Applying ML

Debugging a learning algo

- get more training examples
- try smaller set of features (carefully select)
- try getting additional features
- try adding polynomial features
- try decreasing λ
- try increasing λ

Machine Learning Diagnostic

Evaluating a Hypothesis

Dataset split (70/30)

training/test procedure for lin regression

- Learn param θ from training data
- Compute test set error

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

procedure for logistic regression

- learn param θ from training data
- Compute error

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

- Misclassification Error (0/1 ME)

$$err(h_\theta(x), y) = \begin{cases} 1 & \text{if } h_\theta(x) \ge 0.5,\ y = 0 \\ & \text{or if } h_\theta(x) < 0.5,\ y = 1 \\ 0 & \text{otherwise} \end{cases} \Big\}\ \text{error}$$

Model Selection Problems

d = degree of polynomial

1. $h_\theta(x) = \theta_0 + \theta_1 x \longrightarrow \theta^{(1)} \rightarrow J_{test}(\theta^{(1)})$ $\quad d=1$
2. $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \rightarrow \theta^{(2)} \rightarrow J_{test}(\theta^{(2)})$ $\quad d=2$
3. $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \longrightarrow \theta^{(3)} \rightarrow J_{test}(\theta^{(3)})$ $\quad d=3$

$\vdots$

10. $h_\theta(x) = \theta_0 + \theta_1 x + \dots + \theta_{10} x^{10} \rightarrow \theta^{(10)} \rightarrow J_{test}(\theta^{(10)})$ $\quad d=10$

Choose $\theta_0 + \dots \theta_5 x^5$

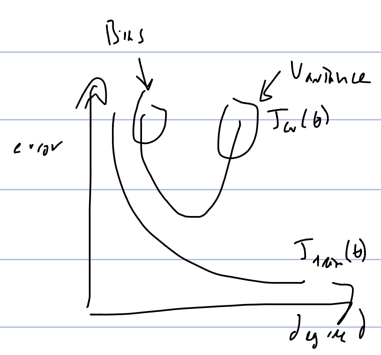
How well generalized? → fit to test set

How fix?

Split dataset into 3 pieces: training, cross validation (CV), test (60/20/20)

train/validation/test error

Diagnosing Bias vs. Variance



Bias (underfit):

$J_{train}(\theta)$ high

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

Variance (overfit).

$J_{train}(\theta)$ low

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

Regularization and Bias/Variance

Choosing λ

1. try $\lambda = 0 \rightarrow \min_\theta J(\theta) \rightarrow \theta^{(1)} \rightarrow J_{cv}(\theta^{(1)})$

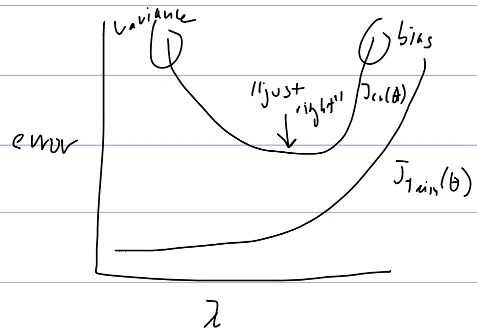
2. try $\lambda = 0.01 \rightarrow \min_\theta (J\theta) \rightarrow \theta^{(2)} \rightarrow J_{cv}(\theta^{(2)})$

$\lambda = 0.02 \longrightarrow \theta^{(3)} \rightarrow J_{cv}(\theta^{(3)})$

$\lambda = 0.04$

$\downarrow$

12 $\lambda = 10.24 \longrightarrow \theta^{(12)} \rightarrow J_{cv}(\theta^{(12)})$



Learning Curves

Deciding What To Do Next

- get more training examples → fixes high variance
- try smaller set of features (carefully select) → fixes high variance
- try getting additional features → fixes high bias

- try adding polynomial features → fixes high bias
- try decreasing λ → fixes high bias
- try increasing λ → fixes high variance

Spam Classifiers

Error Analysis

Precision: $\frac{\text{True pos}}{\text{true pos + false pos}} = \frac{\text{true positives}}{\text{\# predicted positive}}$

$y=1$ in 'rare' class

Recall: $\frac{\text{true pos}}{\text{true pos + false neg}} = \frac{\text{true positives}}{\text{\# actual positives}}$

F_1 Score: $2\frac{PR}{P+R}$

$P = 0.087R$ $.14824$
$R = 0.85$