error = bias^2 + variance + noise

Noise:

other cr*p

Bias:

how far away you fall from actual

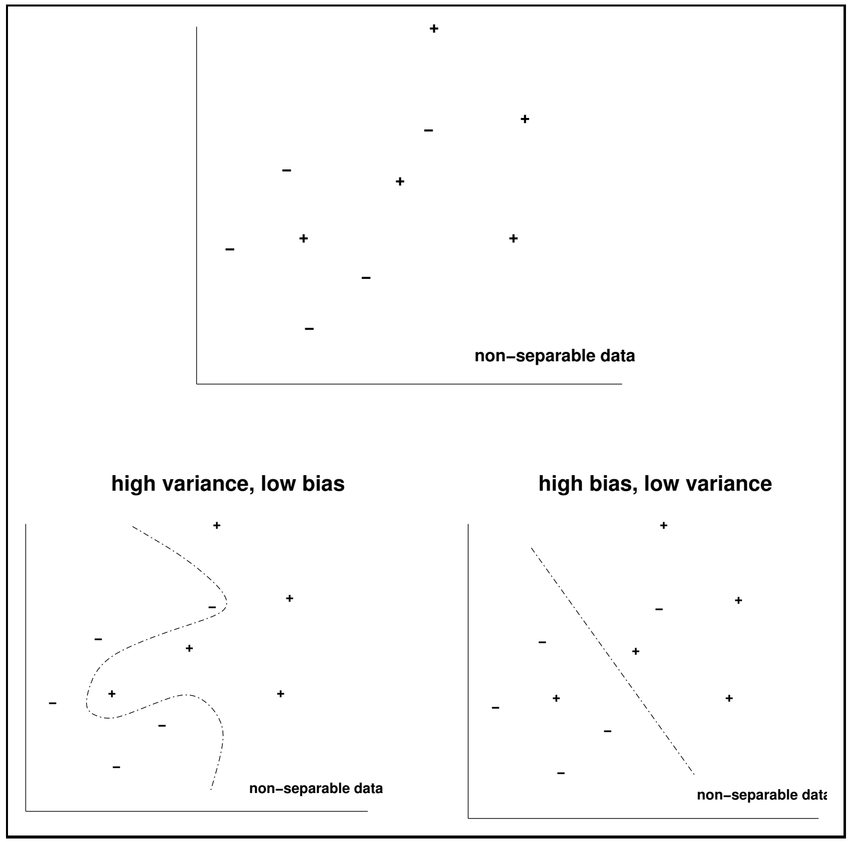
Variance:

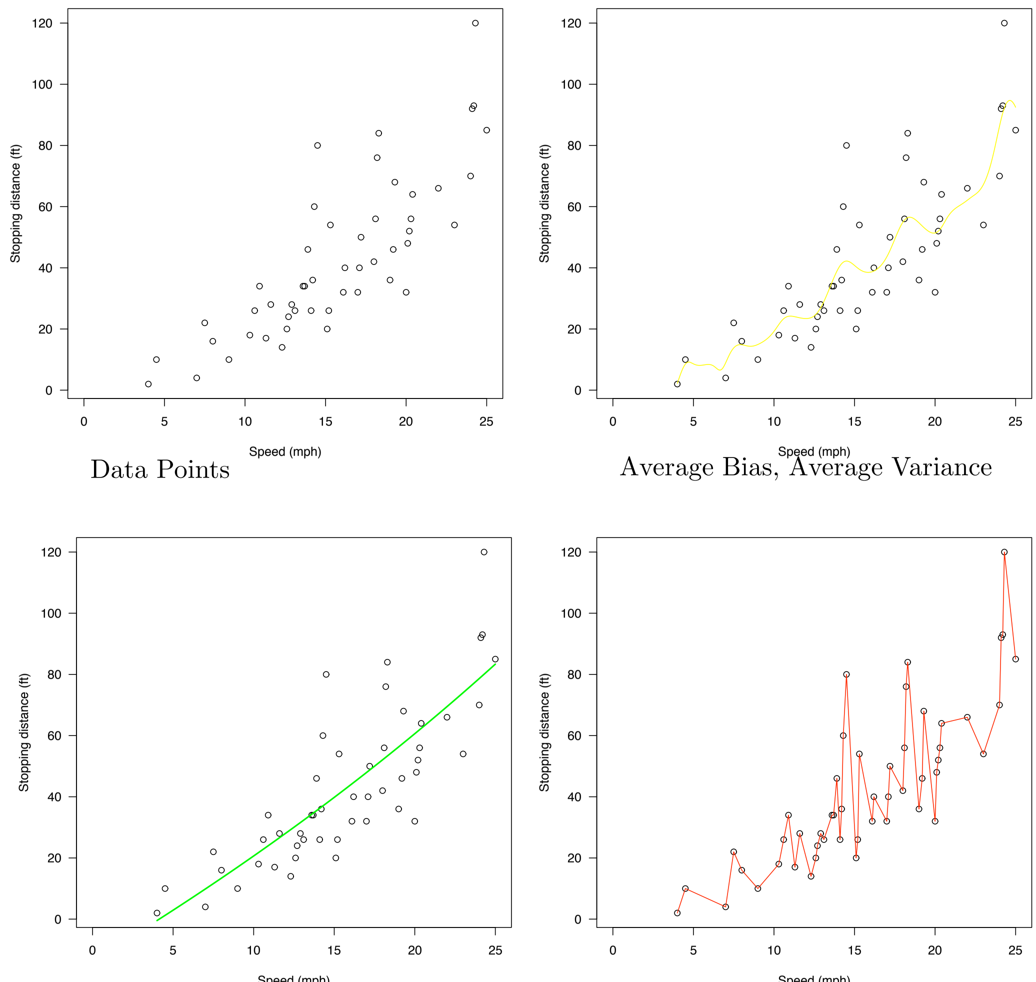
how much you dance around

btw, standard deviation = sqrt(variance)

Bias Variance in Classification

- There are several ways to define the concept of bias and variance for classification problems, but the key tradeoff between them remains.





Data Points

Average Bias, Average Variance

Very High Bias, Very Low Variance

Very Low Bias, Very High Variance

Figure 1: Examples of function estimations with different levels of bias and variance

Dependence of Bias-Variance on Model Complexity

- *$h(x)=sin(2\pi x)$*
- Regularization parameter λ
- $L=100$ data sets
- Each with $N=25$
- *24* Gaussian Basis functions
  - No of parameters $M=25$
- Total Error function:

$$\boxed{\frac{1}{2}\sum_{n=1}^{N}\{\ t_n - \mathrm{w}^T\phi(x_n)\ \}^2 + \frac{\lambda}{2}\mathrm{w}^T\mathrm{w}}$$

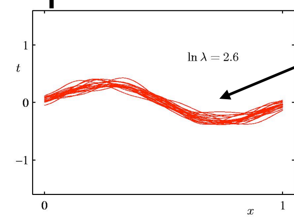
where ϕ is a vector of basis functions

Result of averaging multiple solutions with complex model gives good fit

Weighted averaging of multiple solutions is at heart of Bayesian approach: not wrt multiple data sets but wrt posterior distribution of parameters

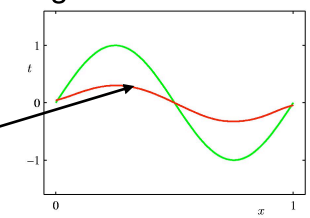
20 Fits for 25 data points each

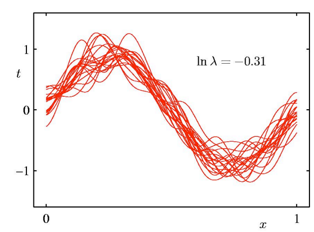
Red: Average of Fits
Green: Sinusoid from which data was generated

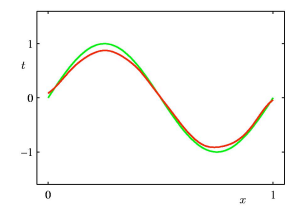


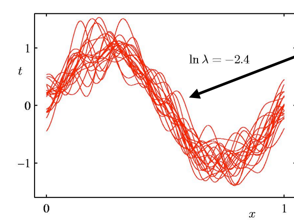
High λ

Low Variance
High bias



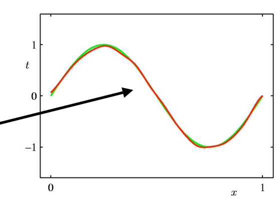






Low λ

High Variance
Low bias



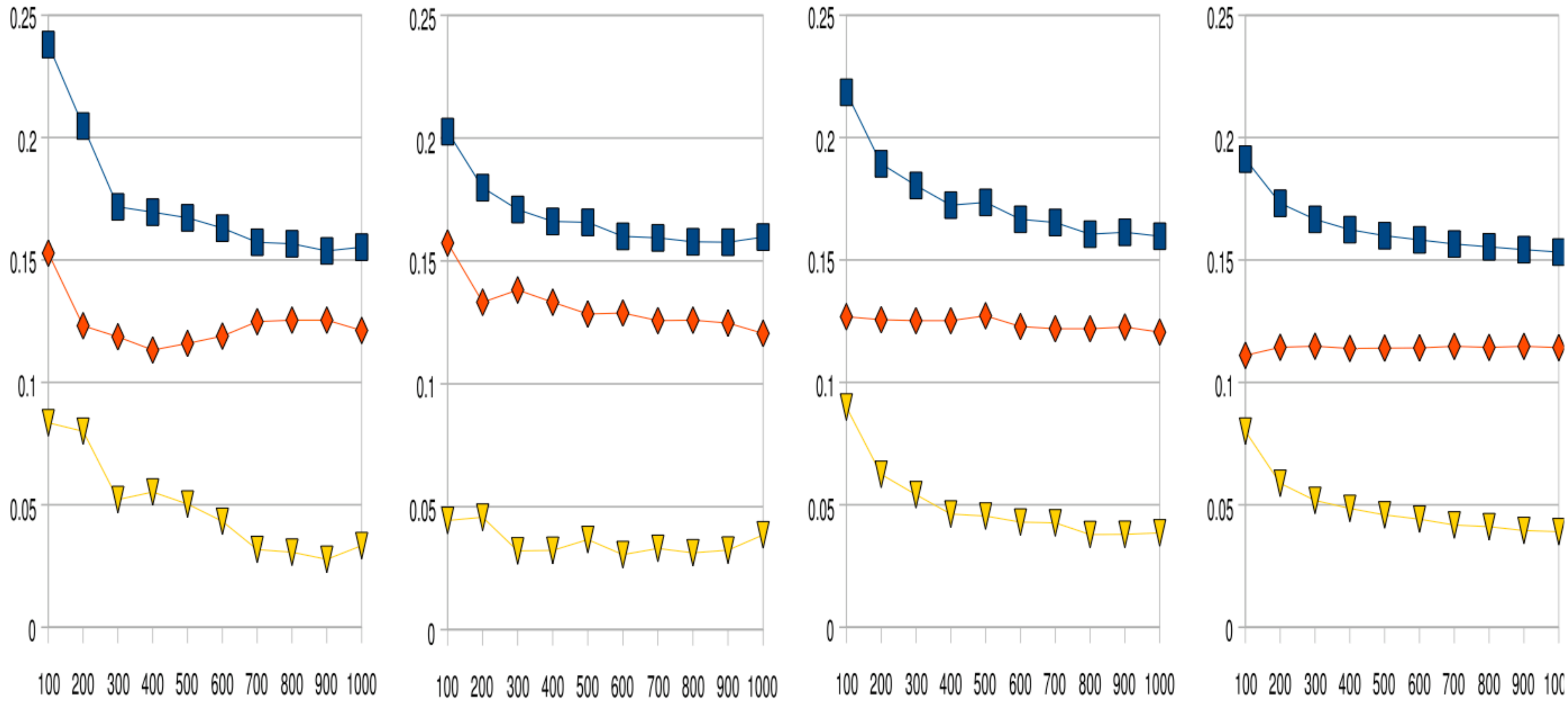
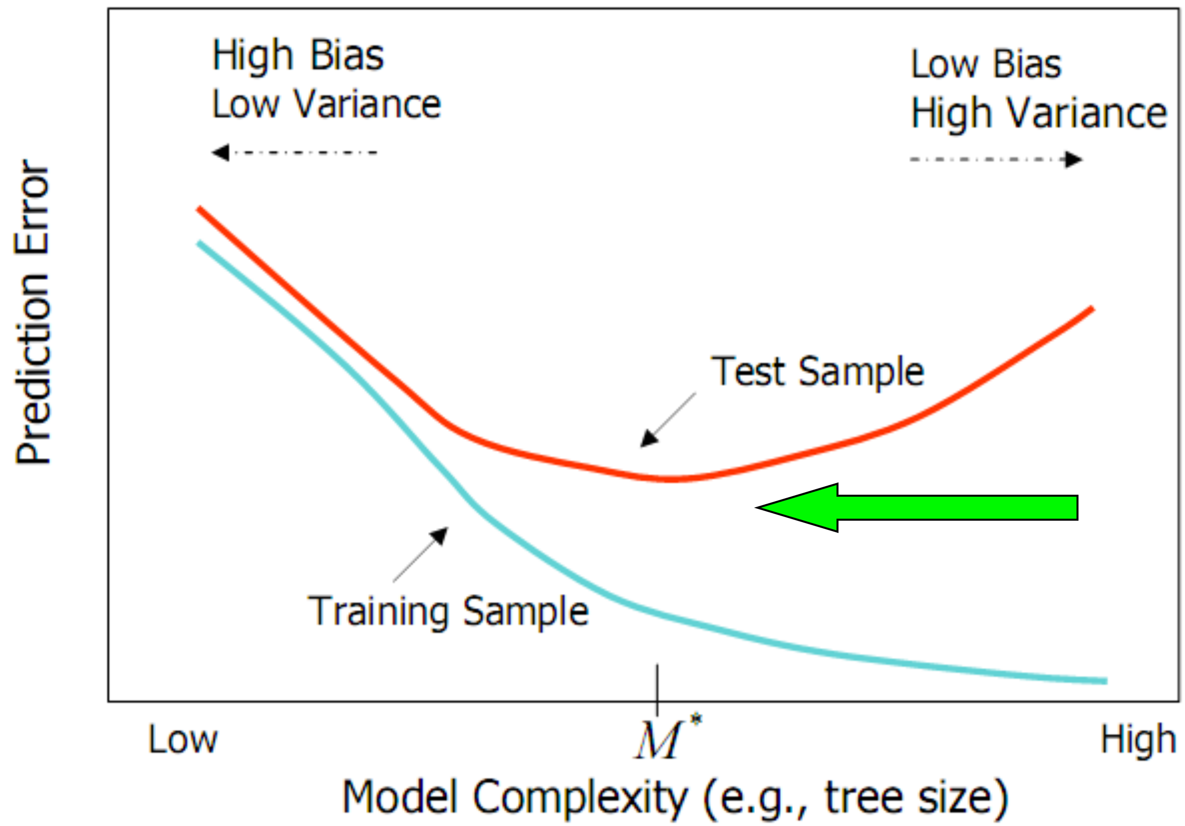


Fig. 1. Various bias variance decompositions on the same problem (see text for description). Training set size on x-axis, error (top line) bias (middle line) and variance (bottom line) on y-axis.

Bias and Variance

- Ensemble methods
  - Combine learners to reduce variance



from Elder, John. From Trees to Forests and Rule Sets - A Unified Overview of Ensemble Methods. 2007.