Neural Networks and Deep Learning

Bias/Variance Tradeoff and Regularization I

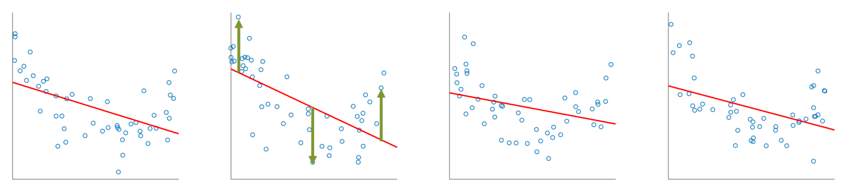
Adam Bloniarz

Department of Computer Science
adam.bloniarz@colorado.edu

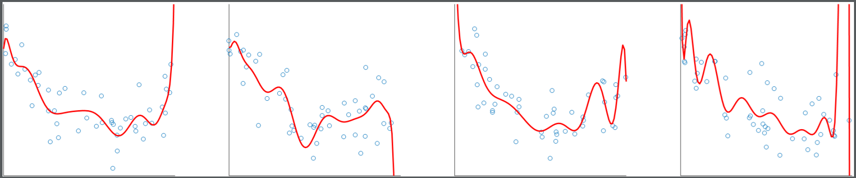
March 16, 2020

University of Colorado Boulder

Bias: The estimate is consistently wrong in the same way. Its expected value is different from the true value.



Variance: The estimate is all over the place on different training sets. Its expected value is correct, though.

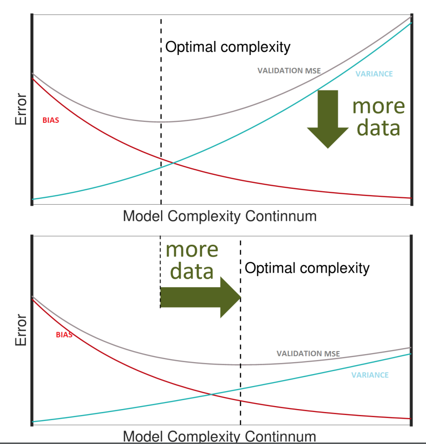


Bias-Variance Decomposition

Consider the prediction at a single point $\mathbf{x}$. Calculate expected mean-squared error, where expectation is over training data and a new test pair $(\mathbf{x}, y)$, where $y = f(\mathbf{x}) + \epsilon$.

$$\begin{aligned}
\mathbb{E}\left(y - \hat{f}(\mathbf{x})\right)^2 &= \mathbb{E}\left(f(\mathbf{x}) + \epsilon - \hat{f}(\mathbf{x})\right)^2 \\
&= \mathbb{E}\left(f(\mathbf{x}) - \hat{f}(\mathbf{x})\right)^2 + \mathbb{E}(\epsilon)^2 \\
&= \mathbb{E}\left(f(\mathbf{x}) - \mathbb{E}\hat{f}(\mathbf{x}) + \mathbb{E}\hat{f}(\mathbf{x}) - \hat{f}(\mathbf{x})\right)^2 + \mathbb{E}(\epsilon)^2 \\
&= \underbrace{\mathbb{E}\left(f(\mathbf{x}) - \mathbb{E}(\hat{f}(\mathbf{x}))\right)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left(\mathbb{E}(\hat{f}(\mathbf{x})) - \hat{f}(\mathbf{x})\right)^2}_{\text{Variance}} + \mathbb{E}(\epsilon)^2
\end{aligned}$$

- More complex model: usually decreases bias, increases variance.
- More data: usually decreases variance thanks to errors cancelling, law of large numbers.


Optimal complexity
VALIDATION MSE
VARIANCE
BIAS
more data
Error
Model Complexity Continnum
more data
Optimal complexity
BIAS
VALIDATION MSE
VARIANCE
Error
Model Complexity Continnum

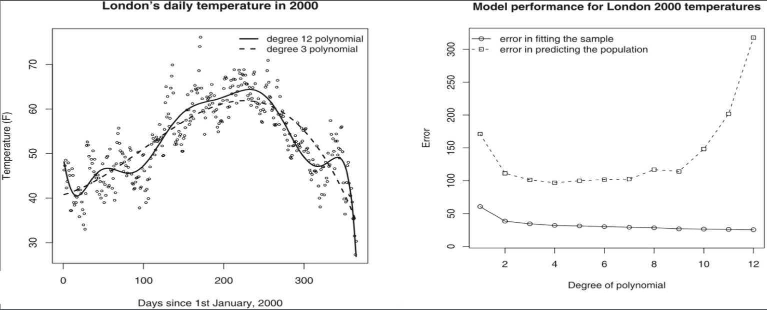
Model selection

Measures of model complexity

- Degree of polynomial expansion
- Number of neighbors in k-NN
- Width of kernel in local polynomial estimation
- Number of neurons in MLP
- Depth of decision tree
- RKHS norm of function (Kernel methods)

Model selection

Increased model complexity usually the training error. There are more degrees of freedom to interpolate training data.



Source: Gigerenzer & Brighton (2009)

Model selection

Choosing model complexity always requires something besides training error:

- Test-set error
- Cross-validated error
- AIC, BIC, MDL: Quantify how complex the model is, find optimal tradeoff with training error

Definition in Goodfellow, Bengio, Courville:

> *Regularization is any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error.*

In a statistical sense, regularization aims to reduce **variance**, at the cost (sometimes) of a relatively small increase in bias.