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

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \]

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

Source: Andrew Ng
Debugging a learning algorithm

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \]

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

Source: Andrew Ng
Debugging a learning algorithm

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

• You have built your awesome linear regression model predicting price
• Work perfectly on your 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
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

- 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

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

• Validation error

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

• Test error

\[ J_{\text{test}}(\theta) = \frac{1}{2m_{\text{test}}} \sum_{i=1}^{m_{\text{test}}} (h_{\theta}(x_{\text{test}}^{(i)}) - y_{\text{test}}^{(i)})^2 \]
Model Selection

• Training error

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

• Validation error

\[ J_{\text{cv}}(\theta) = \frac{1}{2m_{\text{cv}}} \sum_{i=1}^{m_{\text{cv}}} (h_{\theta}(x^{(i)}_{\text{cv}}) - y^{(i)}_{\text{cv}})^2 \]

• Test error

\[ J_{\text{test}}(\theta) = \frac{1}{2m_{\text{test}}} \sum_{i=1}^{m_{\text{test}}} (h_{\theta}(x^{(i)}_{\text{test}}) - y^{(i)}_{\text{test}})^2 \]

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

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

- **Cross-validation error**
  \[ J_{\text{cv}}(\theta) = \frac{1}{2m_{\text{cv}}} \sum_{i=1}^{m_{\text{cv}}} (h_{\theta}(x_{\text{cv}}^{(i)}) - y_{\text{cv}}^{(i)})^2 \]

Source: Andrew Ng
Bias / Variance Trade-off

- Training error

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

- Cross-validation error

\[ J_{\text{cv}}(\theta) = \frac{1}{2m_{\text{cv}}} \sum_{i=1}^{m_{\text{cv}}} (h_{\theta}(x^{(i)}_{\text{cv}}) - y^{(i)}_{\text{cv}})^2 \]
Bias / Variance Trade-off with Regularization

- Training error
  \[ J_{\text{train}}(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^{m} \theta_j^2 \]

- Cross-validation error
  \[ J_{\text{cv}}(\theta) = \frac{1}{2m_{\text{cv}}} \sum_{i=1}^{m_{\text{cv}}} (h_{\theta}(x_{\text{cv}}^{(i)}) - y_{\text{cv}}^{(i)})^2 + \frac{\lambda}{2m_{\text{cv}}} \sum_{j=1}^{m_{\text{cv}}} \theta_j^2 \]

Source: Andrew Ng
Bias / Variance Trade-off with Regularization

- Training error
  \[ J_{\text{train}}(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^{m} \theta_j^2 \]

- Cross-validation error
  \[ J_{\text{cv}}(\theta) = \frac{1}{2m_{\text{cv}}} \sum_{i=1}^{m_{\text{cv}}} (h_\theta(x^{(i)}_{\text{cv}}) - y^{(i)}_{\text{cv}})^2 + \frac{\lambda}{2m_{\text{cv}}} \sum_{j=1}^{m_{\text{cv}}} \theta_j^2 \]

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

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

\[ J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} \left( h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)} \right)^2 \]
Learning Curve

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

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

\[ \theta_0 + \theta_1 x + \theta_2 x^2 \]
Learning Curve

- $J_{cv}(\theta)$
- $J_{train}(\theta)$
Learning Curve

Underfit
High bias

Overfit
High Variance
Learning Curve

Does adding more data help?

$J_{cv}(\theta)$
$J_{train}(\theta)$

Loss

$m$ (training size)

Underfit

High bias

Price ($\$$)

Size (ft)
Learning Curve

Underfit

High bias

Does adding more data help?

$J_{\text{test}}(\theta)$

$J_{\text{train}}(\theta)$

Loss

m (training size)

Underfit

High bias

Price ($\) vs. Size (ft)
Learning Curve

Does adding more data help?

$J_{cv}(\theta)$

$J_{train}(\theta)$

High bias

High error

Price (\$)

Size (ft)
Does adding more data help?

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

Does adding more data help?

\[ J_{cv}(\theta) \]
\[ J_{train}(\theta) \]

m (training size)

Overfit

High Variance

Gap

Price ($)

Size (ft)
Does adding more data help?

\[ J_{cv}(\theta) \]
\[ J_{train}(\theta) \]
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