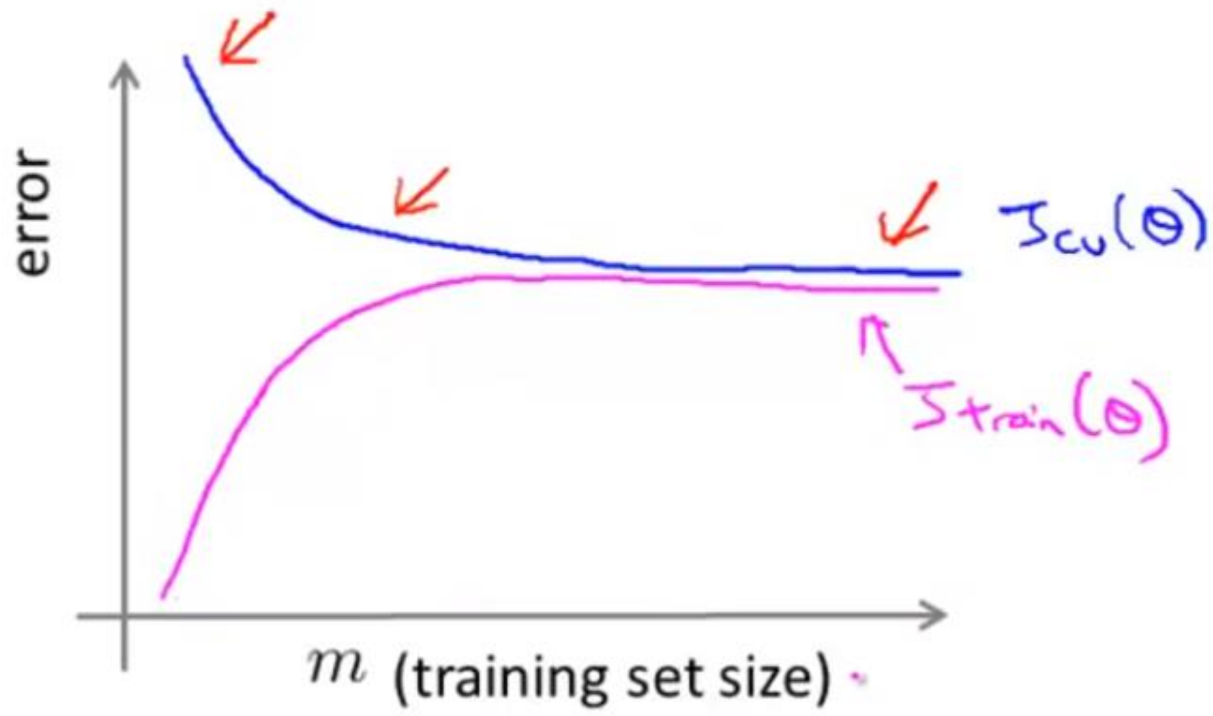


$$h_\theta(x) = \theta_0 + \theta_1 x$$

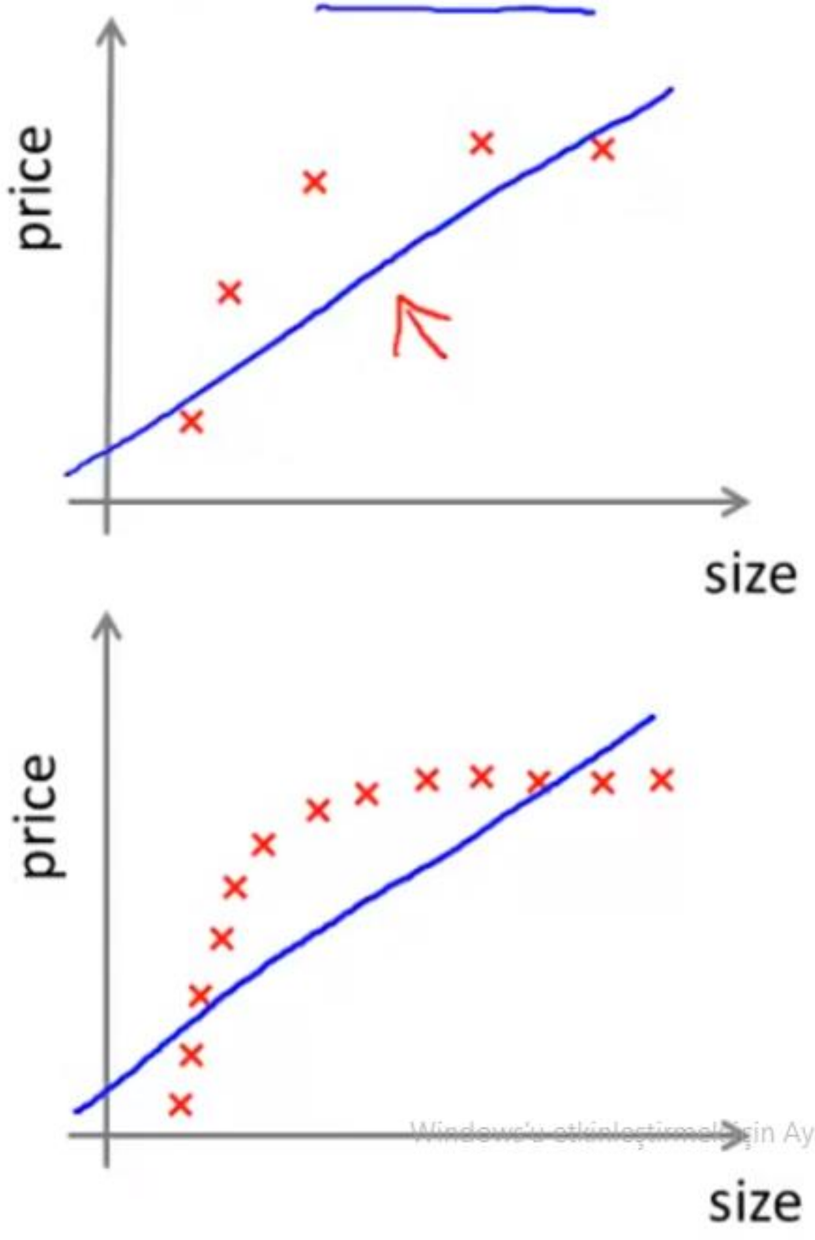




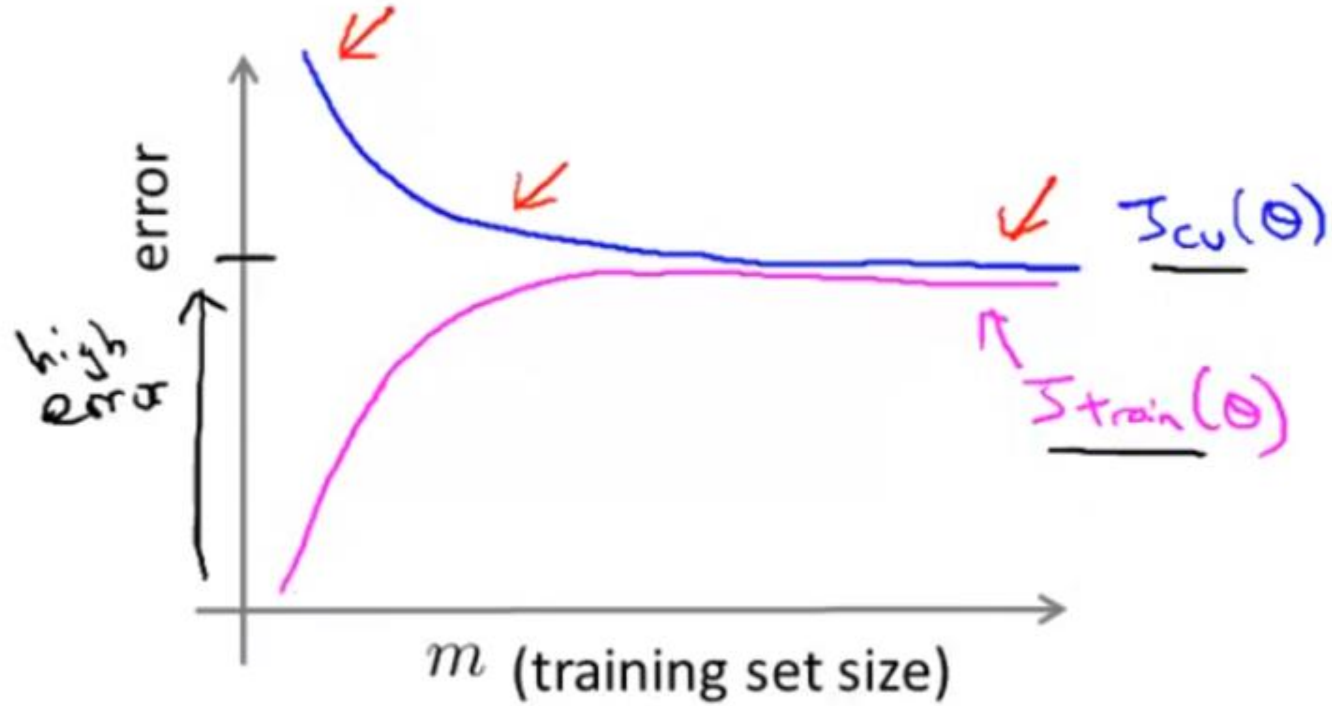
High bias



$$h_\theta(x) = \theta_0 + \theta_1 x$$

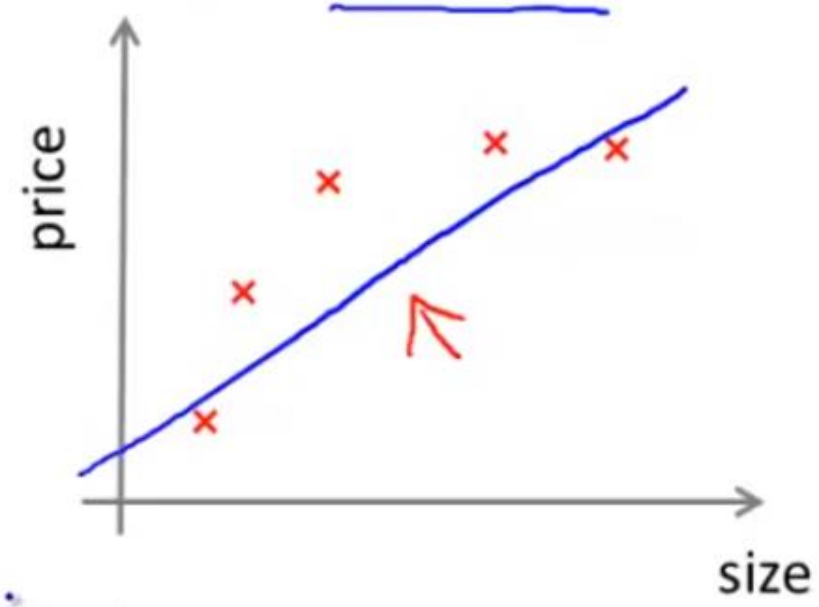


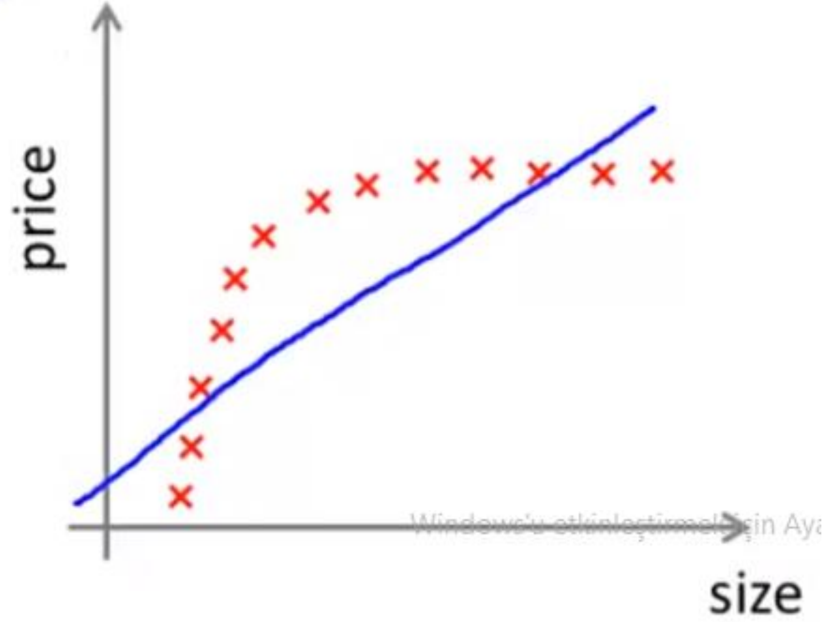
High bias



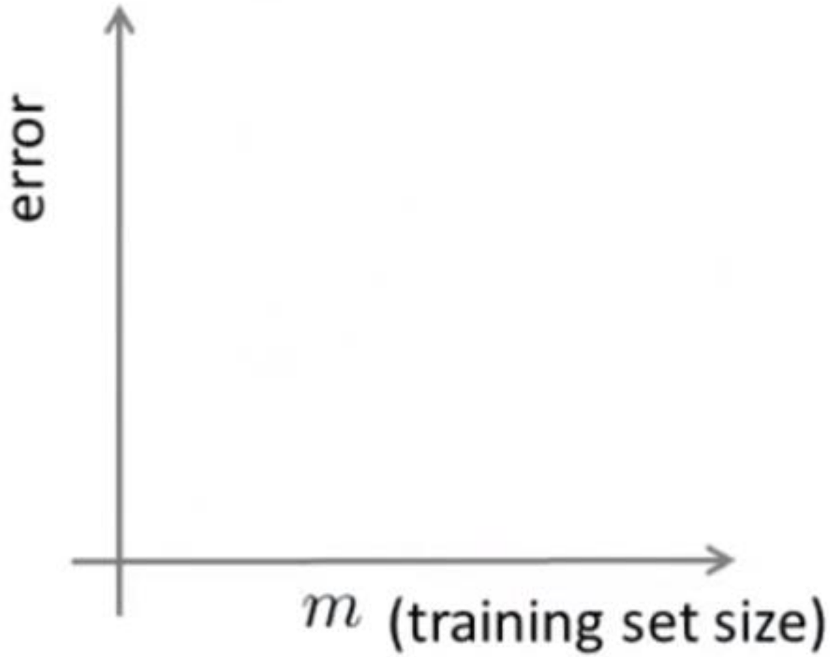
If a learning algorithm is suffering from high bias, getting more training data will not (by itself) help much.

$$h_\theta(x) = \theta_0 + \theta_1 x$$



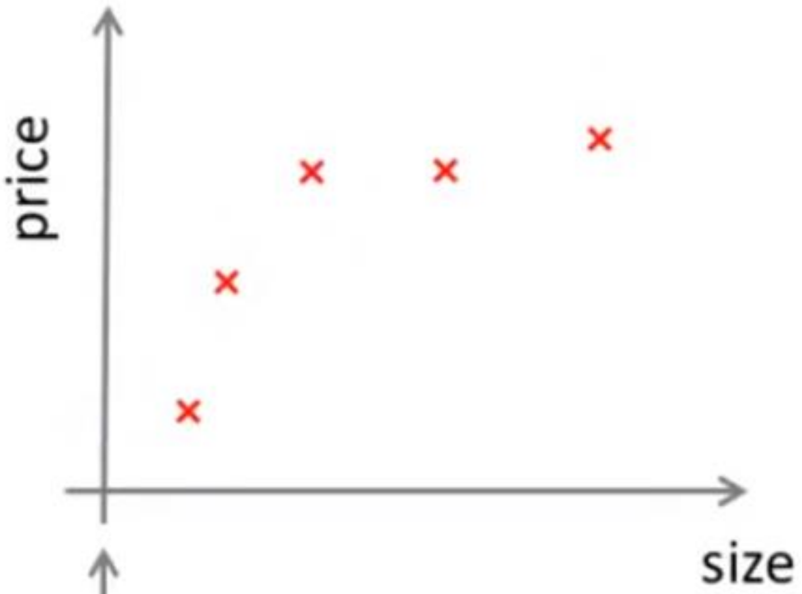


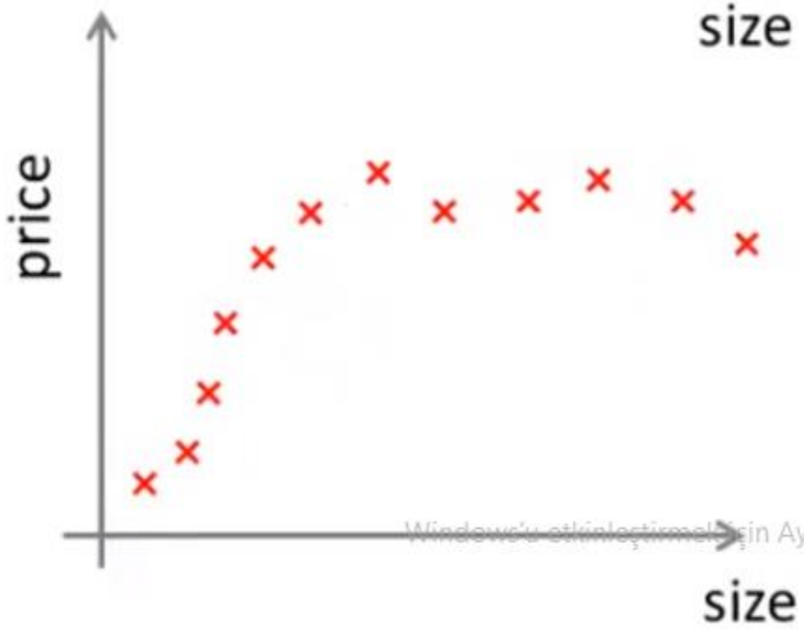
High variance



$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$

(and small λ)

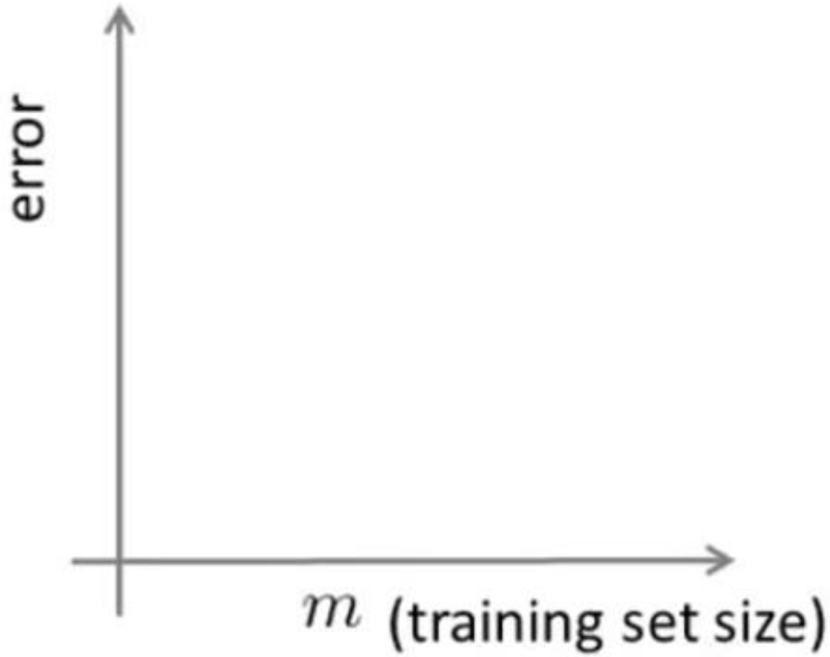


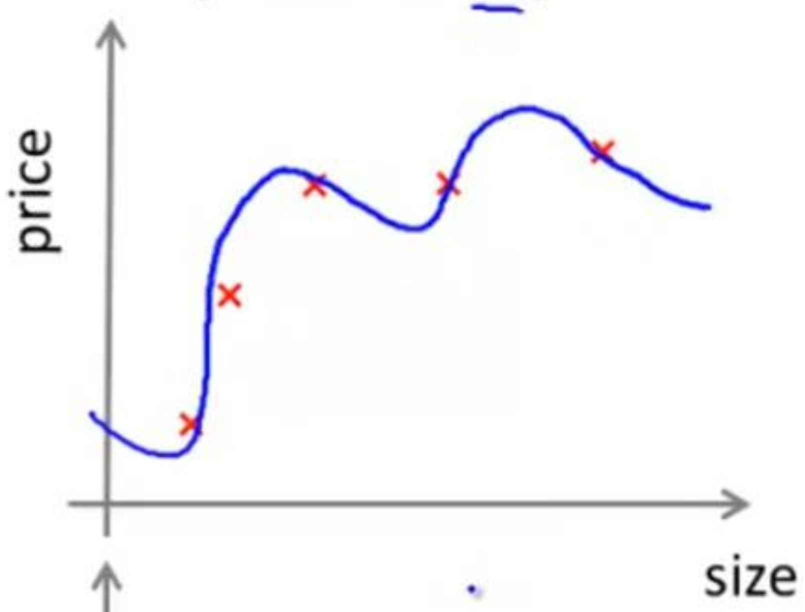


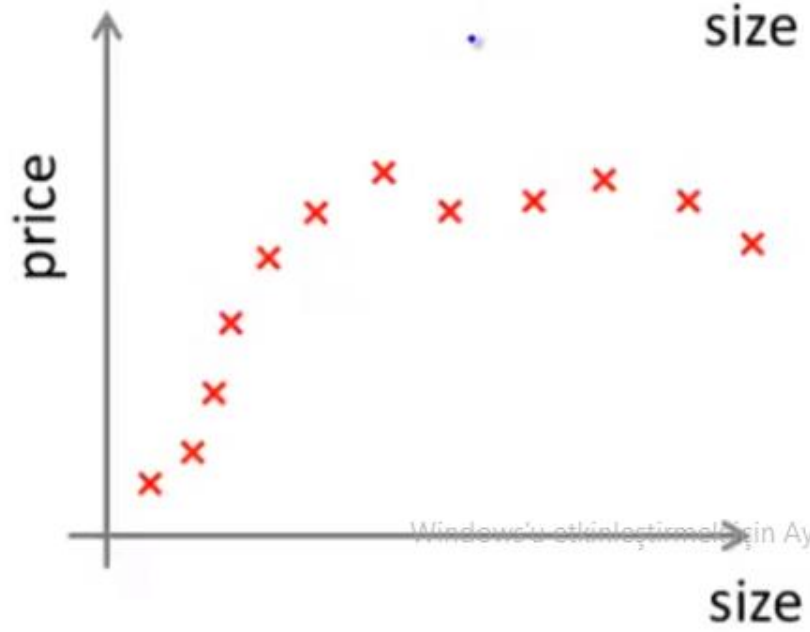
High variance

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$

(and small λ)



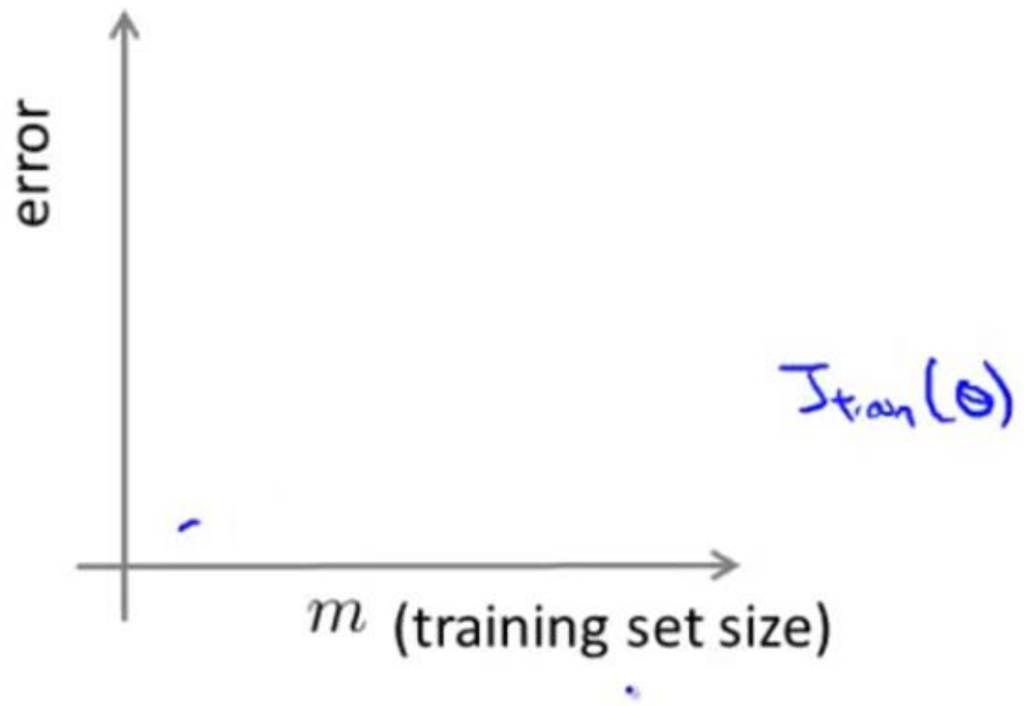


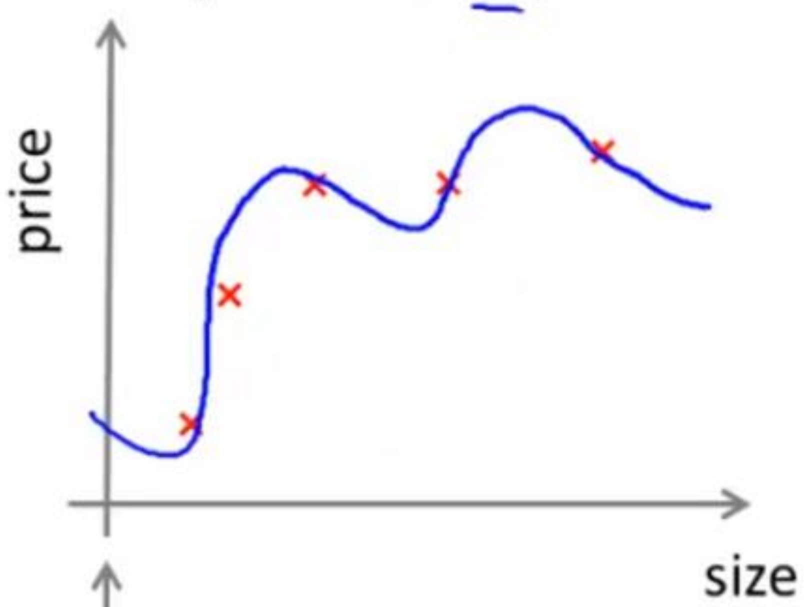


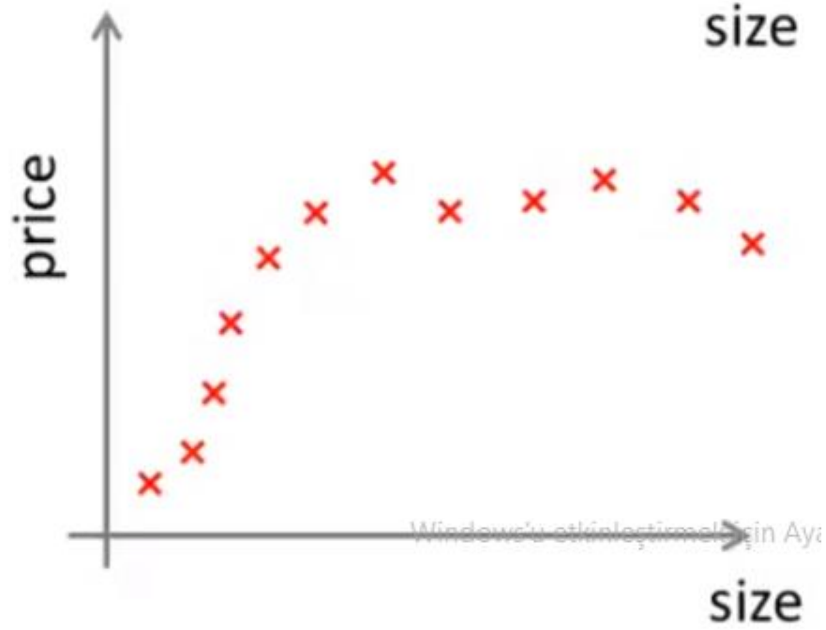
High variance

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$

(and small λ)



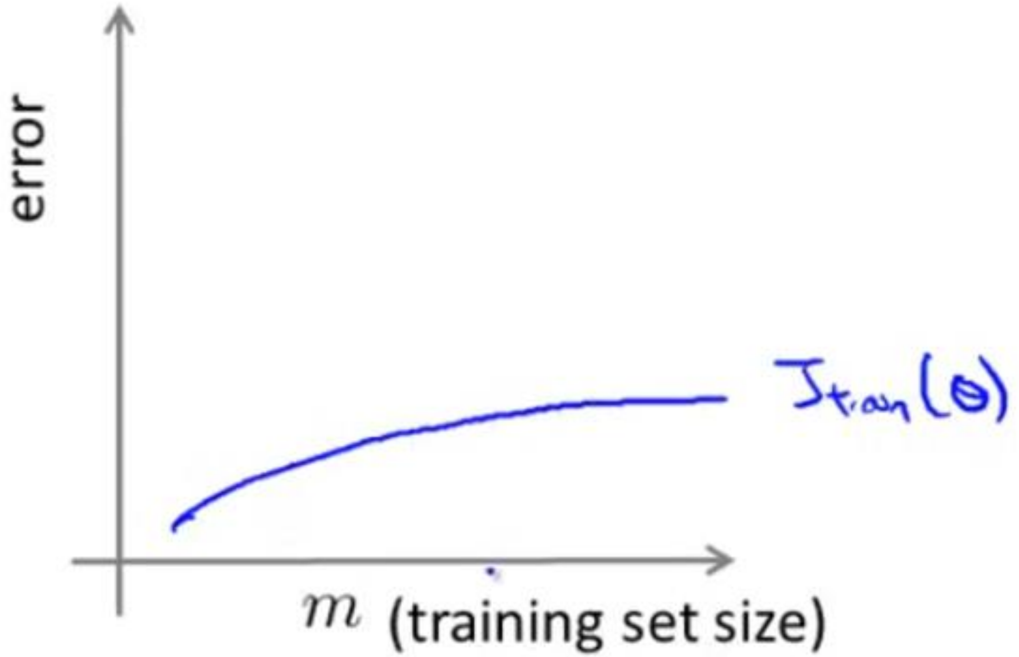


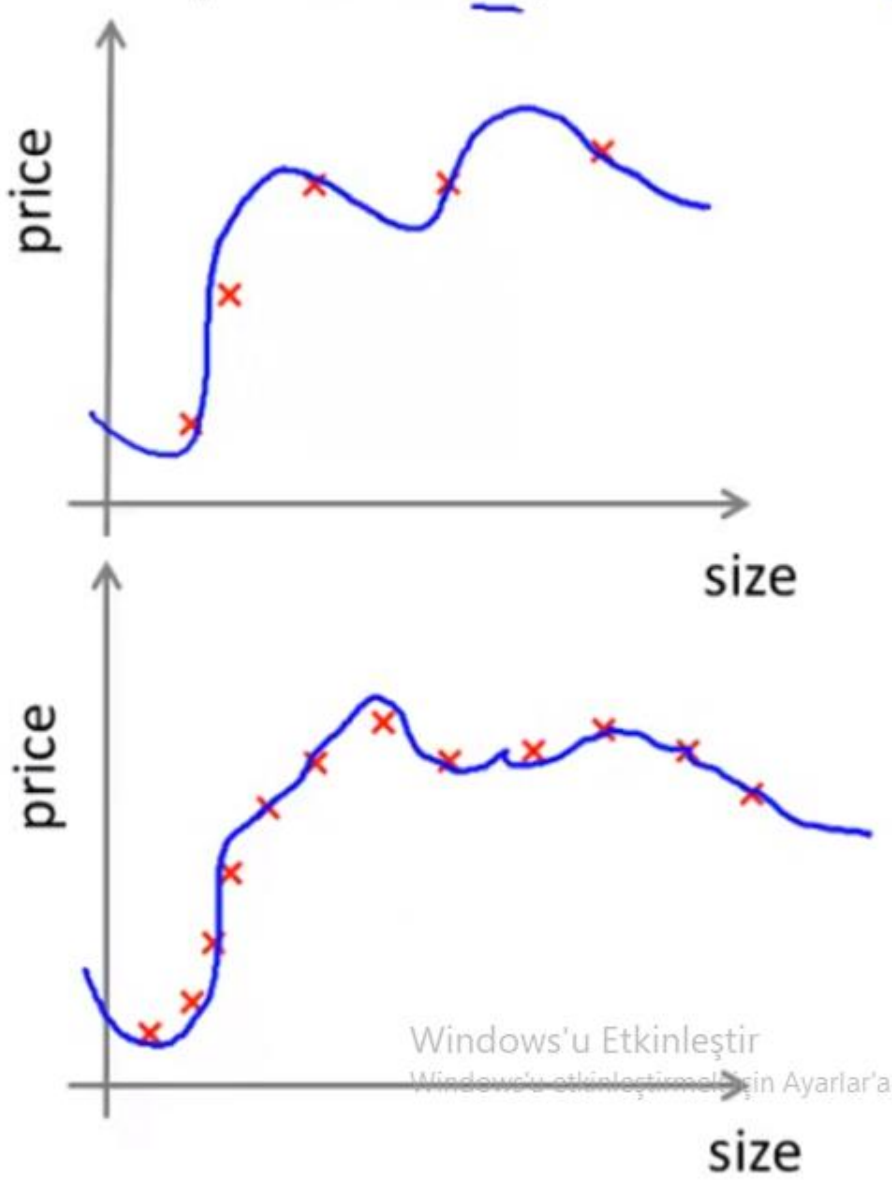


High variance

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$

(and small λ)



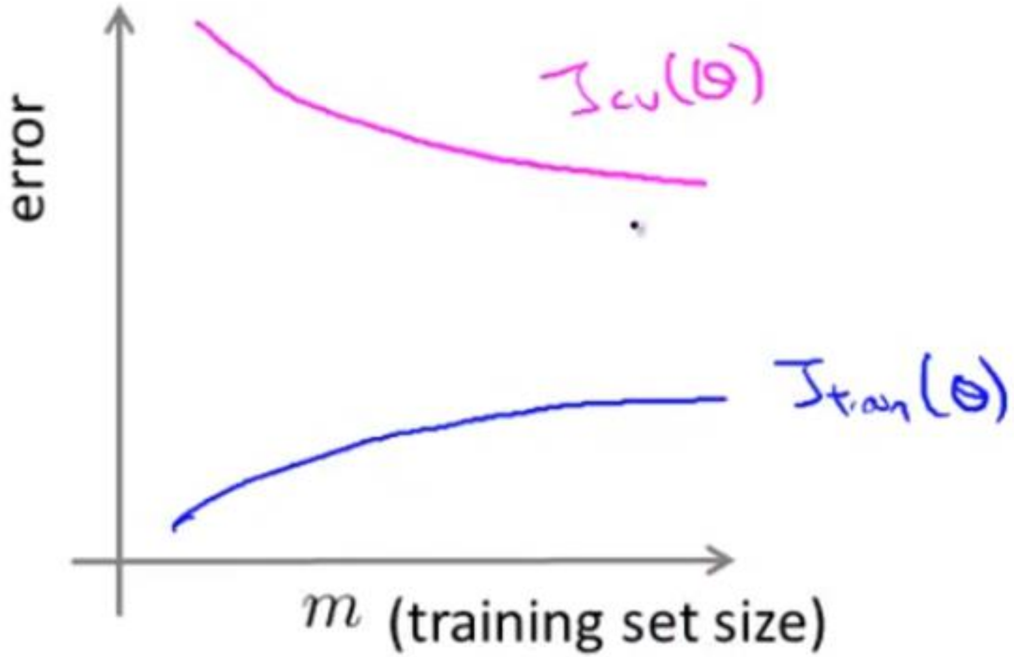


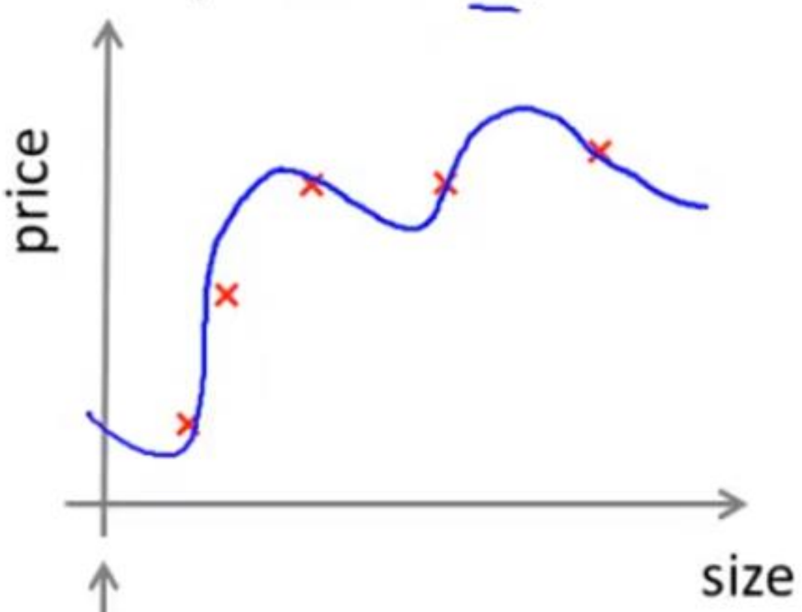
Andrew Ng

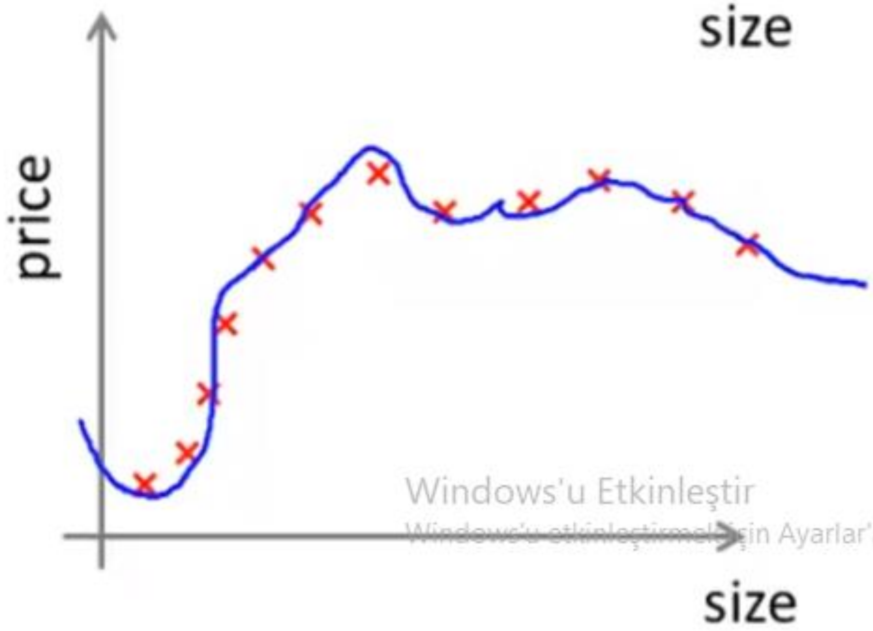
High variance

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$

(and small λ)



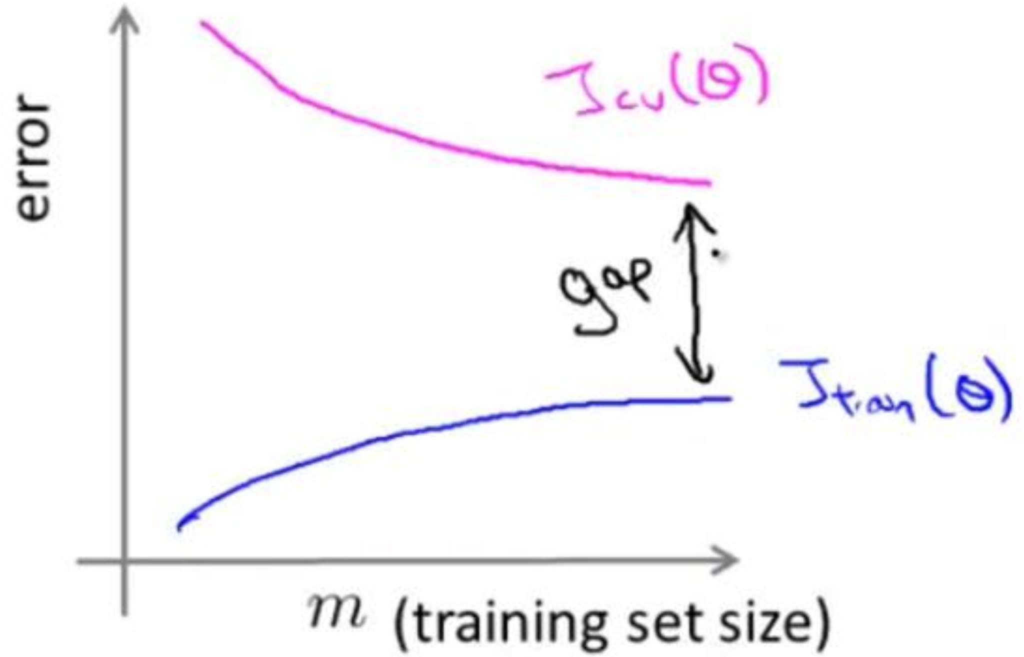


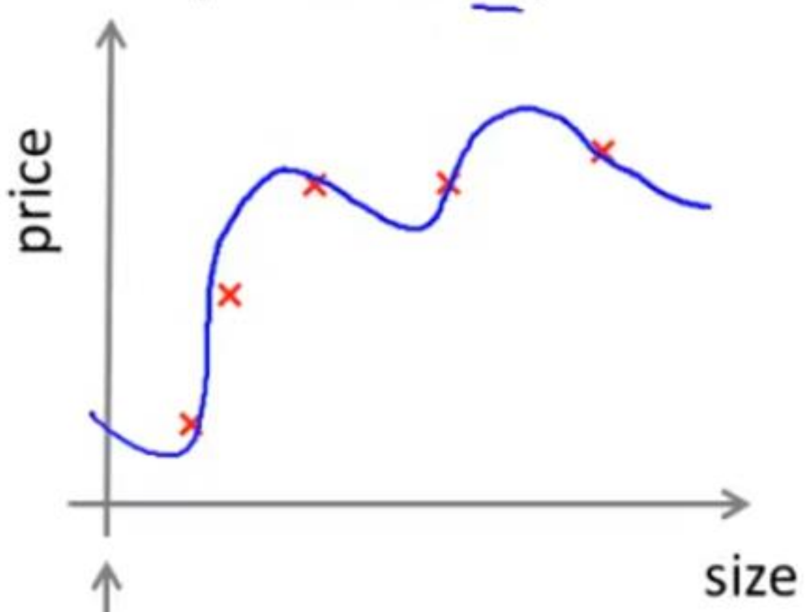


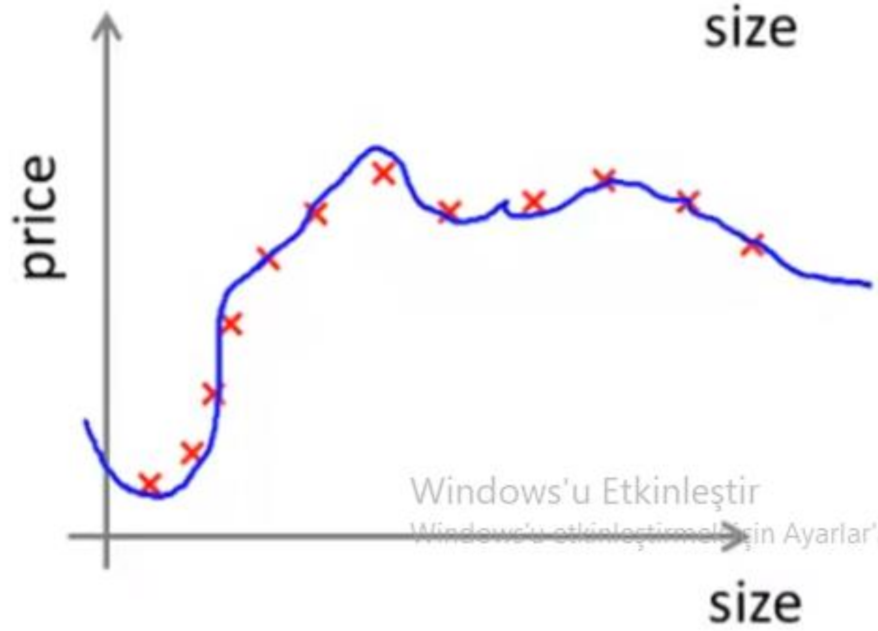
Andrew Ng

High variance

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$
(and small λ)



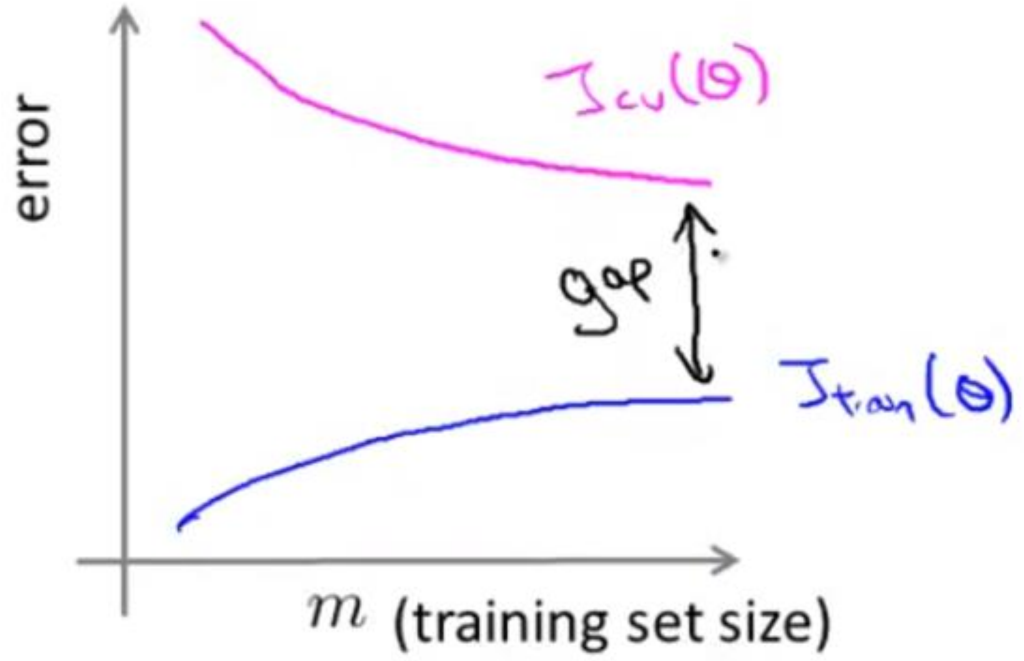


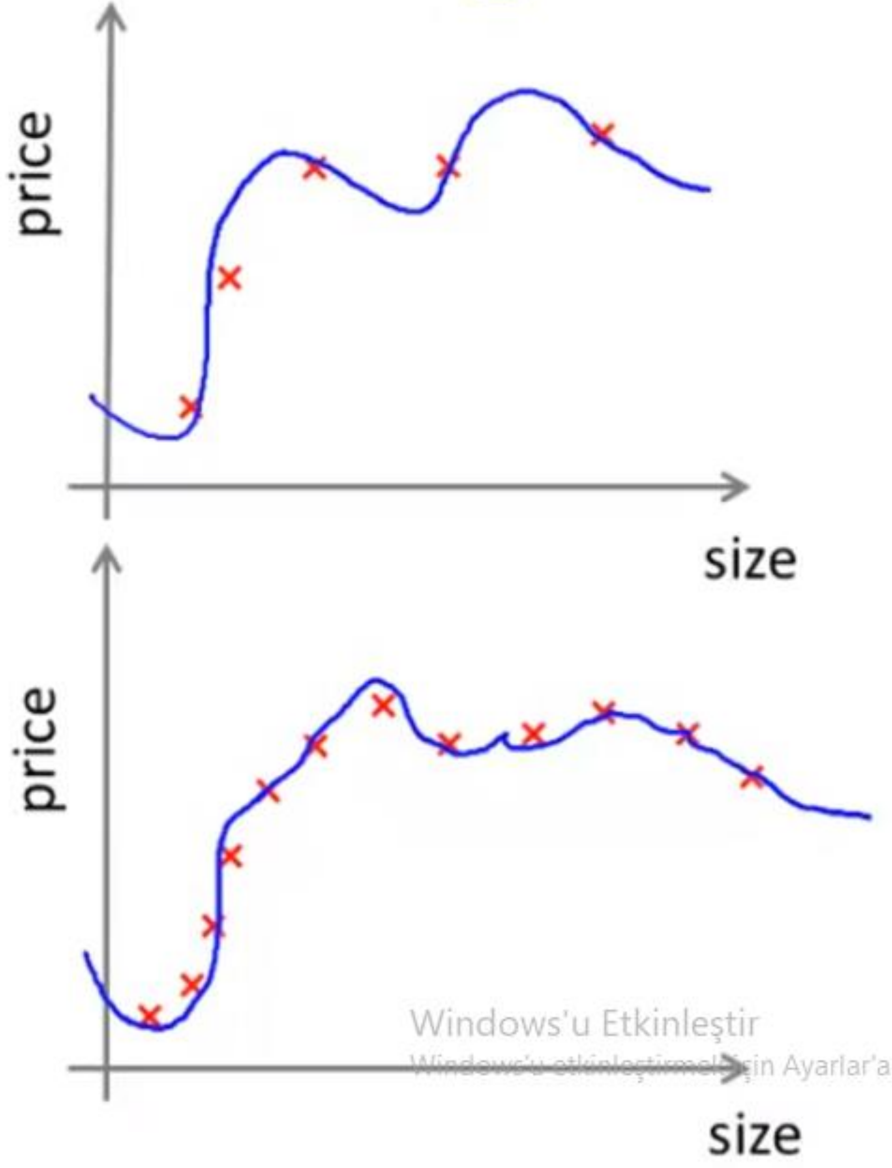


High variance

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$

(and small λ)



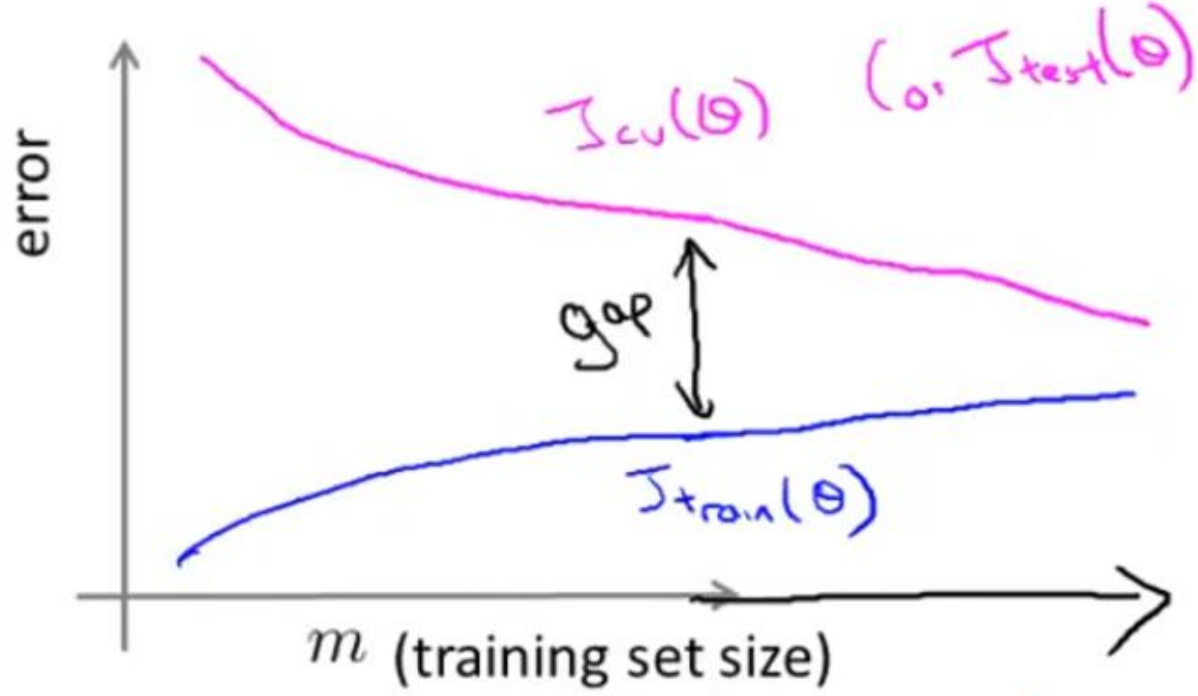


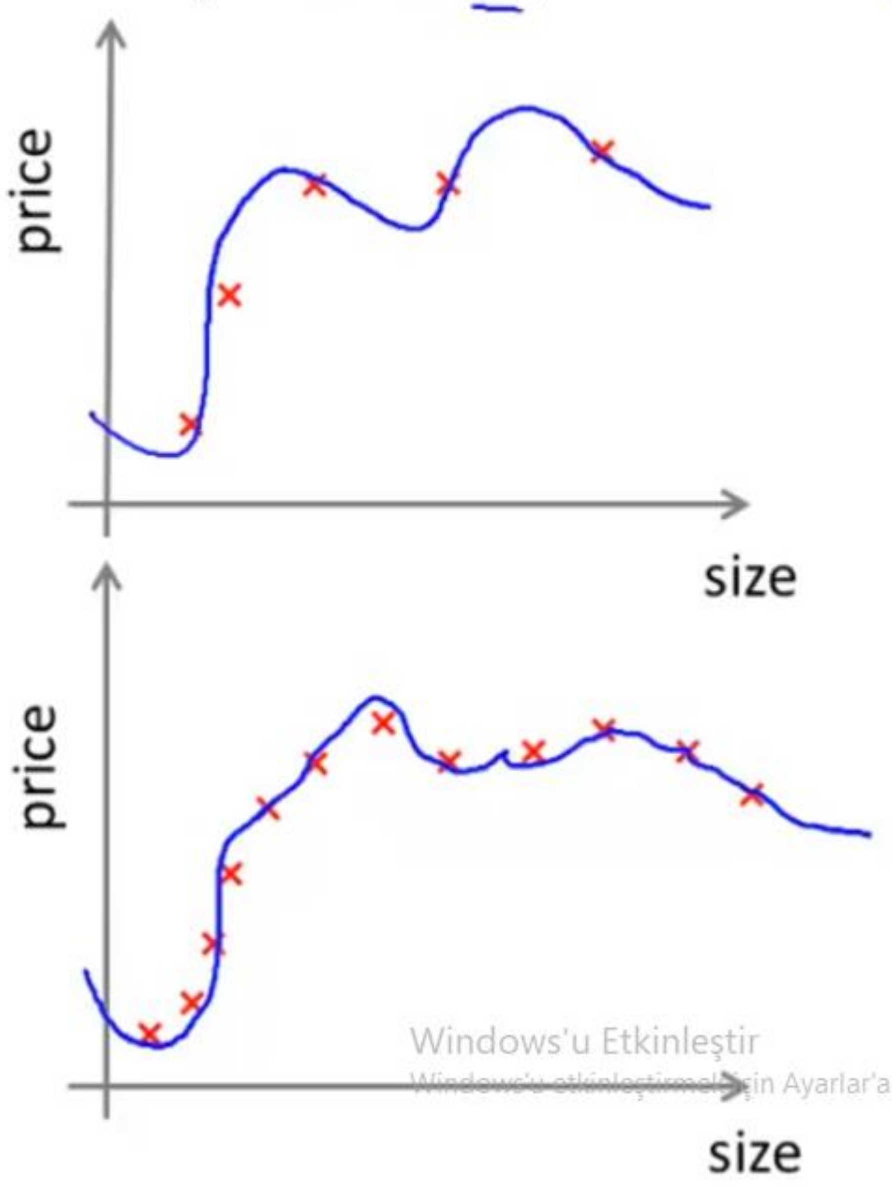
If a learning algorithm is suffering from high variance, getting more training data is likely to help.

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100}x^{100}$$

(and small λ)

High variance





If a learning algorithm is suffering from high variance, getting more training data is likely to help.

Exercise

- In which of the following circumstances is getting more training data likely to significantly help a learning algorithm's performance?
  - ☐ Algorithm is suffering from high bias.
  - ☐ Algorithm is suffering from high variance.
  - ☐ $J_{\text{CV}}(\theta)$ (cross validation error) is much larger than $J_{\text{train}}(\theta)$ (training error).
  - ☐ $J_{\text{CV}}(\theta)$ (cross validation error) is about the same as $J_{\text{train}}(\theta)$ (training error).