Diagnosing ML System

Shih-Yang Su
Virginia Tech

Today's Lectures

- Advice on how getting learning algorithms to different applications
- How to fix your learning algorithm
- Basically ZERO MATH

Debugging a learning algorithm

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- You have built you awesome linear regression model predicting price
- Work perfectly on you testing data

Source: Andrew Ng

Debugging a learning algorithm

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- You have built you awesome linear regression model predicting price
- Work perfectly on you testing data
- Then it fails miserably when you test it on the new data you collected

Source: Andrew Ng

Debugging a learning algorithm

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2$$

- You have built you awesome linear regression model predicting price
- Work perfectly on you testing data
- Then it fails miserably when you test it on the new data you collected
- What to do now?

Source: Andrew Ng

Things You Can Try

- Get more data
- Try different features
- Try tuning your hyperparameter

Things You Can Try

- Get more data
- Try different features
- Try tuning your hyperparameter

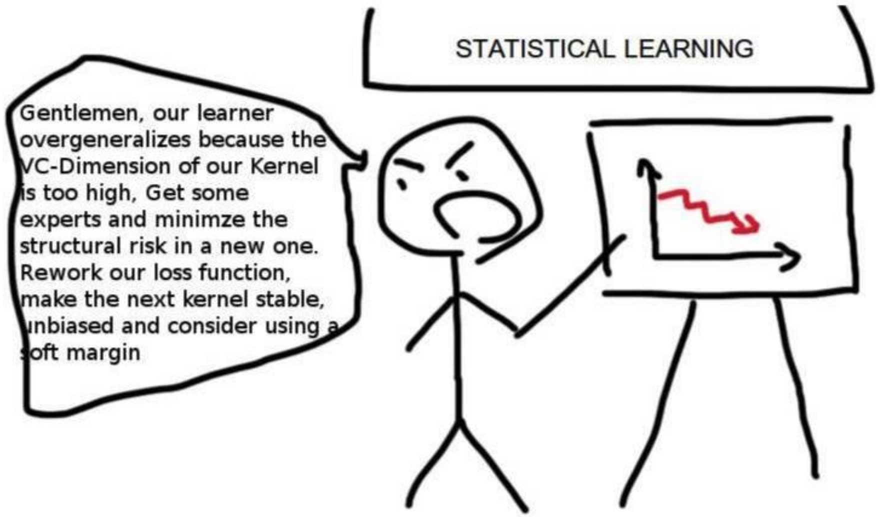
- But which should I try first?

Diagnosing Machine Learning System

- Figure out what is wrong first
- Diagnosing your system takes time, but it can save your time as well
- Ultimate goal: low generalization error

Diagnosing Machine Learning System

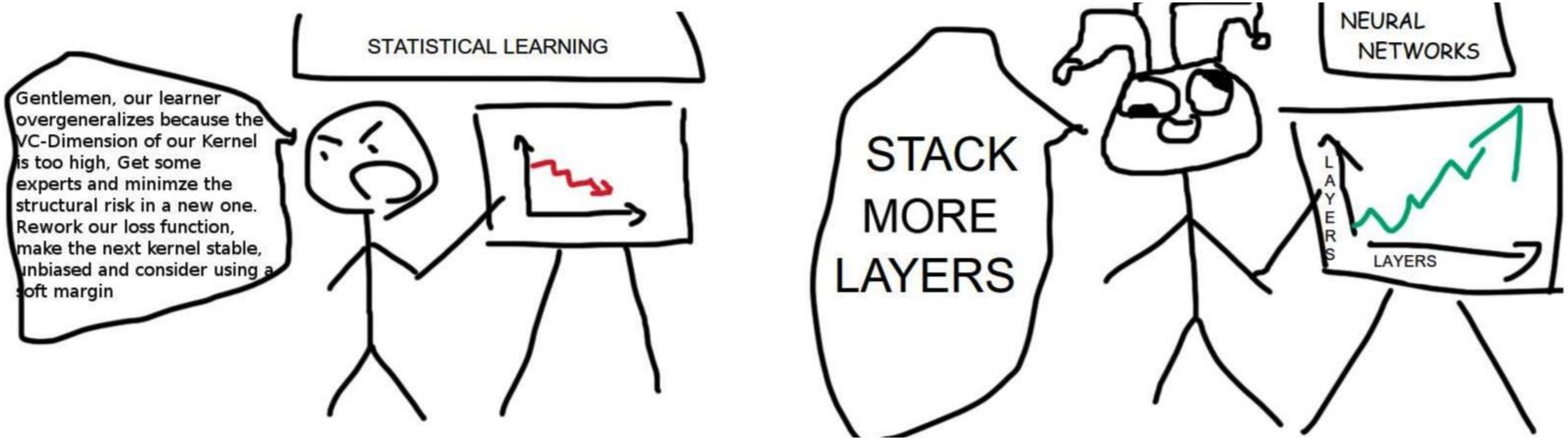
- Figure out what is wrong first
- Diagnosing your system takes time, but it can save your time as well
- Ultimate goal: low generalization error



Source: reddit?

Diagnosing Machine Learning System

- Figure out what is wrong first
- Diagnosing your system takes time, but it can save your time as well
- Ultimate goal: low generalization error

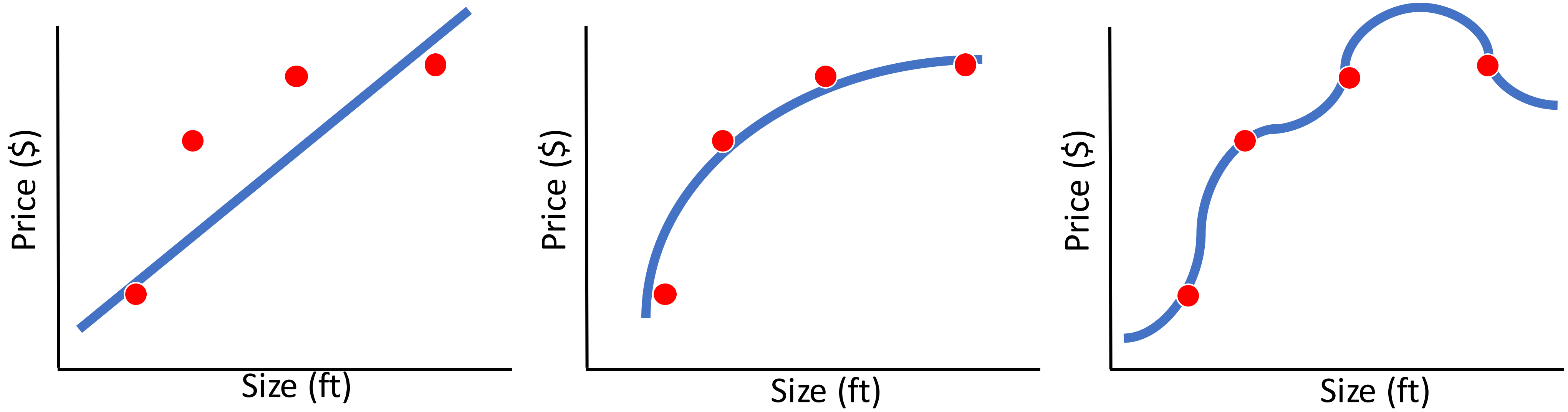


Source: reddit?

Problem: Fail to Generalize

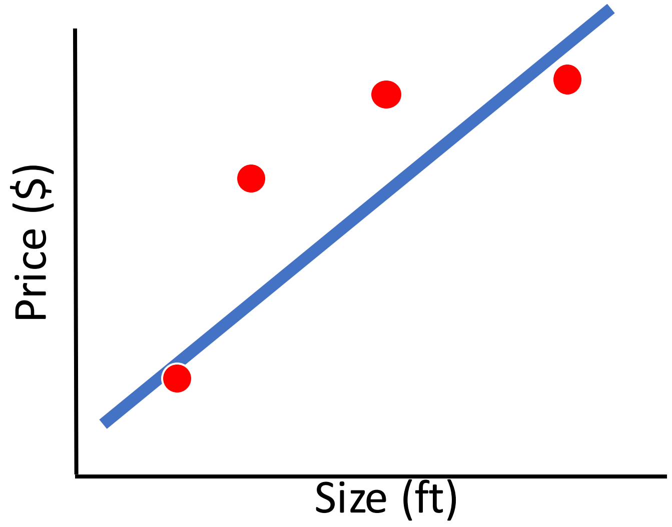
- Model does not generalize to unseen data
  - Fail to predict things that are not in training sample
  - Pick a model that has **lower generalization error**

Evaluate Your Hypothesis



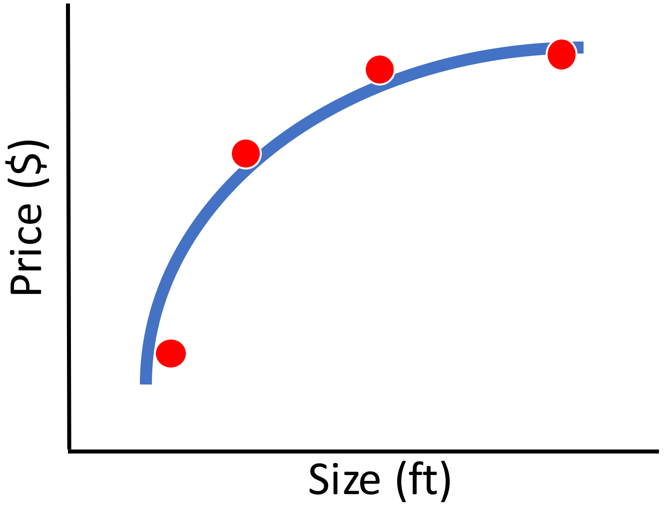
Source: Andrew Ng

Evaluate Your Hypothesis



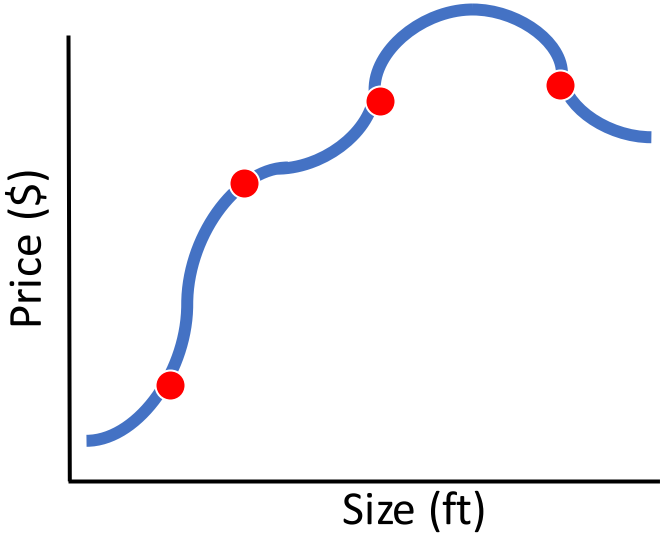
$$\theta_0 + \theta_1 x$$

Underfit



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

Just right

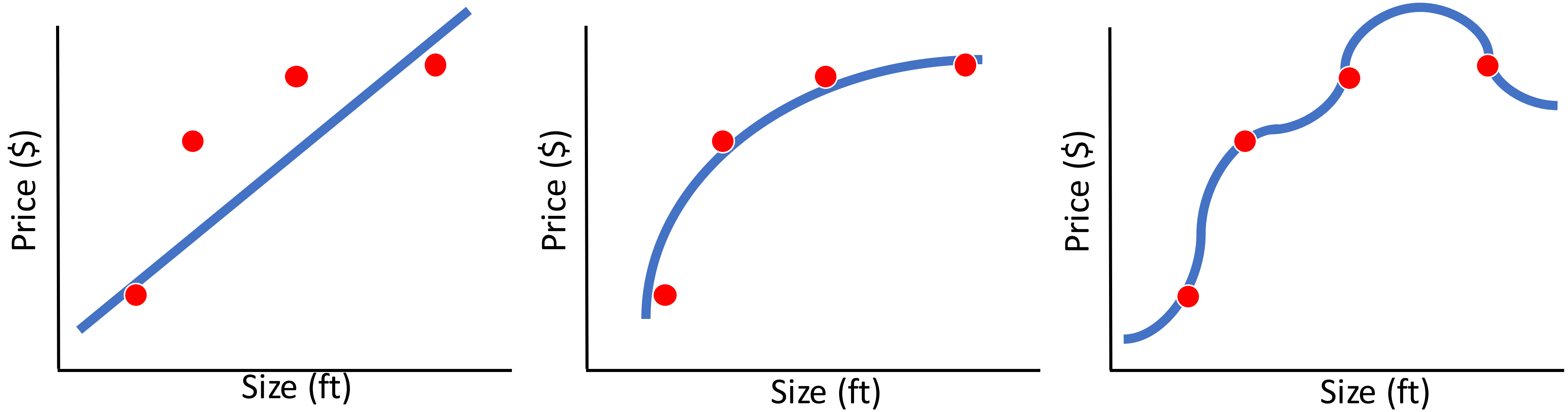


$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

Overfit

Source: Andrew Ng

Evaluate Your Hypothesis



$\theta_0 + \theta_1 x$

Underfit

$\theta_0 + \theta_1 x + \theta_2 x^2$

Just right

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Overfit

- What if the feature dimension is too high?

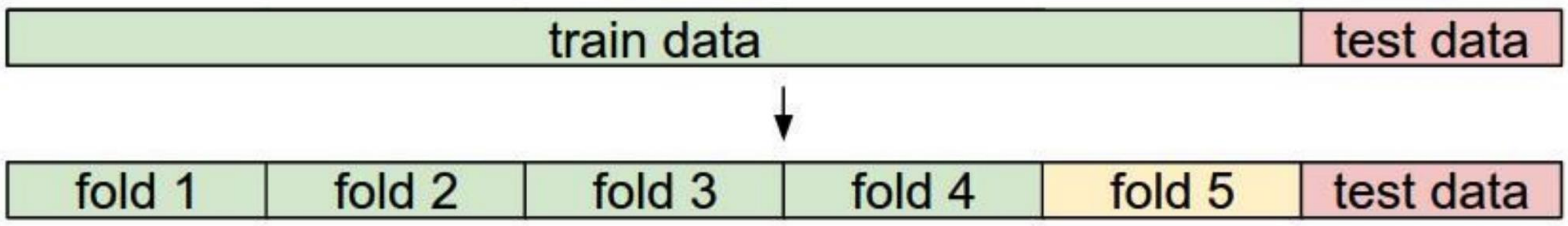
Source: Andrew Ng

Model Selection

- Model does not generalize to unseen data
  - Fail to predict things that are not in training sample
  - Pick a model that has **lower generalization error**

Model Selection

- Model does not generalize to unseen data
  - Fail to predict things that are not in training sample
  - Pick a model that has **lower generalization error**



Model Selection

- Model does not generalize to unseen data
  - Fail to predict things that are not in training sample
  - Pick a model that has **lower generalization error**

- How to evaluate generalization error?
  - Split your data into *train*, *validation*, and *test set*.
  - Use test set error as an *estimator* of generalization error

Model Selection

- Training error

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

- Validation error

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

- Test error

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

Model Selection

- Training error

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

- Validation error

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

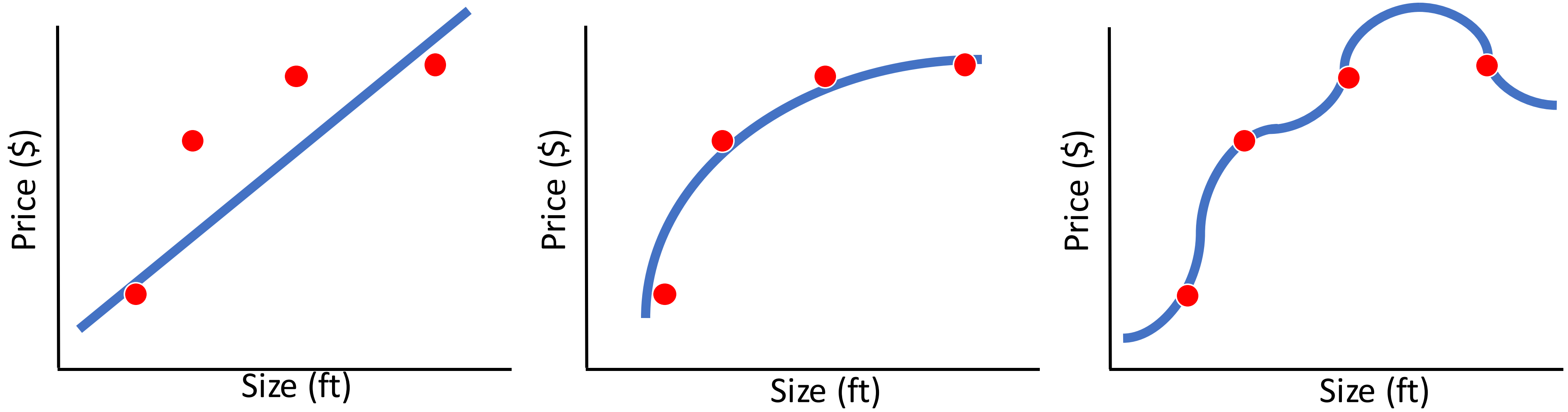
- Test error

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

Procedure:

Step 1. Train on training set
Step 2. Evaluate validation error
Step 3. Pick the best model based on Step 2.
Step 4. Evaluate the test error

Bias/Variance Trade-off



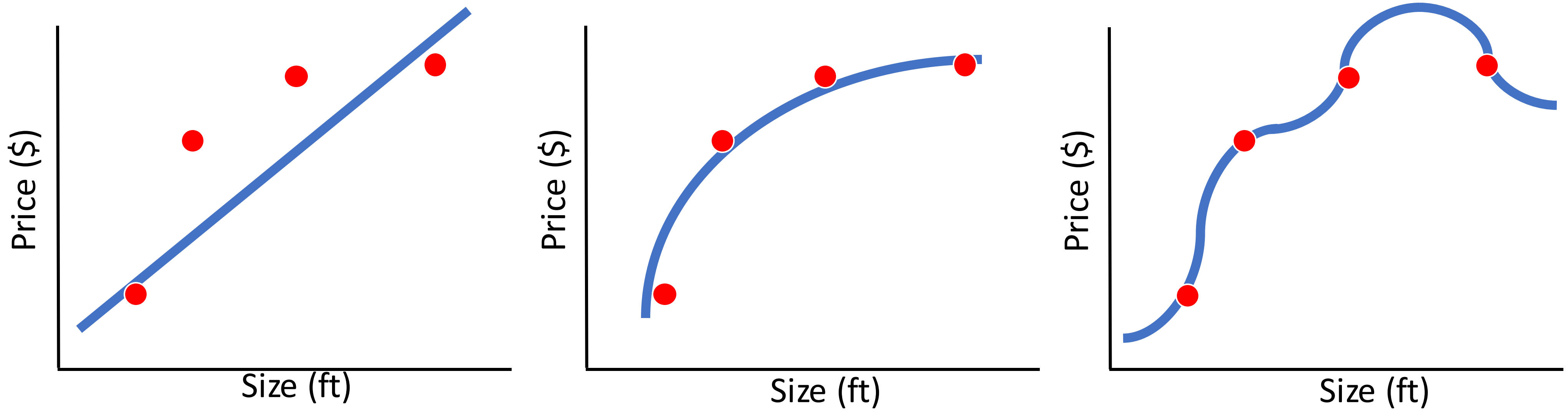
Underfit

Just right

Overfit

Source: Andrew Ng

Bias/Variance Trade-off



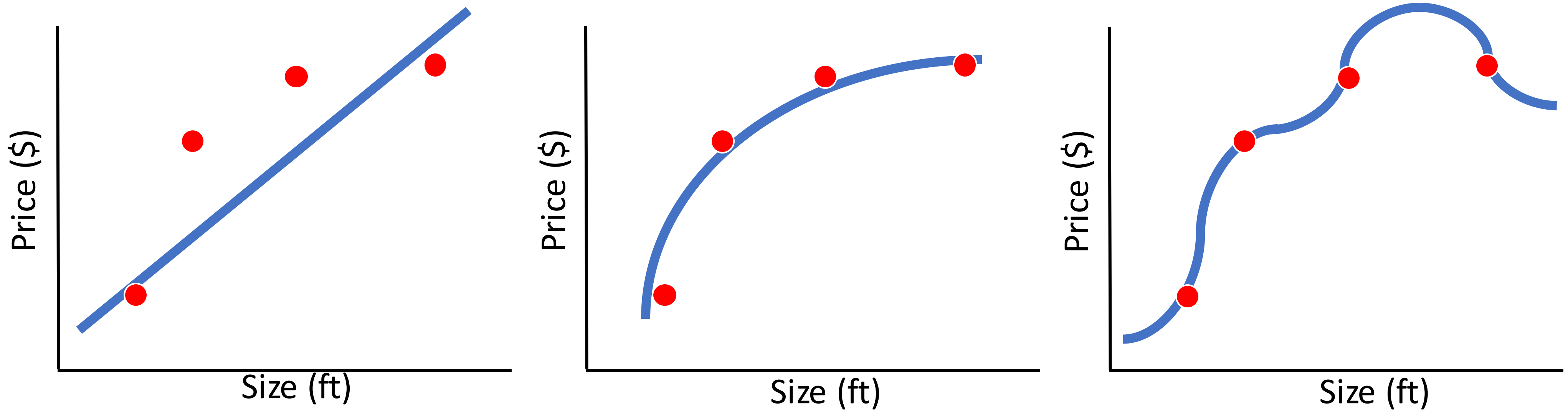
Underfit
High bias

Just right

Overfit
High Variance

Source: Andrew Ng

Bias/Variance Trade-off



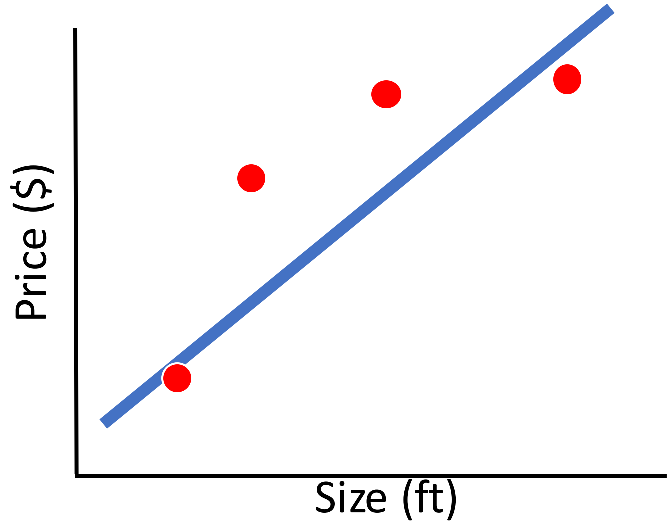
Underfit
High bias
Too simple

Just right

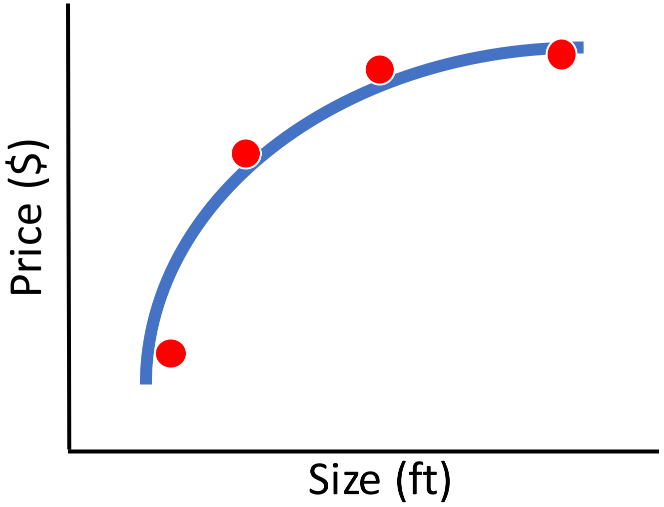
Overfit
High Variance
Too Complex

Source: Andrew Ng

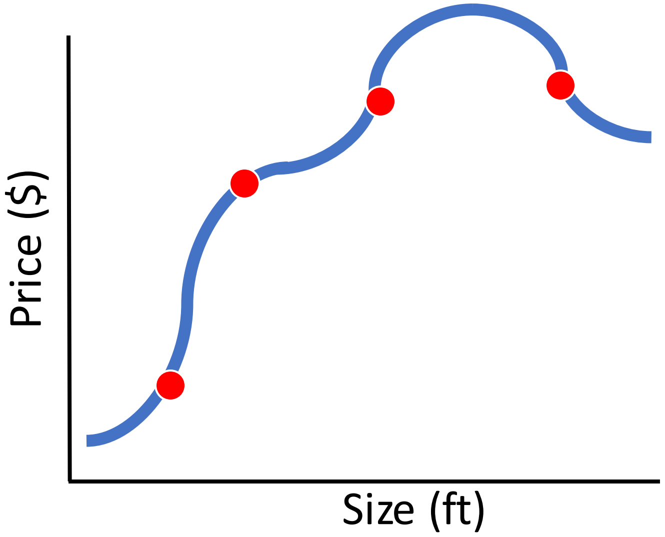
Linear Regression with Regularization



Underfit
High bias
Too simple
Too much regularization



Just right

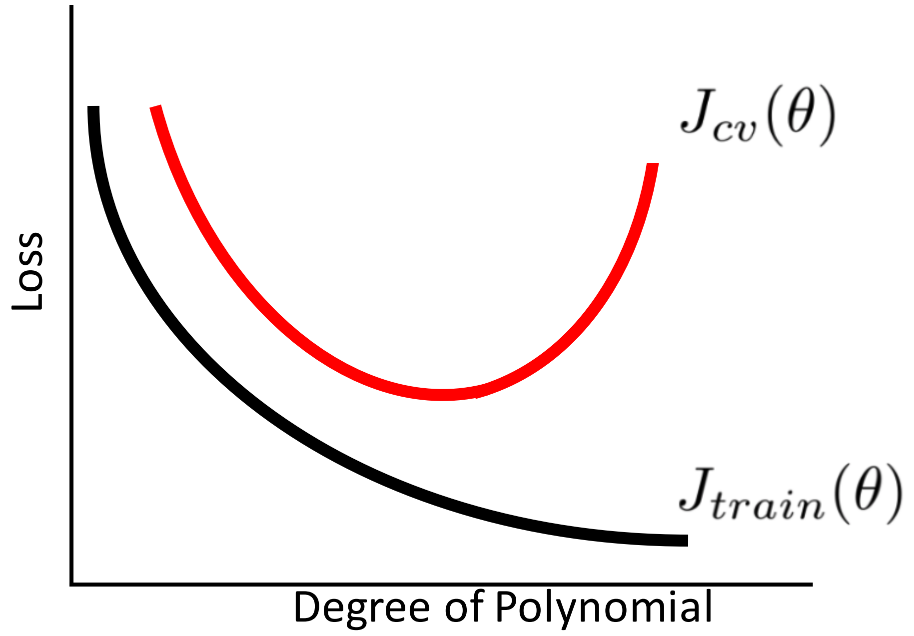


Overfit
High Variance
Too Complex
Too little regularization

Source: Andrew Ng

Bias / Variance Trade-off

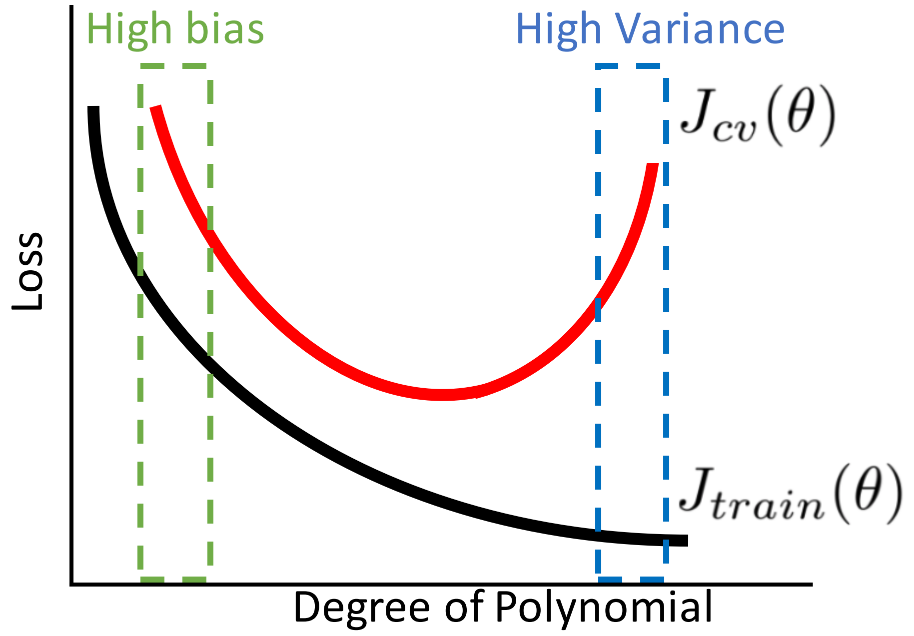
- Training error $J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$
- Cross-validation error $J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$



Source: Andrew Ng

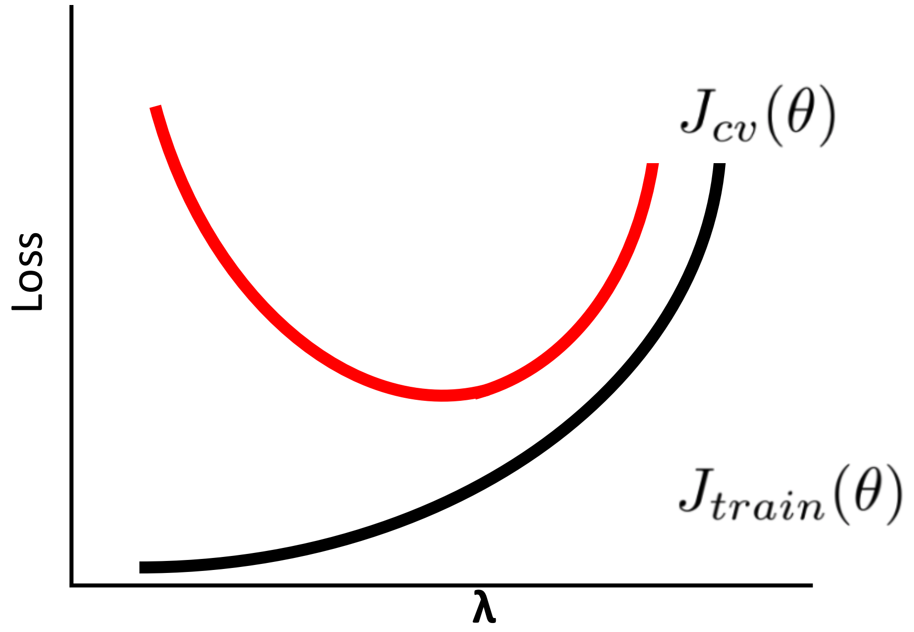
Bias / Variance Trade-off

- Training error $J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$
- Cross-validation error $J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$



Bias / Variance Trade-off with Regularization

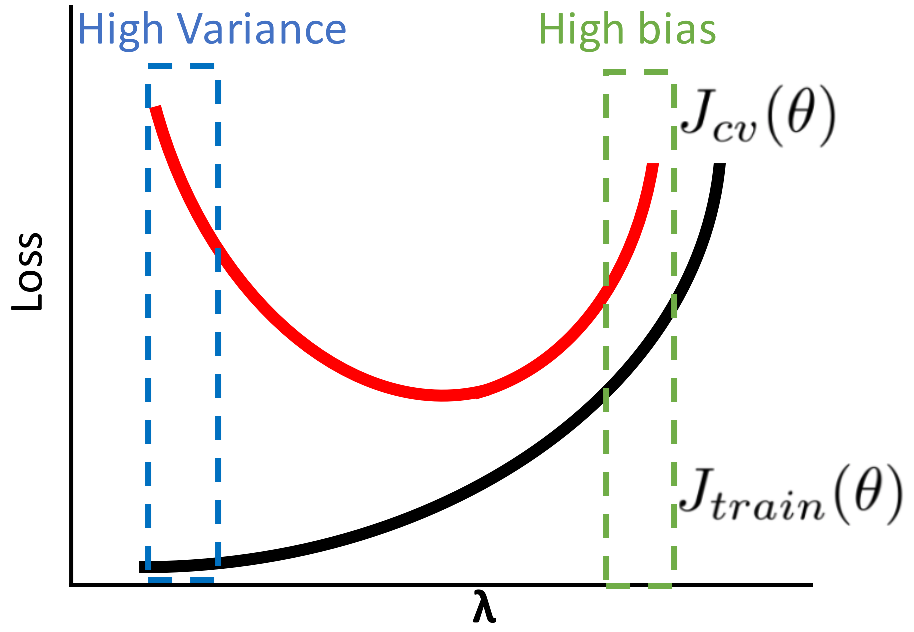
- Training error $J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m}\sum_{j=1}^{m}\theta_j^2$
- Cross-validation error $J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2 + \frac{\lambda}{2m_{cv}}\sum_{j=1}^{m_{cv}}\theta_j^2$



Source: Andrew Ng

Bias / Variance Trade-off with Regularization

- Training error $J_{train}(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m}\sum_{j=1}^{m}\theta_j^2$
- Cross-validation error $J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2 + \frac{\lambda}{2m_{cv}}\sum_{j=1}^{m_{cv}}\theta_j^2$



Source: Andrew Ng

Problem: Fail to Generalize

- Should we get more data?

Problem: Fail to Generalize

- Should we get more data?

- Getting more data does not always help

Problem: Fail to Generalize

- Should we get more data?

- Getting more data does not always help

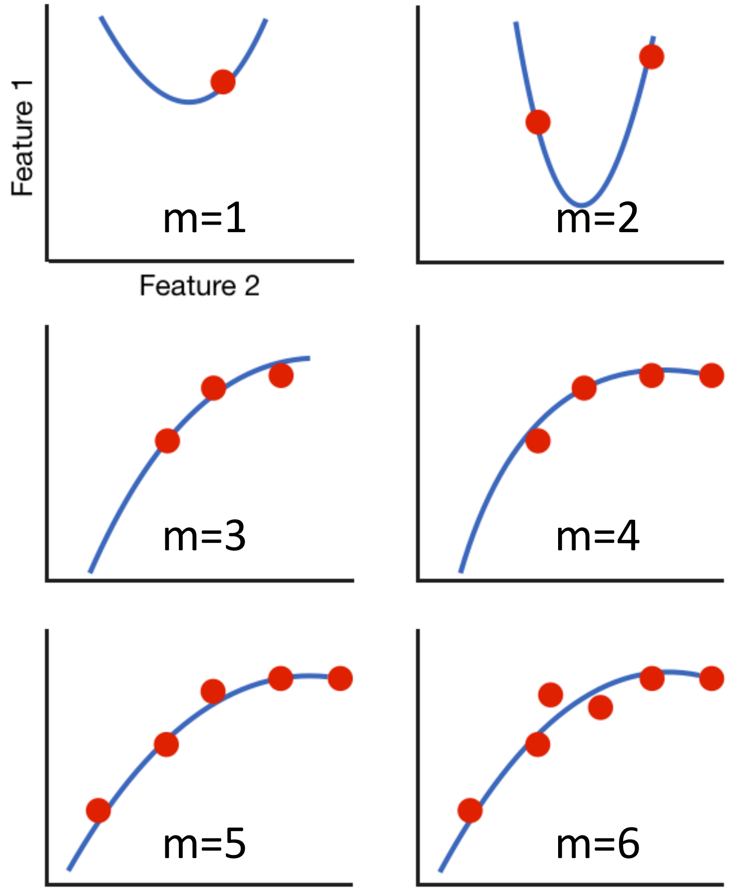
- How do we know if we should collect more data?

Learning Curve

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

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

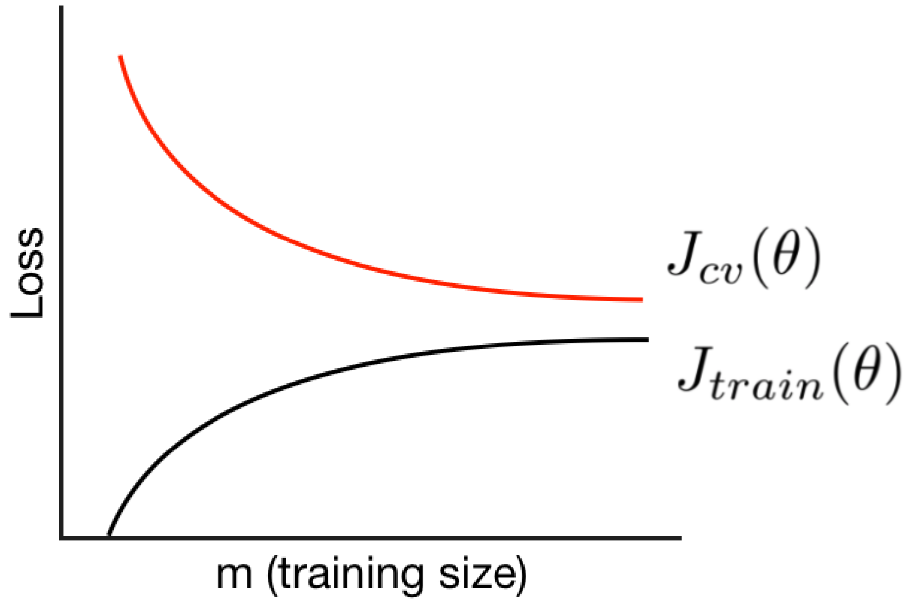
$$\theta_0 + \theta_1 x + \theta_2 x^2$$



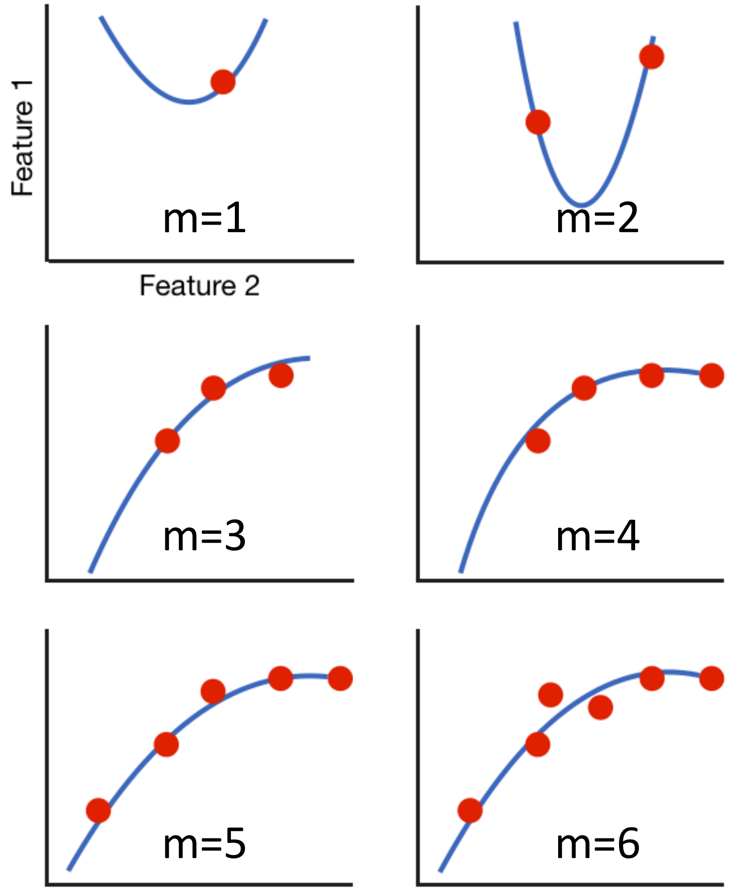
Learning Curve

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

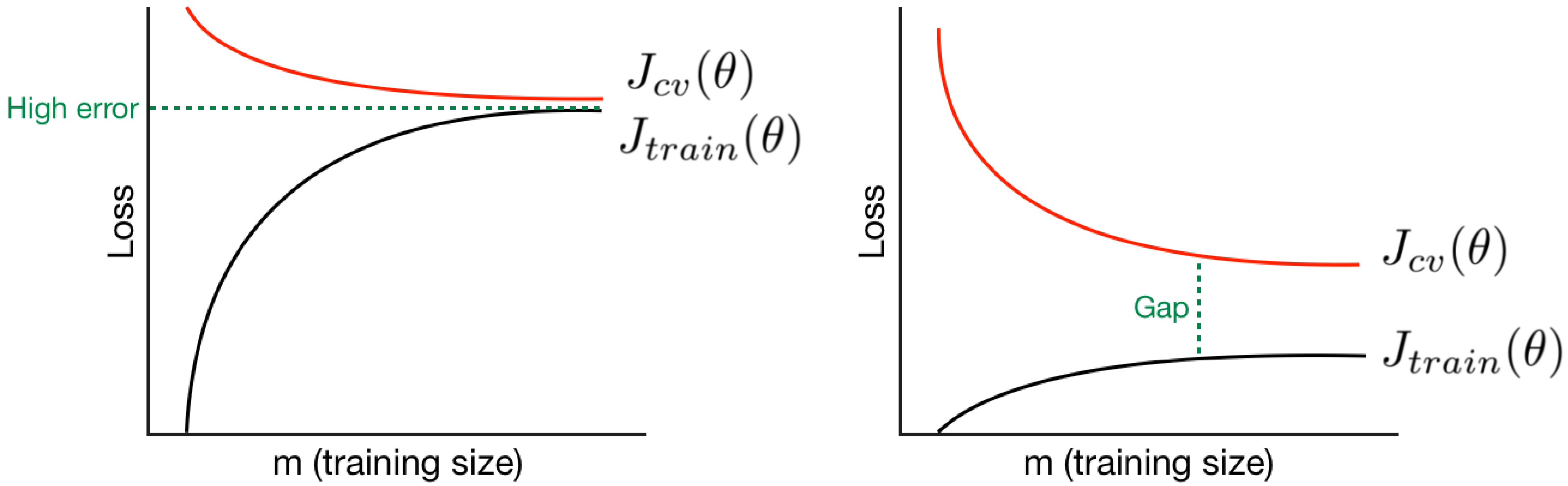
$$J_{cv}(\theta) = \frac{1}{2m_{cv}}\sum_{i=1}^{m_{cv}}(h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



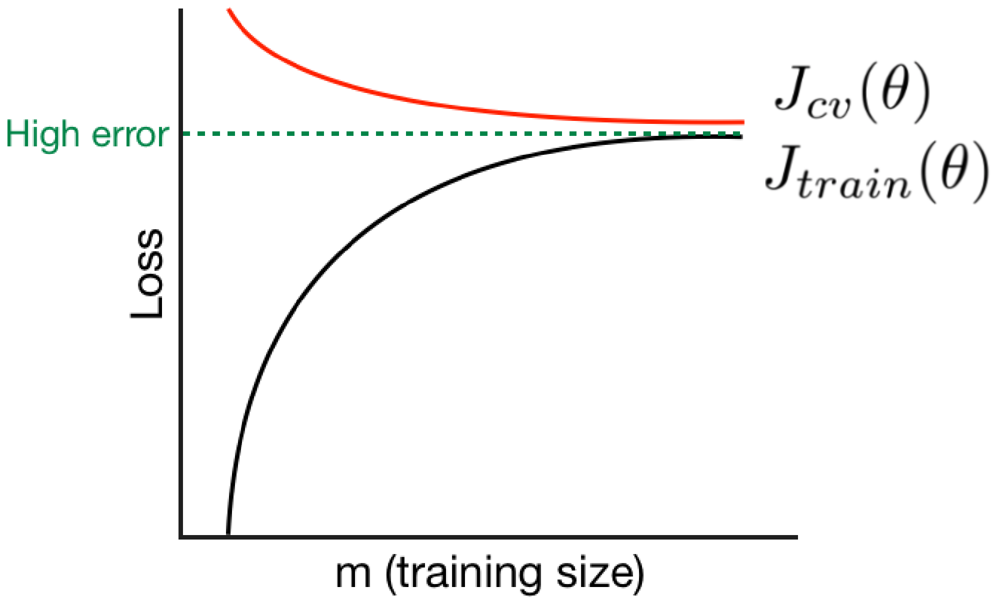
$\theta_0 + \theta_1 x + \theta_2 x^2$



Learning Curve

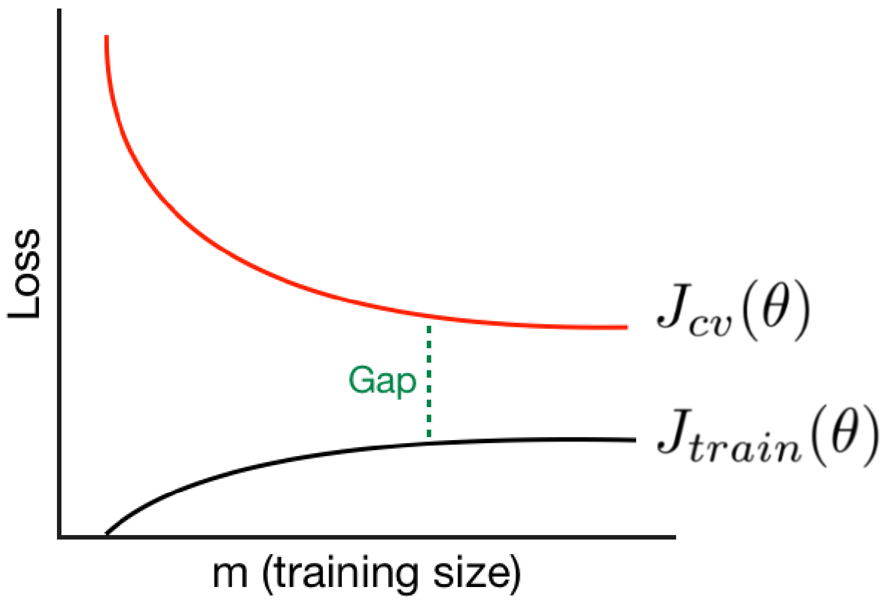


Learning Curve



Underfit

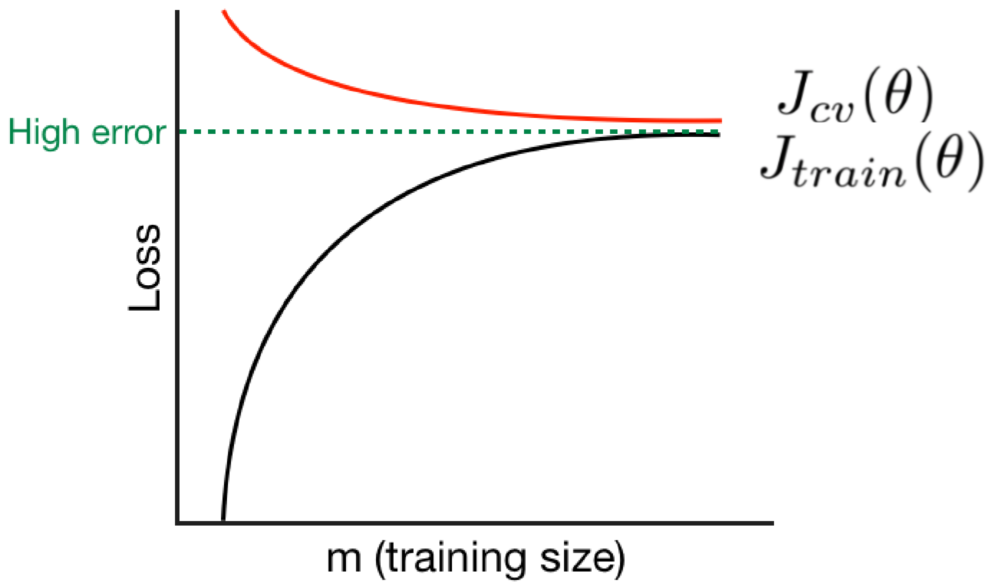
High bias



Overfit

High Variance

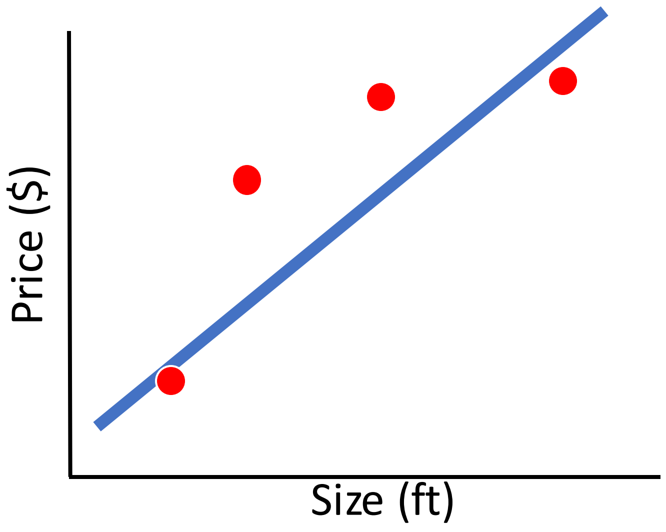
Learning Curve



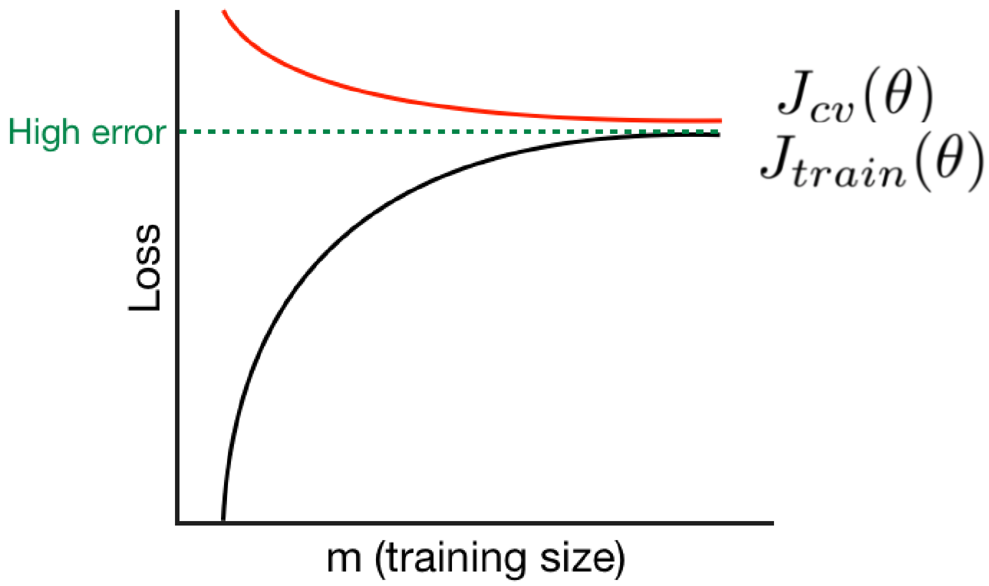
Underfit

High bias

Does adding more data help?



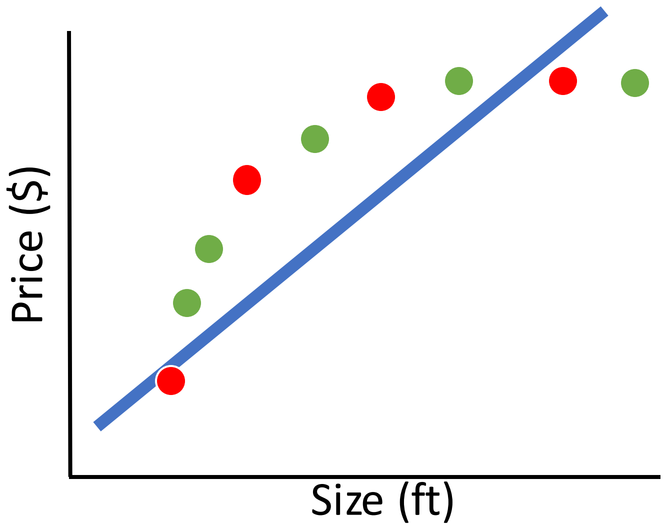
Learning Curve



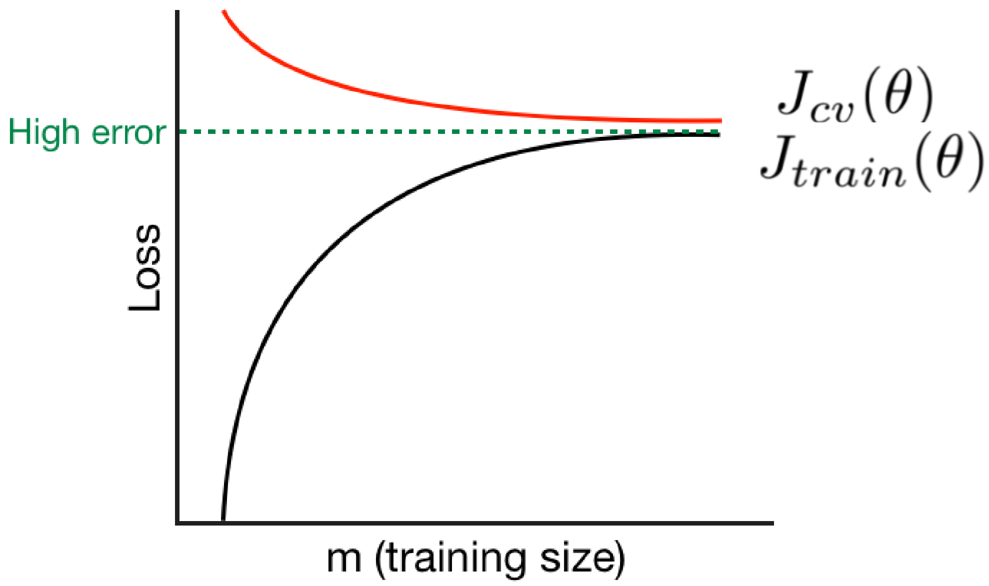
Underfit

High bias

Does adding more data help?



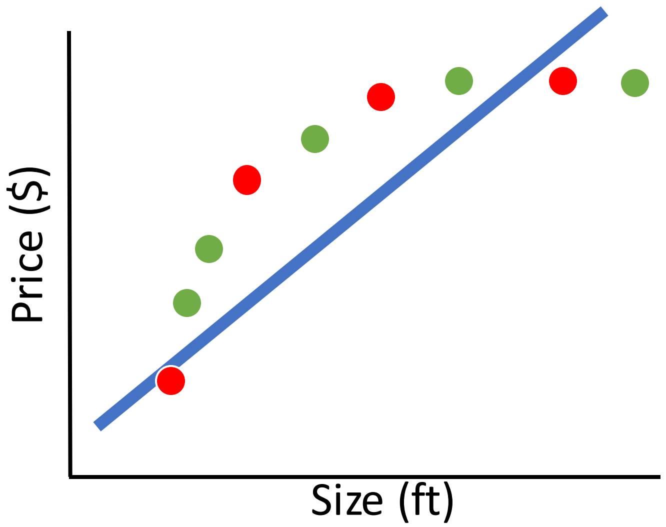
Learning Curve



Underfit

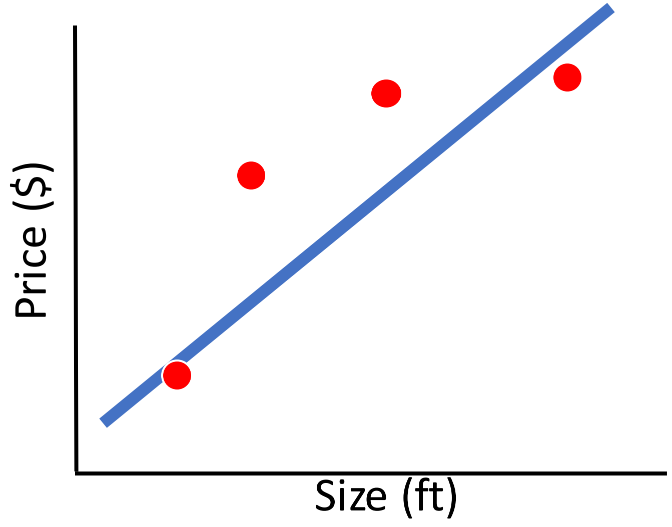
High bias

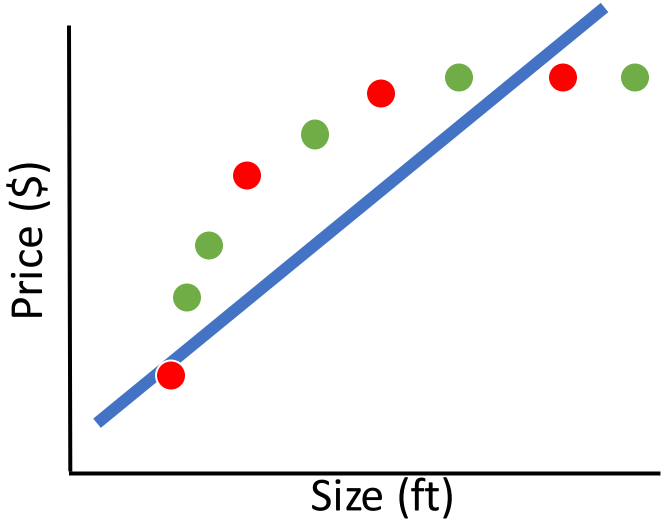
Does adding more data help?



Learning Curve

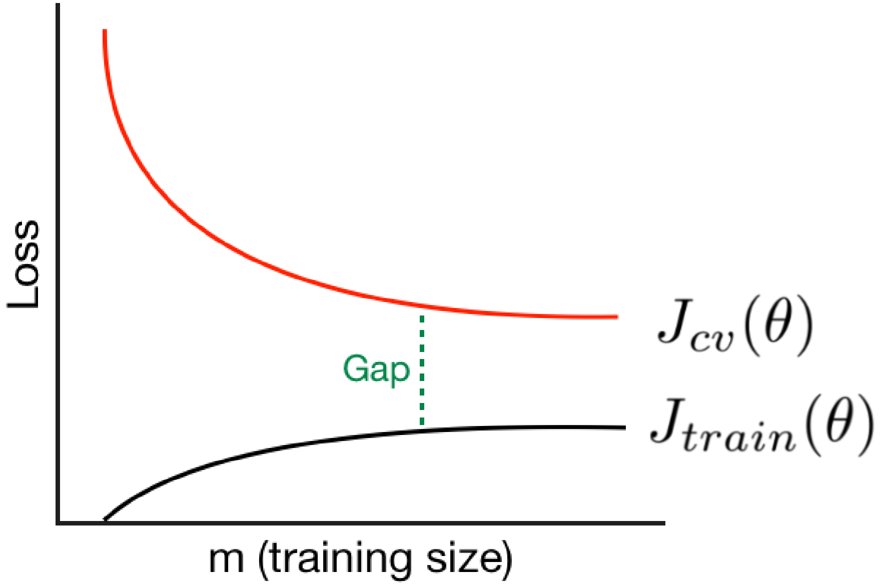
Does adding more data help?





More data doesn't help when your model has high bias

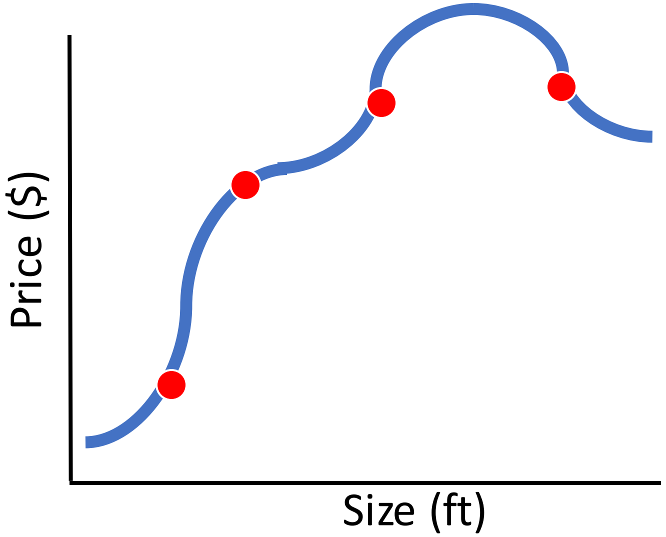
Learning Curve



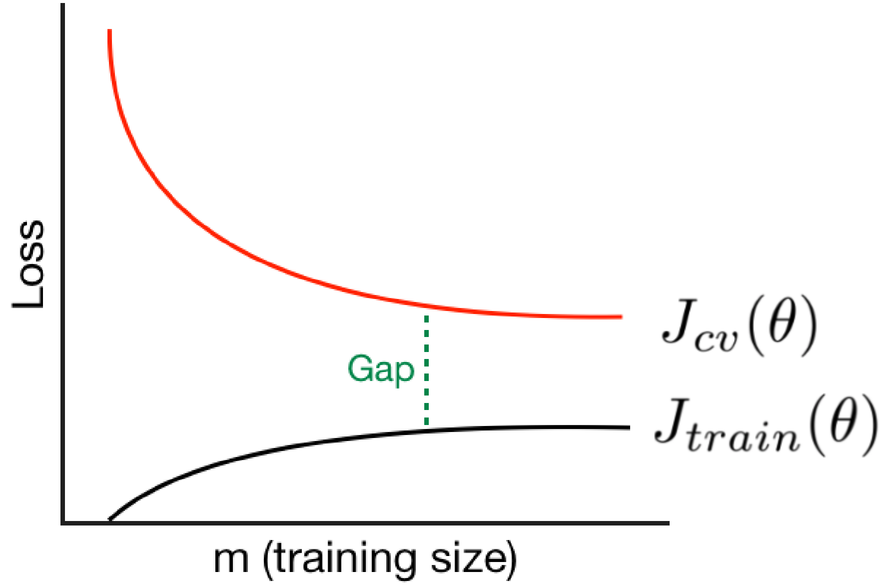
Overfit

High Variance

Does adding more data help?

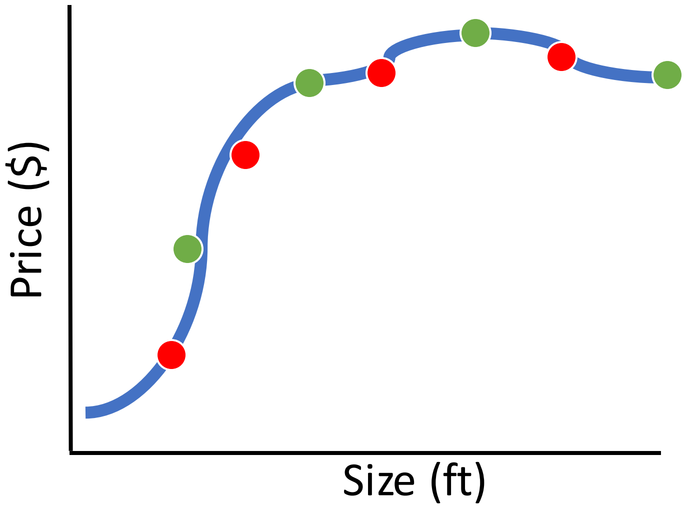


Learning Curve



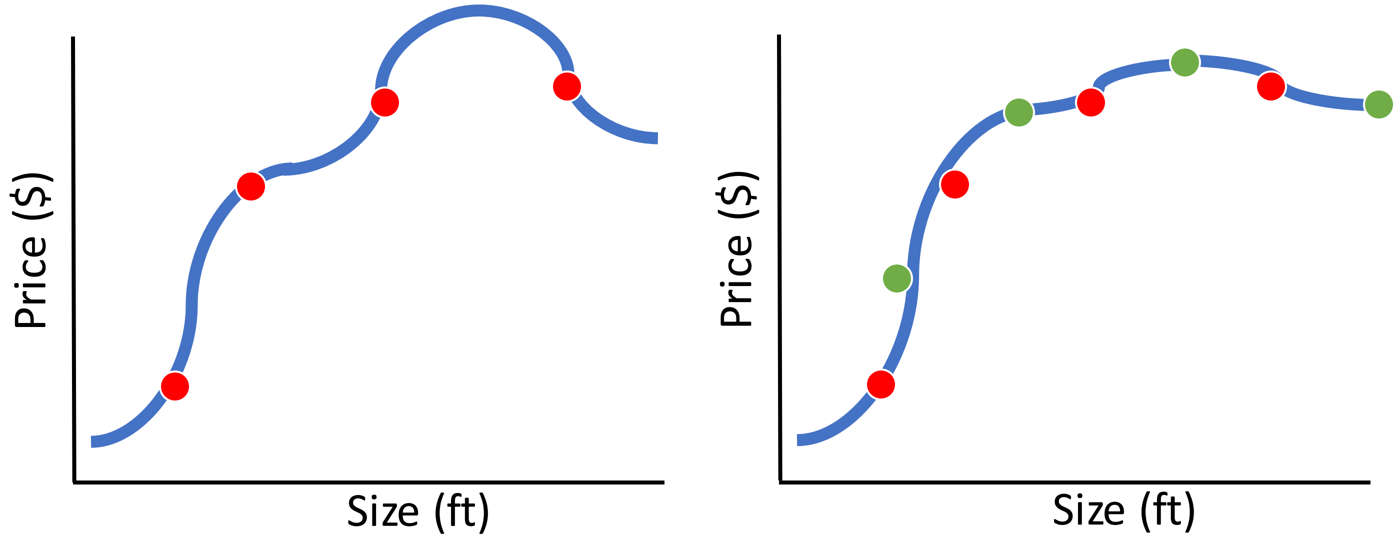
Overfit
High Variance

Does adding more data help?



Learning Curve

Does adding more data help?



More data is likely to help when your model has high variance

Things You Can Try

- Get more data
  - When you have high variance
- Try different features
  - Adding feature helps fix high bias
  - Using smaller sets of feature fix high variance
- Try tuning your hyperparameter
  - Decrease regularization when bias is high
  - Increase regularization when variance is high

Things You Can Try

- Get more data
  - When you have high variance
- Try different features
  - Adding feature helps fix high bias
  - Using smaller sets of feature fix high variance
- Try tuning your hyperparameter
  - Decrease regularization when bias is high
  - Increase regularization when variance is high

Analyze your model before you act