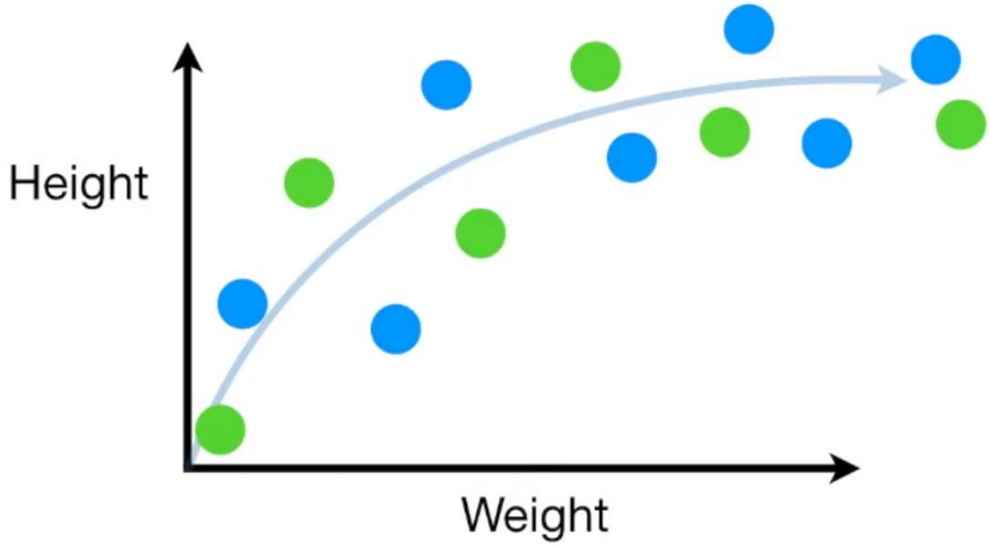
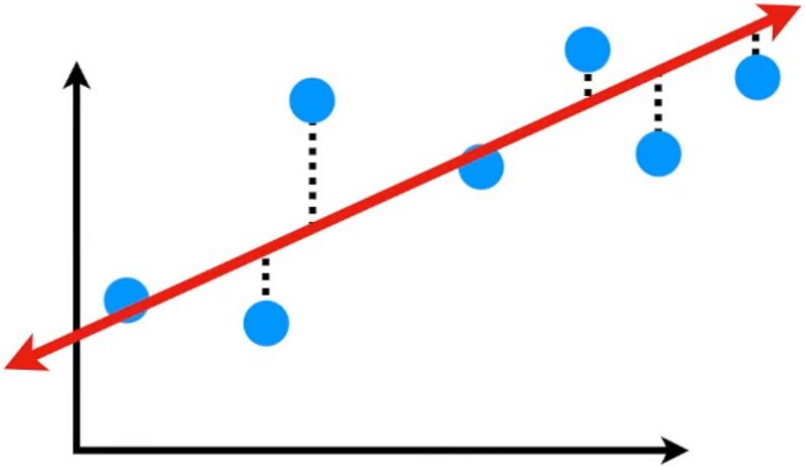
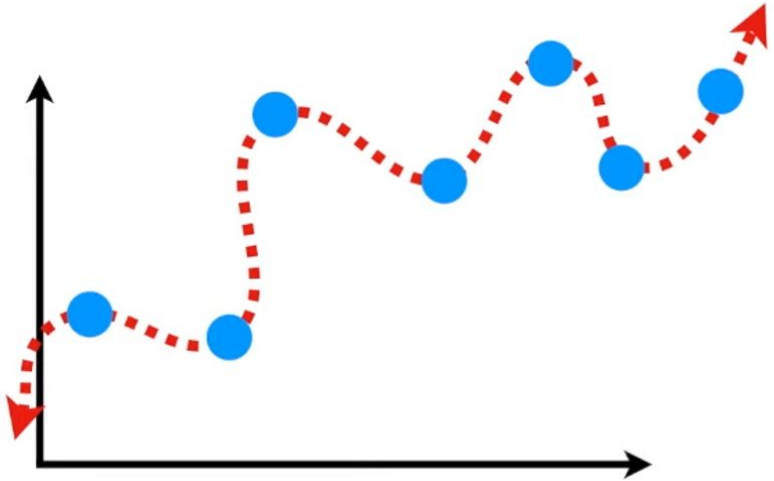
Bias - Variance Trade-off

Vishnuram Ayyavu Vijayakumar

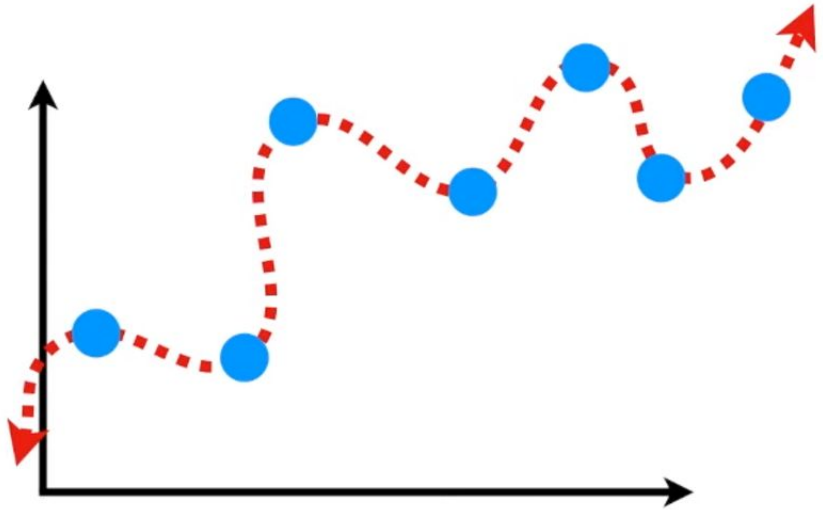

Height
Weight

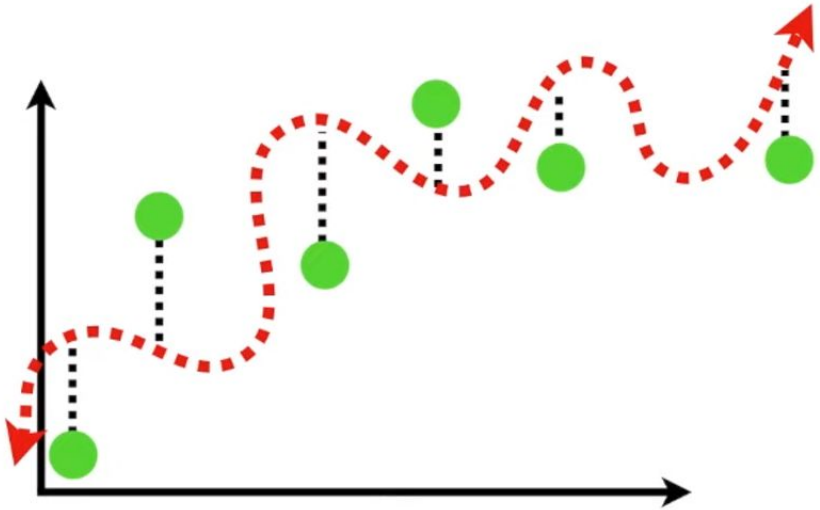
Bias





Variance





Graphical illustration of Bias-Variance

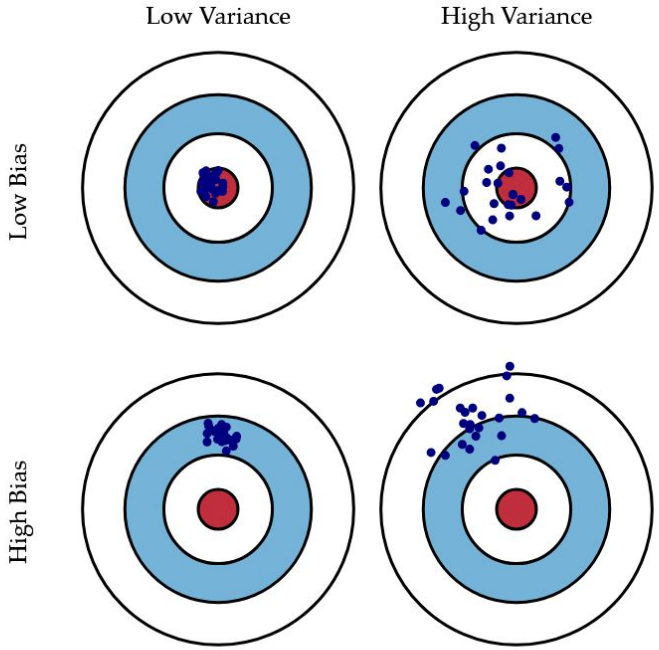


Fig. 1 Graphical illustration of bias and variance.

Example: Voting intentions

Voting Republican	Voting Democratic	Non-Respondent	Total
13	16	21	50

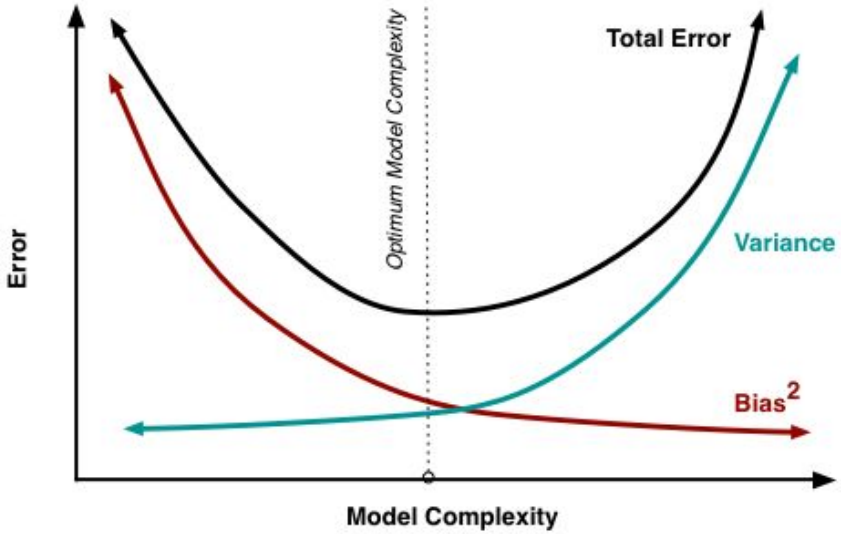
Source of bias:

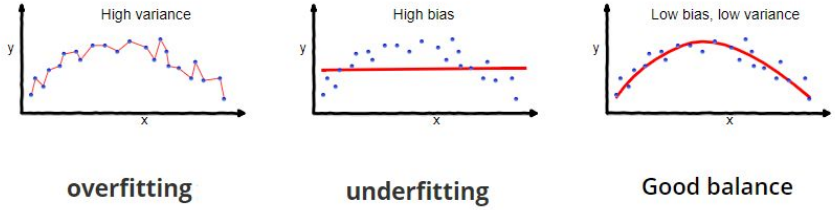
- Selecting from phone directories
- Not checking on the non-respondents

Source of variance:

- Size of the sample is small (50)

Model Complexity vs Bias and Variance





Sweet spot:

$$\frac{dBias}{dComplexity} = -\frac{dVariance}{dComplexity}$$

Ensemble Learning

Vishnuram Ayyavu Vijayakumar

Ensemble Learning

where:

- F1, . . . , Fm are Individual models
- β_i is a weight for each model's prediction

$$f(x) = \sum_{i=1}^{m} \beta_i F_i(x),$$

Bootstrap **agg**regat**ing** : Bagging

Bootstrapping is a general resampling technique where we aim to simulate drawing a new sample from the true underlying distribution from which our training set is generated (since we don't have access to the true distribution
outright).

Bootstrap sampling example:

- **Original Dataset**: {10, 20, 30, 40, 50}
- **Original Statistic**: Mean = 30
- **Bootstrap Samples**:
  - Sample 1: {10, 20, 20, 50, 40} → Mean = 28
  - Sample 2: {30, 30, 50, 10, 40} → Mean = 32
  - Sample 3: {50, 20, 10, 10, 30} → Mean = 24
- **Distribution of Means**: {28, 32, 24, ...}

Original Dataset

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
Yes	No	Yes	167	Yes

Random forest