Stanford University

CS 329P : Practical Machine Learning (2021 Fall)

5. Model Combination

Qingqing Huang, Mu Li, Alex Smola

https://c.d2l.ai/stanford-cs329p

So far…

- Data
- ML Models for different types of data
- Good models perform well on unseen data
  - Model specific metrics VS business metrics
  - Generalization error depends on model / data complexity
  - TODAY: Methods for reducing generalization error

Stanford University

CS 329P : Practical Machine Learning (2021 Fall)

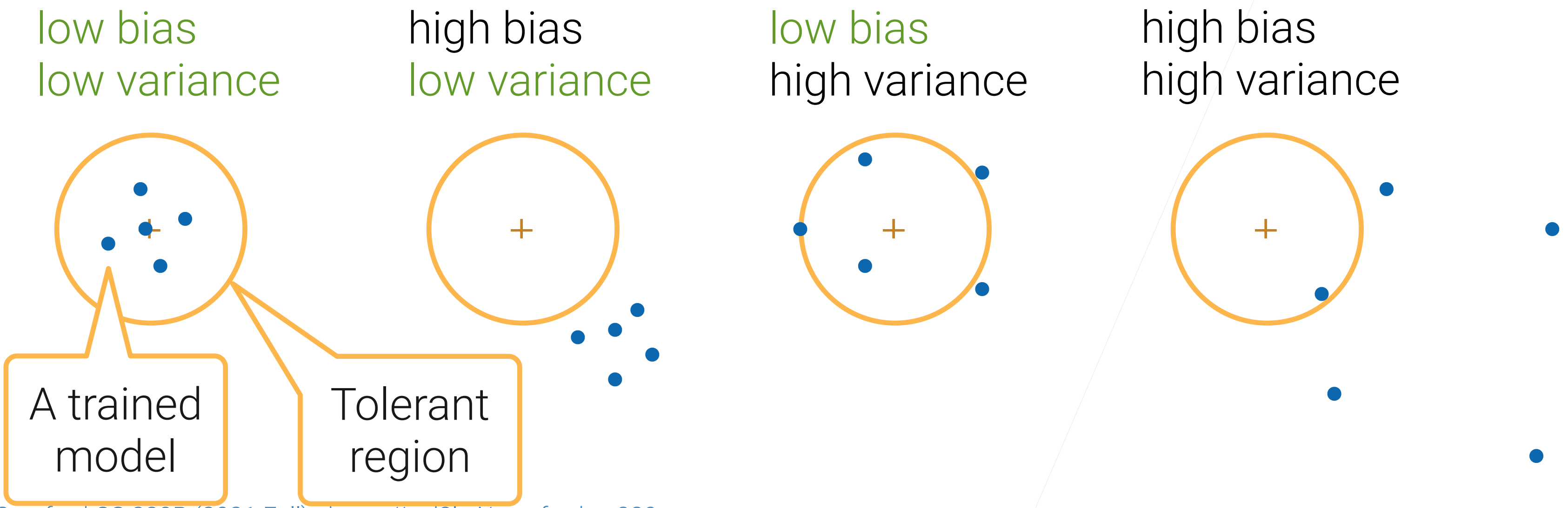
5.1 Bias & Variance

Qingqing Huang, Mu Li, Alex Smola

https://c.d2l.ai/stanford-cs329p

Bias & Variance

- In statistic learning, assume a round truth $y = f(x) + \varepsilon$
- Given a set of randomly sampled data, we can train a model $\hat{f}(x)$
- Ideally, $\hat{f}(x)$ is close to $f(x)$ for any sample train dataset.



Bias-Variance Decomposition

- Sample data $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ from $y = f(x) + \varepsilon$
- Learn $\hat{f}$ from D by minimizing MSE

  we want it generalizes well to an unseen data point (x, y)

$$\mathrm{E}_D\left[(y - \hat{f}(x))^2\right] = \mathrm{E}\left[\left((f - \mathrm{E}[\hat{f}]) - (\hat{f} - \mathrm{E}[\hat{f}]) + \varepsilon\right)^2\right]$$

$$= (f - \mathrm{E}[\hat{f}])^2 + \mathrm{E}\left[(\hat{f} - \mathrm{E}[\hat{f}]))^2\right] + \mathrm{E}[\varepsilon^2]$$

$$= \mathrm{Bias}[\hat{f}]^2 + \mathrm{Var}[\hat{f}] + \sigma^2$$

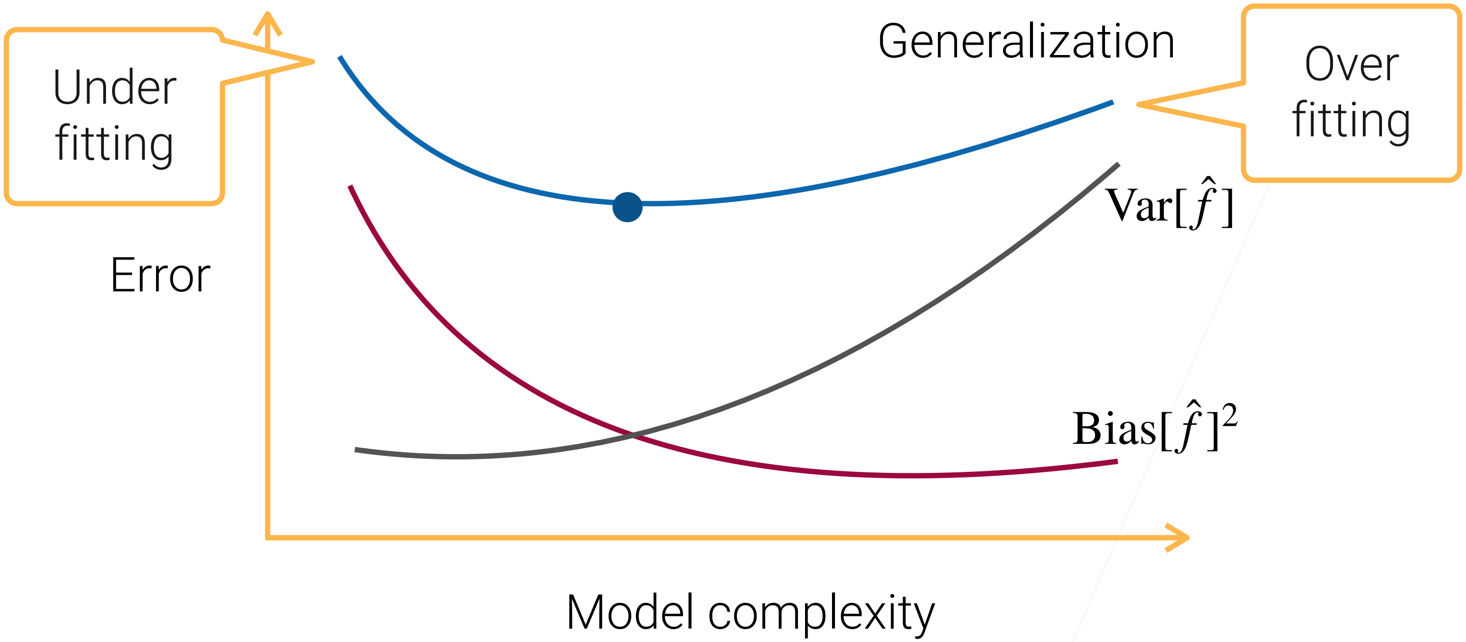
$\mathrm{E}[f] = f$

$\mathrm{E}[\varepsilon] = 0, \mathrm{Var}[\varepsilon] = \sigma^2$

ε is independent of $\hat{f}$

Bias-Variance Tradeoff

$$\mathrm{E}_D\left[(y-\hat{f}(x))^2\right] = \mathrm{Bias}[\hat{f}]^2 + \mathrm{Var}[\hat{f}] + \sigma^2$$



Reduce Bias & Variance

$$E_D\left[(y-\hat{f}(x))^2\right] = \mathrm{Bias}[\hat{f}]^2 + \mathrm{Var}[\hat{f}] + \sigma^2$$

- Reduce bias
  - A more complex model
    - e.g. increase #layers, #hidden units of MLP
  - Boosting
  - Stacking
- Reduce variance
  - A simpler model
    - e.g. regularization
  - Bagging
  - Stacking
- Reduce σ^2
  - Improve data

Ensemble learning: train and combine multiple models to improve predictive performance