

CS444

Nathan Sprague

February 27, 2017

Supervised Learning

- Stationarity Assumption...
- Data is i.d.d.
 - Independent and identically distributed.
- Goal is to minimize loss on unobserved data from the same distribution.

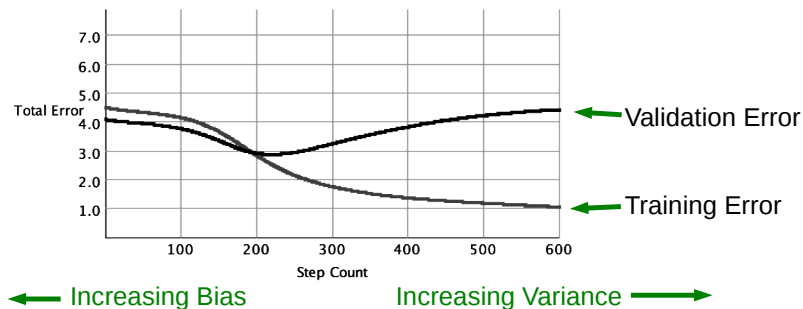
Why Do Errors Happen?

- (Informally) Three reasons:
 - 1 Bias - Strong assumptions / excessive generalization
 - 2 Variance - Too much sensitivity to insignificant detail in the training data
 - 3 Unavoidable/Irreducible error
- Bias and Variance are two sides of a coin: decreasing one tends to increase the other
 - “Bias/Variance Dilemma”

Evaluating Learning

- Training Set / Test Set
- Validation Set

Using A Validation Set



Artificial Intelligence: Foundations of Computational Agents, Poole and Mackworth, 2010

Not Enough Data?

- K-fold cross-validation
- Leave-one-out cross-validation

Model Selection vs. Optimization

- Model selection - Defines the hypothesis space.
- Optimization - Finding the best hypothesis in that space.

Regularization

- Sort of like addressing the model selection problem through optimization.
- Change your loss metric to penalize complexity...

$$E(\mathbf{w}) = \sum_j (y_j - \mathbf{w}^T \mathbf{x}_j)^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

Solution:

$$\mathbf{w}^* = (X^T X + \lambda I)^{-1} X^T \mathbf{y}$$