Machine Learning

Master Degree in Computer Science - IAS Curriculum
Bias-Variance Trade-off

Marco Piangerelli
marco.piangerelli@unicam.it

10 January 2023

Bias-Variance Trade-off

- Small E_{err}: good approximation of f out sample
- More complex $\mathcal{H} \Rightarrow$ better chance of approximating f
- Less complex $\mathcal{H} \Rightarrow$ better chance of generalizing of out sample
- Ideal $\mathcal{H} = \{f\}$ OUR DESIRE....

Bias Variance Trade-off

- First approach
  VC analysis $\rightarrow Err_{out}(g) \leq Err_{in}(g) + \Omega(N, \mathcal{H}, \delta)$

Bias Variance Trade-off

- First approach
  VC analysis $\rightarrow Err_{out}(g) \leq Err_{in}(g) + \Omega(N, \mathcal{H}, \delta)$
- Second approach
  Bias-variance analysis is another $\rightarrow$????

Bias Variance Trade-off

- First approach
  VC analysis $\rightarrow Err_{out}(g) \leq Err_{in}(g) + \Omega(N, \mathcal{H}, \delta)$
- Second approach
  Bias-variance analysis is another $\rightarrow$????
  - How well $\mathcal{H}$ can approximate f ?
  - Enabling the data to zoom in on (or to pin down) the right $h \in \mathcal{H}$ (or How close can you get to that approximation with a finite data set)?

Bias Variance Trade-off

- First approach
  VC analysis $\rightarrow Err_{out}(g) \leq Err_{in}(g) + \Omega(N, \mathcal{H}, \delta)$

- Second approach
  Bias-variance analysis is another $\rightarrow$????

  - How well $\mathcal{H}$ can approximate f ?
  - Enabling the data to zoom in on (or to pin down) the right $h \in \mathcal{H}$ (or How close can you get to that approximation with a finite data set)?

Applies to *real-valued target* and uses *squared error*

Bias Variance Trade-off

$$Err_{out}(g)^{(\mathcal{D})} = \mathbb{E}_x[(g^{(\mathcal{D})}(x) - f(x))^2]$$

Bias Variance Trade-off

$$Err_{out}(g)^{(\mathcal{D})} = \mathbb{E}_x[(g^{(\mathcal{D})}(x) - f(x))^2]$$

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}[Err_{out}(g)^{(\mathcal{D})}] &= \mathbb{E}_{\mathcal{D}}[\mathbb{E}_x[(g^{(\mathcal{D})}(x) - f(x))^2]] \\
&= \mathbb{E}_x[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - f(x))^2]] \\
&= \mathbb{E}_x[\underbrace{\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - f(x))^2]}_{\text{We focus on this term}}]
\end{aligned}$$

Bias Variance Trade-off

$$\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - f(x))^2]$$

$$\bar{g}(x) = \mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x)]$$

Image many data sets $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$

$$\bar{g}(x) \approx \frac{1}{K} \sum_{k=1}^{K} (g^{(\mathcal{D})_k})$$

$$\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - f(x))^2] = \mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x) + \bar{g}(x) - f(x))^2]$$

$$= \mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2 + (\bar{g}(x) - f(x))^2$$

$$+ 2(g^{(\mathcal{D})}(x) - \bar{g}(x))(\bar{g}(x) - f(x))]$$

$$= \mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2] + (\bar{g}(x) - f(x))^2$$

$$\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - f(x))^2] = \underbrace{\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2]}_{\text{var(X)}} + \underbrace{(\bar{g}(x) - f(x))^2}_{\text{bias(x)}}$$

Putting all together,

$$\mathbb{E}_{\mathcal{D}}[Err_{out}(g)^{(\mathcal{D})}] = \mathbb{E}_x[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - f(x))^2]]$$

$$= \mathbb{E}_x[bias(x) + var(x)]$$

$$\text{bias} = \mathbb{E}_x[(\bar{g}(x) - f(x))^2] \ \text{var} = \mathbb{E}_x[\mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2]]$$

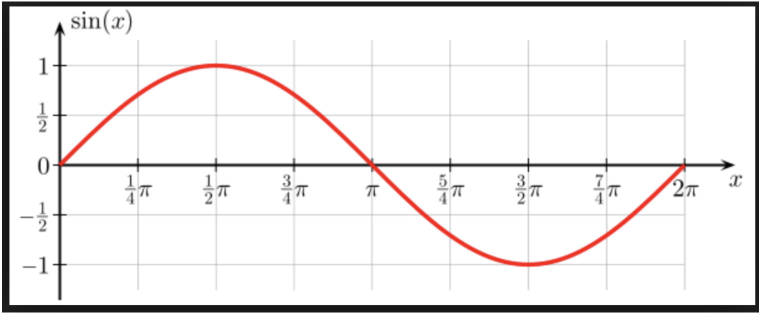
$$f : [-1, +1] \to \mathbb{R} \qquad f(x) = sin(x)$$

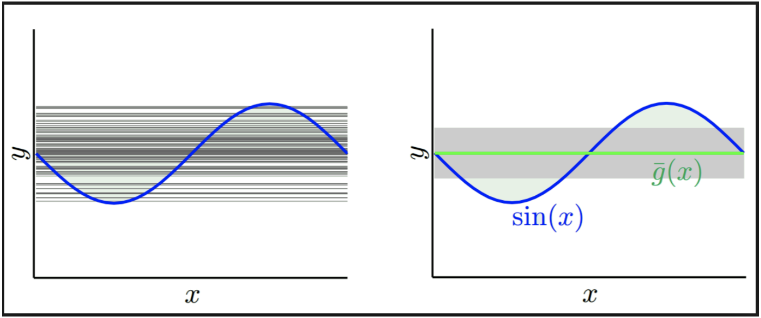
Only two training samples $\qquad N = 2$

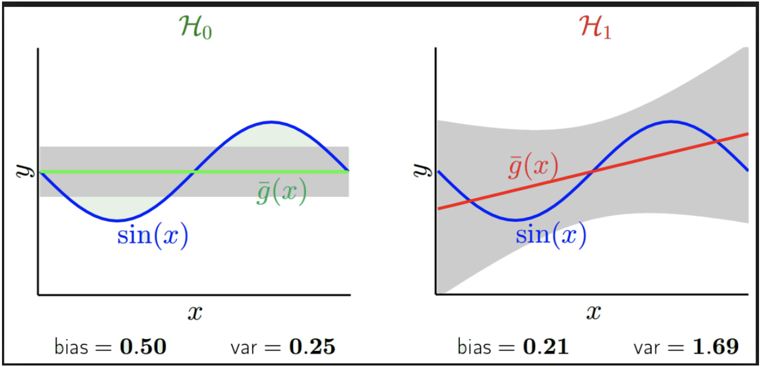
Two hypothesis:

$\mathcal{H}_0 : h(x) = b$

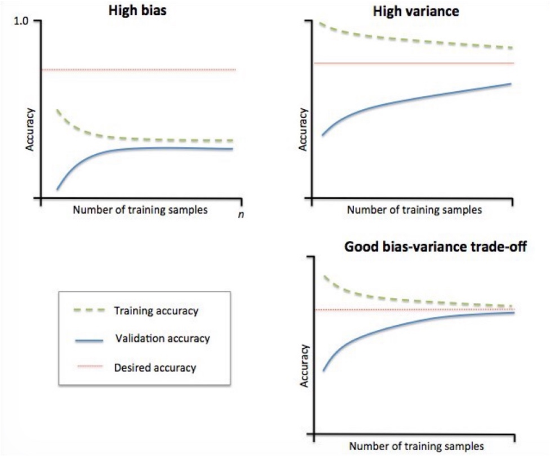
$\mathcal{H}_1 : h(x) = ax + b$


sin(x)
1
1/2
0
−1/2
−1
1/4 π
1/2 π
3/4 π
π
5/4 π
3/2 π
7/4 π
2π
x


y
x
y
$\bar{g}(x)$
$\sin(x)$
x



Remember : $E_{out} = bias + var$


High bias
1.0
Accuracy
Number of training samples
n
High variance
Accuracy
Number of training samples
Training accuracy
Validation accuracy
Desired accuracy
Good bias-variance trade-off
Accuracy
Number of training samples