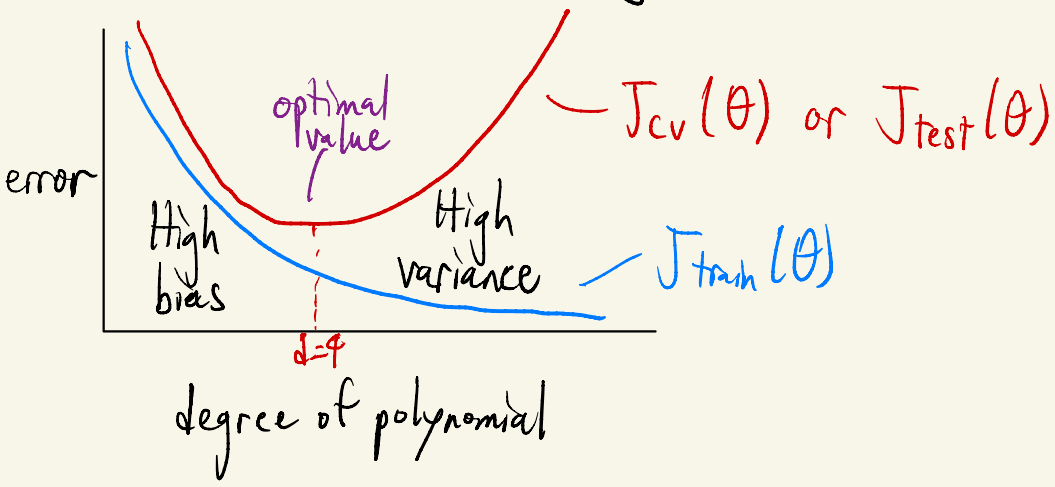
May 20, 2021

Bias/Variance:

High bias = underfitting

High variance = overfitting



High Bias (Underfit):

→ high $J_{train}(\theta)$

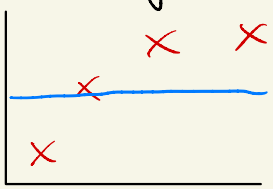
→ $J_{cv}(\theta) \approx J_{train}(\theta)$

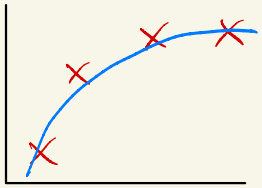
High Variance (Overfit):

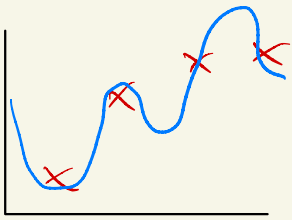
→ low $J_{train}(\theta)$

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

With Regularization:







↳ Large λ

High bias (underfit)

↳ Small λ

High variance (overfit)

Choosing Regularization Parameter λ:

e.g. 4_1 is chosen.

$\therefore h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

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

1. Try $\lambda = 0 \longrightarrow J_{cv}(\theta^{(1)})$
2. Try $\lambda = 0.01 \longrightarrow J_{cv}(\theta^{(2)})$
3. Try $\lambda = 0.04 \longrightarrow J_{cv}(\theta^{(3)})$

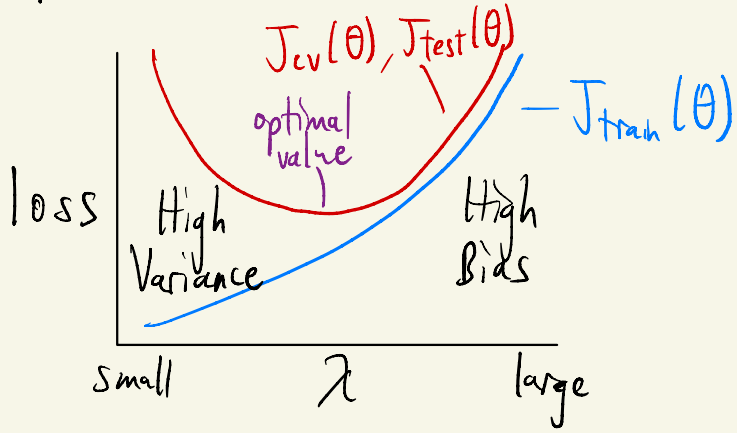
⋮

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

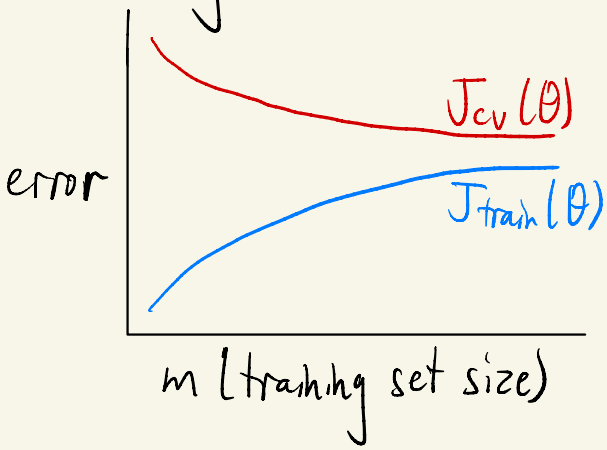
e.g. Pick 5., $\theta^{(5)}$.

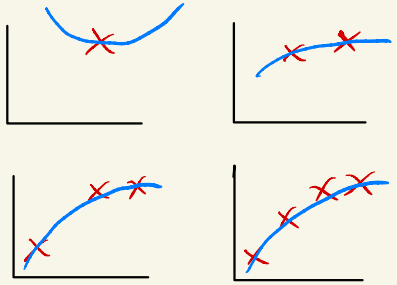
Now estimate generalization error $J_{test}(\theta^{(5)})$.

~~⊗~~ Same steps, more or less.



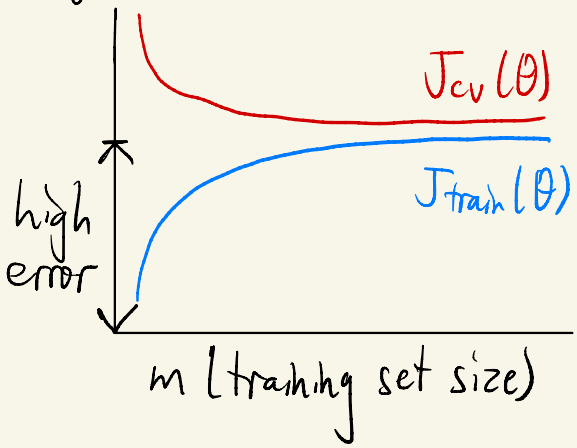
Learning Curves:



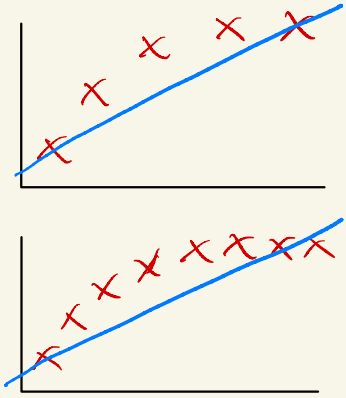


↳ harder to fit to all points if there are more

High Bias:

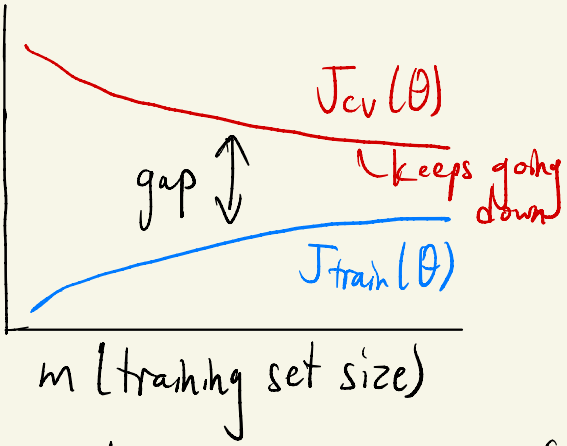


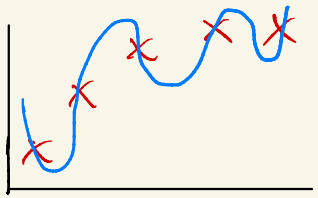
e.g. $h_\theta(x) = \theta_0 + \theta_1 x$

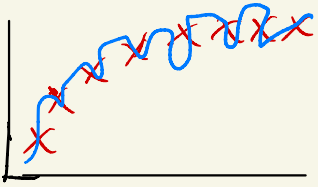


High Variance!

e.g. $h_\theta(x) = \theta_0 + \dots \theta_{100} x^{100}$
(small λ)







- If a learning algo is suffering from high variance, or $J_{cv}(\theta) \gg J_{train}(\theta)$, getting more training data is likely to help.
- Using a smaller set of features could also help.
- Adding more hidden units will NOT help!

Back to Beginning:

Scenario: Your regularized linear regression algorithm is making large, unexpected errors. Now what?

→ Get more training examples → fixes high variance
→ Try smaller sets of features → fixes high variance
→ Try getting additional features → fixes high bias
→ Adding polynomial features (i.e. x_1^2, x_2^2, x_1x_2, etc) ↳ fixes high bias
→ Increase/decrease λ
↳ fixes high variance / fixes high bias

Size of Network:

→ Small NN's have less parameters, more prone to underfitting.
→ Large NN's are more prone to overfitting, but can be addressed w/ regularization.