Deciding What to Try Next

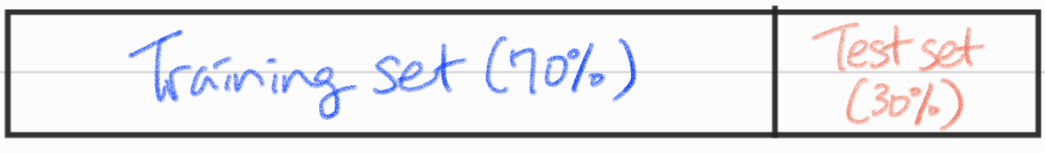
If hypothesis makes unacceptably large errors

- Get more training examples
- Try smaller sets of features
- Try getting additional features
- Try adding polynomial features
- Try decreasing/increasing λ

Evaluating a hypothesis

overfitting: fails to generalize to new examples not in training set.

Dataset



- Learn parameter θ from training set
- Compute set error

low error in training set,
but high error in test set

Model selection and training / validation / test sets

d = degree of polynomial

d = 1. $h_\theta(x) = \theta_0 + \theta_1 x$ $\theta^{(1)} \rightarrow J_{test}(\theta^{(1)})$

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

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

⋮

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

Choose the lowest?

해당 testset에 맞추는 거라 바람직하지 않음

Validation set ⇐

Training set (70%)	Test set (30%)

Training set (60%)	Validation set (20%)	Test set (20%)

Cross Validation 교차 검증 CV