Bias-Variance Trade-Off

Michał Szafraniuk

September 10, 2019

Suppose we model a phenomena whose nature is

$$y(x) = f(x) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.

Suppose the examples in training/test set are all taken from the same distribution as above so:

$$y_i = f(x_i) + \epsilon_i$$

where:

- i iterates over both training and test set
- $\epsilon_i \sim N(0, \sigma^2)$ for all i
- $\epsilon_i \perp \epsilon_j$ for $i \neq j$

Our goal is to compute f. By looking at the training set, we obtain and estimate $\hat{f}$ and now use this estimate over test set, i.e. for each j in test set our prediction for $y_j = f(x_j) + \epsilon_j$ is $\hat{f}(x_j)$.

Observations:

- ϵ_j is real, random variable
- f is just a function, x_j is fixed, real number so $f(x_j)$ is fixed (nonrandom)
- $\hat{f}(x_j)$ is random since it depends on values of ϵ_i from the training set
- $\epsilon_j \perp \hat{f}(x_j)$ because training and test sets are disjoint, $\hat{f}(x_j)$ depends on ϵ_i from the training set and $\epsilon_i \perp \epsilon_j$ for $i \neq j$

Now we can compute **test MSE**:

$$\begin{aligned} MSE^{test} &= \mathbb{E}[(y - \hat{f})^2] \\ &= \mathbb{E}[(\epsilon + f - \hat{f})^2] \\ &= \mathbb{E}[\epsilon^2] + \mathbb{E}[(f - \hat{f})^2] \\ &= \sigma^2 + \left(\mathbb{E}[f - \hat{f}]\right)^2 + \mathbb{V}[f - \hat{f}] \\ &= \sigma^2 + (b(\hat{f}))^2 + \mathbb{V}[\hat{f}] \end{aligned}$$

To sum up:

- **bias** is defined as $b(\hat{f}) = \mathbb{E}[f - \hat{f}]$
- by **variance** we mean $\mathbb{V}[\hat{f}]$
- as complexity/flexibility of our estimate $\hat{f}$ increases then:
  - bias decreases
  - variance increases
- the typical test MSE is U-shaped with minimum representing the best bias-variance tradeoff
- **underfitting** $\leftrightarrow$ **high bias problem**
- **overfitting** $\leftrightarrow$ **high variance problem**
- **noisy data** $\leftrightarrow$ **high** σ^2