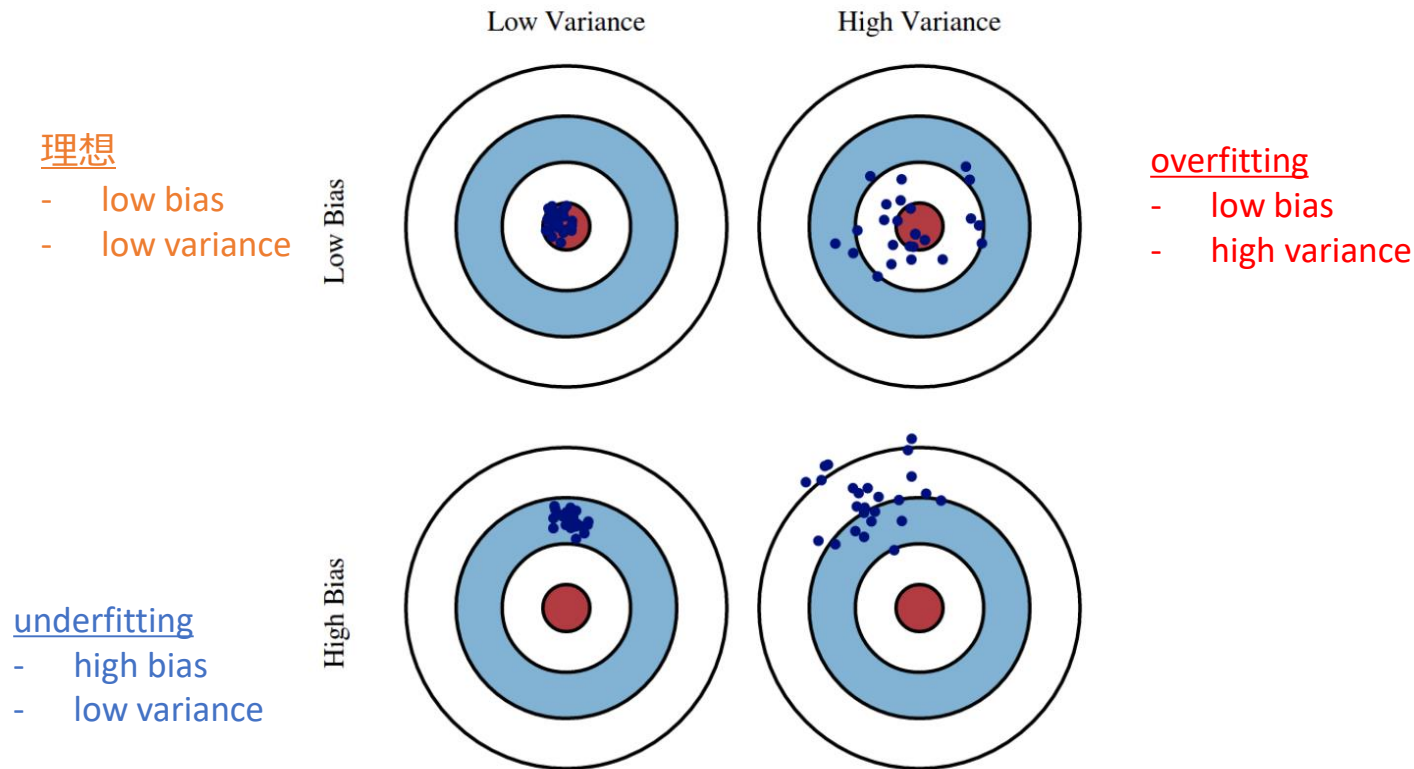


# バイアスとバリエーション

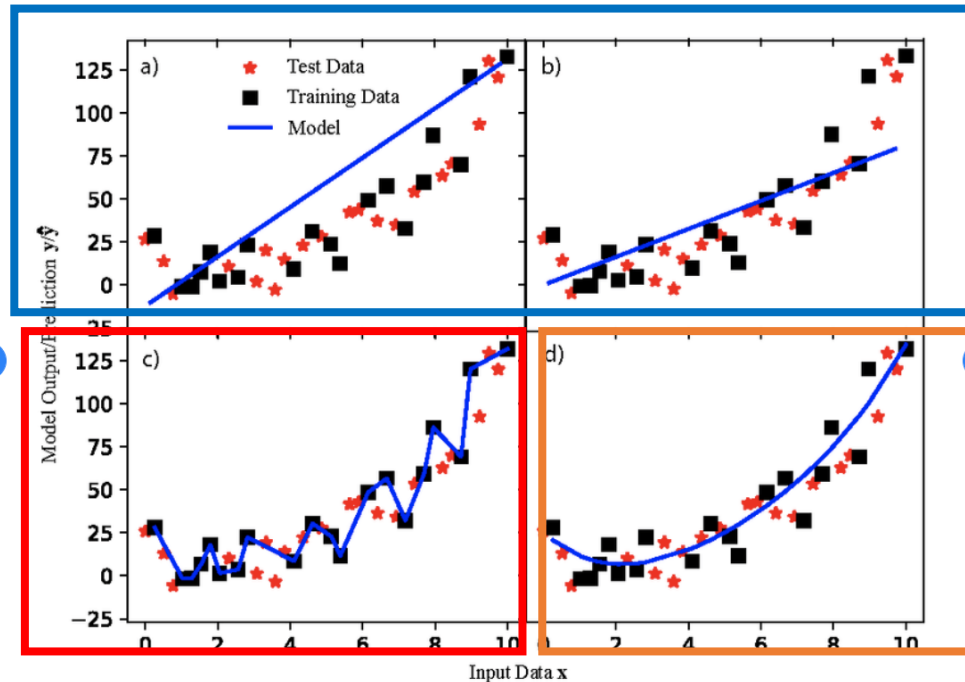
- 汎化誤差 = バイアス + バリエーション + ノイズ (irreducible error)
- バイアス : 偏り。系統的な誤差
- バリエーション : ばらつき。データの小さな変動に対するモデルの応答性



underfittingだとバイアスは高いがバリエーションは低い、overfittingだとバイアスは低いがバリエーションは高い

underfitting

- high bias
- low variance



overfitting

- low bias
- high variance

理想

- low bias
- low variance

Illustrations of high bias and high variance models. A toy dataset was generated from the polynomial  $y = 5 + 0.1x + 0.1x^2 + 0.1x^3 + 0.002x^4 + \text{Random Noise}$ . The fits in a) and b) are both parameterizations of a model. Each model (line) in both fits has approximately the same error but does not accurately capture the behavior of the data due to poor model assumptions; in this case, fitting a first order polynomial to a dataset generated from a fourth order polynomial. This is an example of high bias models. A twentieth order polynomial was fit to a subset of the full dataset in c), shown in blue. While the model has very good predictive error for the training dataset it will not extrapolate well to the data in the testing set; this is overfitting or high variance. A third order polynomial was fit to the data in d) demonstrating a good balance between bias and variance. The model accurately captures trends in the data while not overfitting the training dataset.

一般的にモデルが複雑になれば（自由度が上がれば）、バイアスは低下しバリエーションが増加する

underfitting

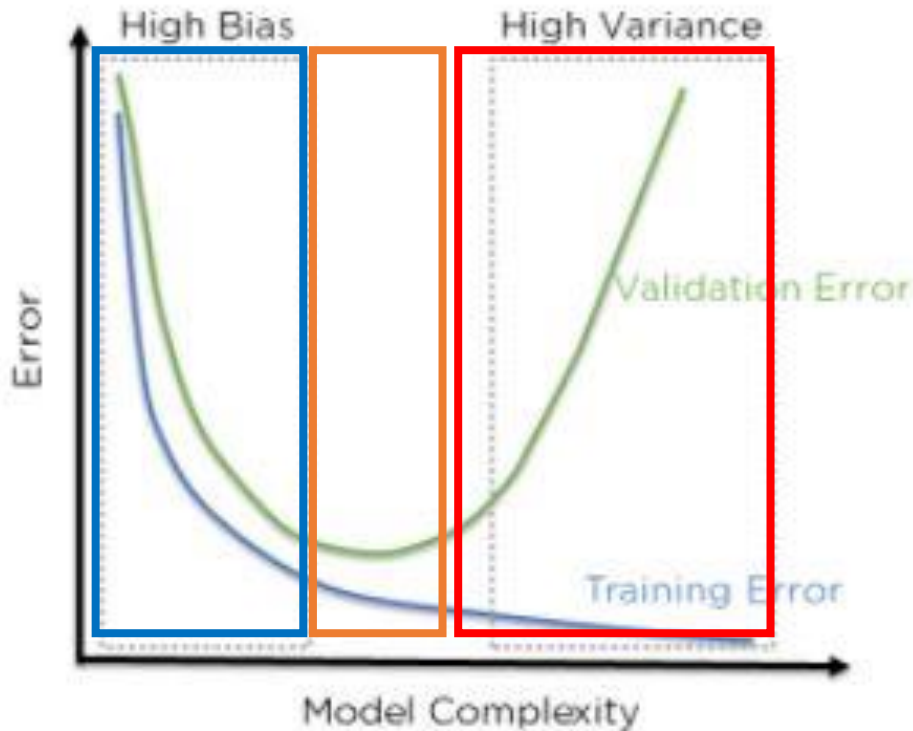
- high bias
- low variance

理想

- low bias
- low variance

overfitting

- low bias
- high variance

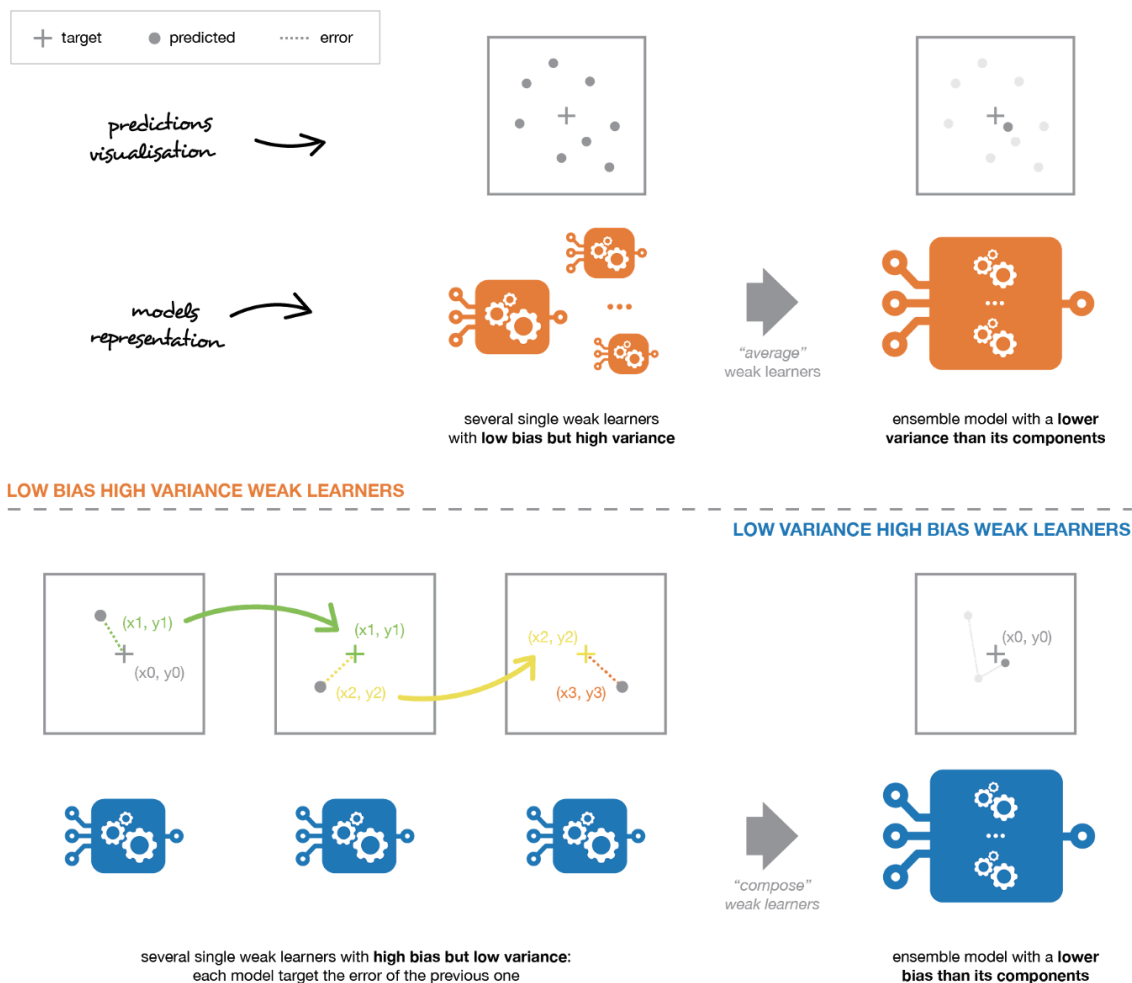


モデルの複雑度（決定木の深さetc）

- モデルを複雑にすると
  - バイアスは下がる
  - バリエーションは増える
- モデルをシンプルにすると
  - バリエーションは減る
  - バイアスは増える
- トレードオフの関係にある  
↓
- 複数のモデルを組み合わせで対応する（アンサンブル）
- 例えば、自由度の低いモデル（バイアス大・バリエーション小）を複数組み合わせることで全体としてバイアス小・バリエーション小を実現する

- ベースモデルとアンサンブルとの関係
  - ベースモデルがバイアスが低くバリエーションが大きい→分散を減らすようなアンサンブルを行う
  - ベースモデルがバリエーションが低くバイアスが高い→バイアスを減らすようなアンサンブルを行う
- バギング
  - バリエーションを低くする
- ブースティング、スタッキング
  - バイアスを低くする
- (※) スタッキングはバリエーションとバイアスの両方を下げるとも言われる

# バギングとブースティングによるバリエーションとバイアスの低減



<https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205>