Foundations of Machine Learning
AI2000 and AI5000

FoML-10
Bias Variance Decomposition

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

So far in FoML

- What is ML and the learning paradigms
- Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Linear Regression with basis functions - and regularization
- Model selection

Breaking down the prediction error of a model

Frequentist interpretation of the model complexity

Expected Loss for Regression

- Regression loss $L(t, y(\mathbf{x})) =$

Expected Loss for Regression

- Regression loss $L(t, y(\mathbf{x})) =$

- If we know the data distribution, we can find the

$$\mathbb{E}[L(t, y((\mathbf{x})))] =$$

Data and prediction distributions

Minimizing the Expected loss at given x

Expected Loss for Regression

$$\mathbb{E}[L] = \int \int (y(\mathbf{x}) - t)^2 p(\mathbf{x}, t) dt d\mathbf{x}$$

Minimizing the expected loss

$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}]\}^2 p(\mathbf{x}) d\mathbf{x} + \int \mathrm{var}[t/\mathbf{x}] p(\mathbf{x}) d\mathbf{x}$$

- Optimal solution is unknown $y(\mathbf{x}) = \mathbb{E}[t/\mathbf{x}]$

Minimizing the expected loss

$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}]\}^2 p(\mathbf{x}) d\mathbf{x} + \int \mathrm{var}[t/\mathbf{x}] p(\mathbf{x}) d\mathbf{x}$$

- Optimal solution is unknown $y(\mathbf{x}) = \mathbb{E}[t/\mathbf{x}]$

- We only have finite dataset (but not the distribution)

Minimizing the expected loss

$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}]\}^2 p(\mathbf{x}) d\mathbf{x} + \int \mathrm{var}[t/\mathbf{x}] p(\mathbf{x}) d\mathbf{x}$$

- Frequentist approach → multiple datasets, multiple models

$$\mathbb{E}_D[(y_D(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}])^2]$$

Minimizing the expected loss

$$\mathbb{E}[\mathbb{E}_D[L]] = \int \mathbb{E}_D[(y_D(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}])^2]p(\mathbf{x})d\mathbf{x} + \int \mathrm{var}[t/\mathbf{x}]p(\mathbf{x})d\mathbf{x}$$

- Bias-Variance decomposition

Minimizing the expected loss

$$\mathbb{E}[\mathbb{E}_D[L]] = \int \mathbb{E}_D[(y_D(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}])^2]p(\mathbf{x})d\mathbf{x} + \int \mathrm{var}[t/\mathbf{x}]p(\mathbf{x})d\mathbf{x}$$

- Bias-Variance decomposition

$$\mathbb{E}_D[(y_D(\mathbf{x}) - \mathbb{E}[t/\mathbf{x}])^2] = \mathbb{E}_D[(y_D(\mathbf{x}) - \mathbb{E}_D[y_D(\mathbf{x})] + \mathbb{E}_D[y_D(\mathbf{x})] - \mathbb{E}[t/\mathbf{x}])^2]$$

$(\text{Bias})^2$ =

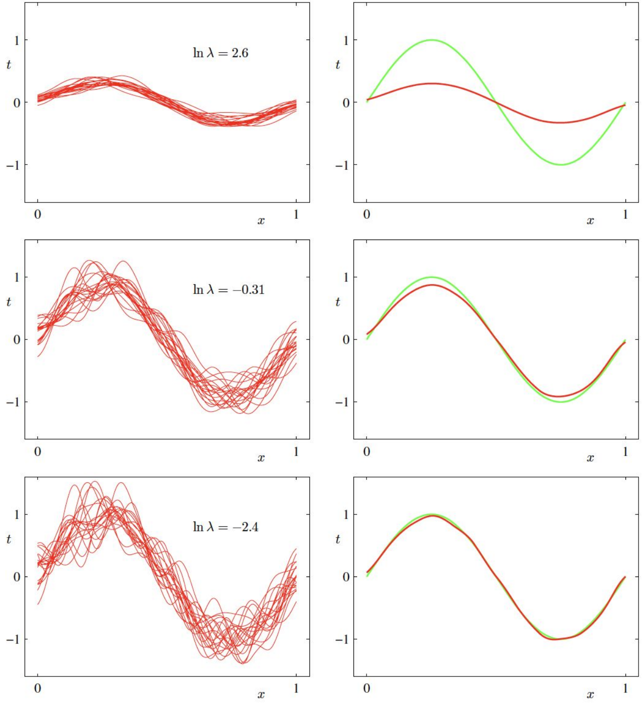
Variance =

Noise =

Example

Bias-Variance Decomposition Example

- 100 datasets of size 25
- x ~ U[0, 1]
- t = sin(2πx) + ϵ



$$\mathbb{E}_D[y_D(x)] = \bar{y}(x)$$

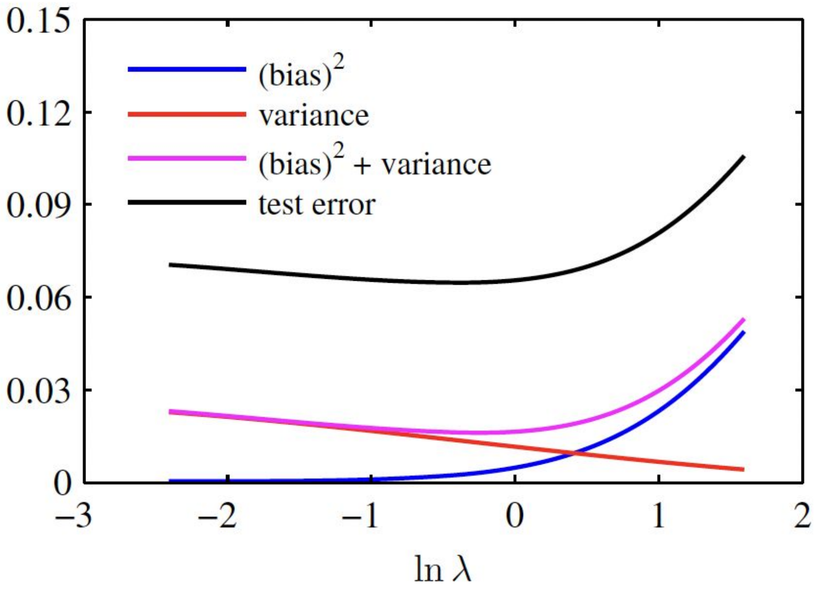
Bias-Variance Decomposition Example

Estimating the bias and variance

$$(\text{bias})^2 = \int \{\mathbb{E}_D[y_D(x) - \mathbb{E}[t/x]\}^2 p(x)dx$$

$$\text{variance} = \mathbb{E}_D[\{y_D(x) - \mathbb{E}_D[y_D(x)]\}]^2 p(x)dx$$

Bias-Variance Decomposition Example



Bias-Variance Decomposition

- In practice - we don't split our dataset to determine the model complexity
  - Large datasets are better
- Bayesian regression!

Rough work

Next
Bayesian Regression