Bias & Variance

Dr. Alireza Aghamohammadi

May 12, 2024

Problem Statement

- There are three potential sources of error: *noise*, *bias*, and *variance*.

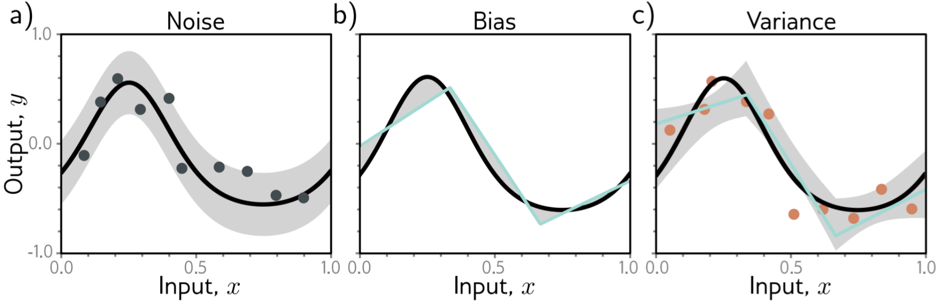

Figure: Possible sources of error.[1]

[1] Adopted from the book, Understanding Deep Learning

Mathematical Analysis (Part 1)

- For simplicity, let's consider a 1D linear regression problem where the data generation process includes additive noise with variance σ^2.
- For a given input x, there can be multiple valid outputs y.
- For each input x, there exists a distribution $\Pr(y|x)$ with an expected value $\mu[x]$ given by:

$$\mu[x] = \mathbb{E}_y[y[x]] = \int y[x] \ \Pr(y|x) \ \mathrm{d}y$$

$$\sigma^2 = \mathbb{E}_y[(\mu[x] - y[x])^2]$$

Mathematical Analysis (Part 2)

- Let's denote our prediction model by $\mathrm{f}[x, \phi]$ and compute the least squares loss.

$$\begin{aligned} L[x] &= (\mathrm{f}[x, \phi] - y[x])^2 \\ &= ((\mathrm{f}[x, \phi] - \mu[x]) + (\mu[x] - y[x]))^2 \\ &= (\mathrm{f}[x, \phi] - \mu[x])^2 + 2(\mathrm{f}[x, \phi] - \mu[x])(\mu[x] - y[x]) + (\mu[x] - y[x])^2 \end{aligned}$$

- The loss depends on the particular $y[x]$:

$$\begin{aligned} \mathbb{E}_y[L[x]] &= \mathbb{E}_y[(\mathrm{f}[x, \phi] - \mu[x])^2 + 2(\mathrm{f}[x, \phi] - \mu[x])(\mu[x] - y[x]) + (\mu[x] - y[x])^2] \\ &= (\mathrm{f}[x, \phi] - \mu[x])^2 + 2(\mathrm{f}[x, \phi] - \mu[x])(\mu[x] - \mathbb{E}_y[y[x]]) + \mathbb{E}_y[(\mu[x] - y[x])^2] \\ &= (\mathrm{f}[x, \phi] - \mu[x])^2 + \underbrace{\sigma^2}_{\text{Noise}} \end{aligned}$$

Mathematical Analysis (Part 3)

- The first term can be further decomposed into bias and variance.
- The parameter ϕ depends on the training dataset $\mathcal{D} = \{x_i, y_i\}_{i=1}^{m}$
- The expected value of the prediction model with respect to all possible training datasets $\mathcal{D}$ is given by:

$$\mathrm{f}_{\mu}[x] = \mathbb{E}_{\mathcal{D}}\Big[\mathrm{f}\big[x, \phi\,[\mathcal{D}]\big]\Big]$$

- Now let's add and subtract $\mathrm{f}_{\mu}[x]$ to $(\mathrm{f}[x, \phi] - \mu[x])^2$ and expand:

$$\begin{aligned}(\mathrm{f}[x, \phi[\mathcal{D}]] - \mu[x])^2 &= \big((\mathrm{f}[x, \phi[\mathcal{D}]] - \mathrm{f}_{\mu}[x]) + (\mathrm{f}_{\mu}[x] - \mu[x])\big)^2 \\ &= (\mathrm{f}[x, \phi[\mathcal{D}]] - \mathrm{f}_{\mu}[x])^2 + 2(\mathrm{f}[x, \phi[\mathcal{D}]] - \mathrm{f}_{\mu}[x])(\mathrm{f}_{\mu}[x] - \mu[x]) + (\mathrm{f}_{\mu}[x] - \mu[x])^2\end{aligned}$$

- We then take the expectation with respect to the training dataset $\mathcal{D}$:

$$\mathbb{E}_{\mathcal{D}}\Big[(\mathrm{f}[x, \phi[\mathcal{D}]] - \mu[x])^2\Big] = \mathbb{E}_{\mathcal{D}}\Big[(\mathrm{f}[x, \phi[\mathcal{D}]] - \mathrm{f}_{\mu}[x])^2\Big] + (\mathrm{f}_{\mu}[x] - \mu[x])^2$$

- Finally, we can obtain this equation:

$$\mathbb{E}_{\mathcal{D}}\Big[\mathbb{E}_{y}[L[x]]\Big] = \underbrace{\mathbb{E}_{\mathcal{D}}\Big[(\mathrm{f}[x, \phi[\mathcal{D}]] - \mathrm{f}_{\mu}[x])^2\Big]}_{\text{Variance}} + \underbrace{(\mathrm{f}_{\mu}[x] - \mu[x])^2}_{\text{Bias}} + \underbrace{\sigma^2}_{\text{Noise}}$$

Reducing Error

- Noise: Unavoidable.
- Variance: Reduced by more data or better representation.
- Bias: Lowered by complex models, but risk of increased variance (Bias-Variance Tradeoff).

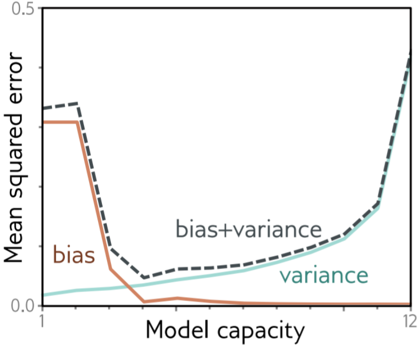

Figure: Bias-Variance Tradeoff.[2]

[2]Adopted from the book, Understanding Deep Learning