Model Selection - how to choose model order

data proportions		
60% training set	J_{TRAIN}	training error
20% Cross-validation set	J_{CV}	cross-validation error
20% Test Set	J_{Test}	test error / generalization error

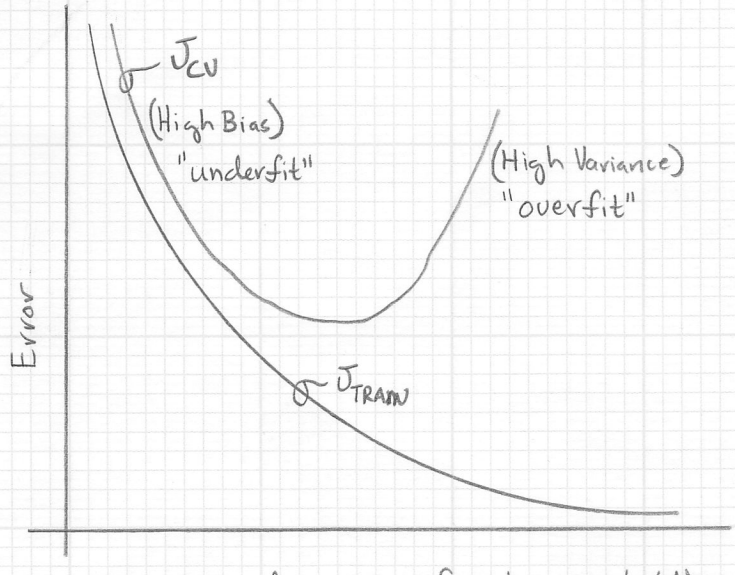
Procedure:

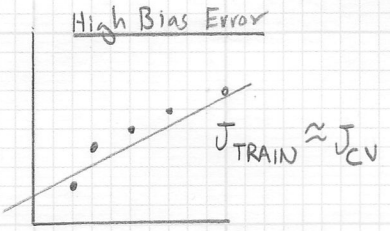
1. Fit several candidate models by minimizing the training error
2. Select candidate that gives the lowest cross-validation error.
3. Estimate generalization error using test set

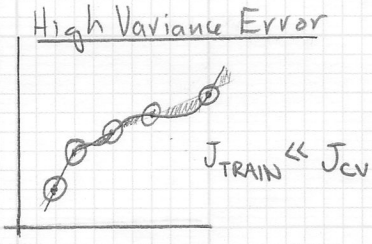
Notes:

* We should expect $J_{CV} < J_{TEST}$ since we have effectively fit a model order parameter to the CV set.

Bias vs. Variance







Bias & Variance - Regularization

* Do not include $\frac{\lambda}{2m}\sum_{j=1}^{m}\theta_j^2$ in cost functions used for selection purposes. *

$$J_{FIT}(\theta) = \frac{1}{2}\left\langle \left| h_\theta(x_{TRAIN}) - y_{TRAIN} \right|^2 \right\rangle + \frac{1}{2}\lambda\left(\langle\theta^2\rangle - \theta_0^2\right) \quad \text{use to fit params}$$

$$J_{TRAIN} = \frac{1}{2}\left\langle \left| h_\theta(x_{TRAIN}) - y_{TRAIN} \right|^2 \right\rangle$$

$$J_{CV} = \frac{1}{2}\left\langle \left| h_\theta(x_{CV}) - y_{CV} \right|^2 \right\rangle$$

$$J_{TEST} = \frac{1}{2}\left\langle \left| h_\theta(x_{TEST}) - y_{TEST} \right|^2 \right\rangle$$

} use for model selection

Methods for choosing λ:

- Brute force:

  $\lambda_0 = 0$
  $\lambda_1 = 1$
  $\lambda_2 = 2$
  $\vdots$
  $\lambda_n =$

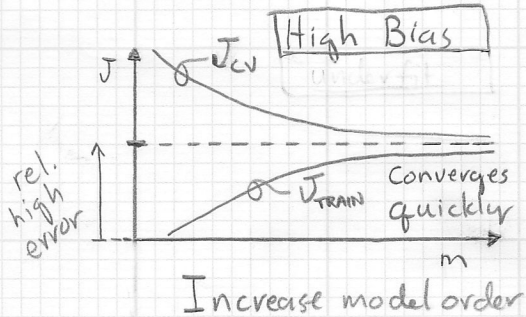
  } run each, choose best J_{CV}, report J_{TEST}

- Bisection search, terminating when bracket distance falls below some threshold.
- Newton-Raphson?

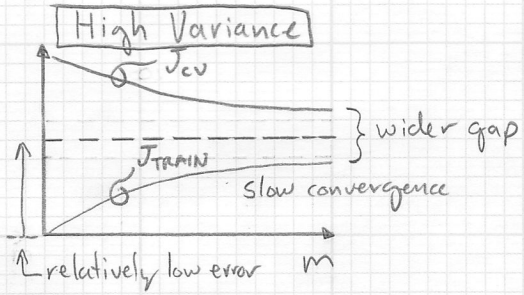
Learning Curves

J_{TRAIN} vs. m

J_{CV} vs. m



Increase model order



Increase training data set size.

Debugging:

High Bias:
- more features
- more polynomial features
- decrease λ

High Variance:
- less features
- more training examples
- increase λ