

AAI/CPE/EE 595-WS1

Applied Machine Learning

Lecture 5: Cross Validation and Bias-Variance Trade-offs

Speaker: Jiarui Li
jli148@stevens.edu

Prediction Errors: Bias-Variance (BV) Trade-off

Statistically, define:

Y: target value (variable we want to predict)

X: input data (covariates)

$$Y = f(X) + \varepsilon$$

where $\varepsilon \sim N(0, \sigma_\varepsilon)$ is the error term.

$\hat{f}(X)$: trained model of $f(X)$

The expected squared error at point x :

$$Err(x) = E \left[\left(Y - \hat{f}(x) \right)^2 \right] = \underbrace{E \left[\left(E[\hat{f}(x)] - f(X) \right)^2 \right]}_{\text{Bias}} + \underbrace{E \left[\left(\hat{f}(x) - E[\hat{f}(x)] \right)^2 \right]}_{\text{Variance}} + \sigma_\varepsilon^2$$

Prediction Errors: Bias-Variance (BV) Trade-off

Bias Problem

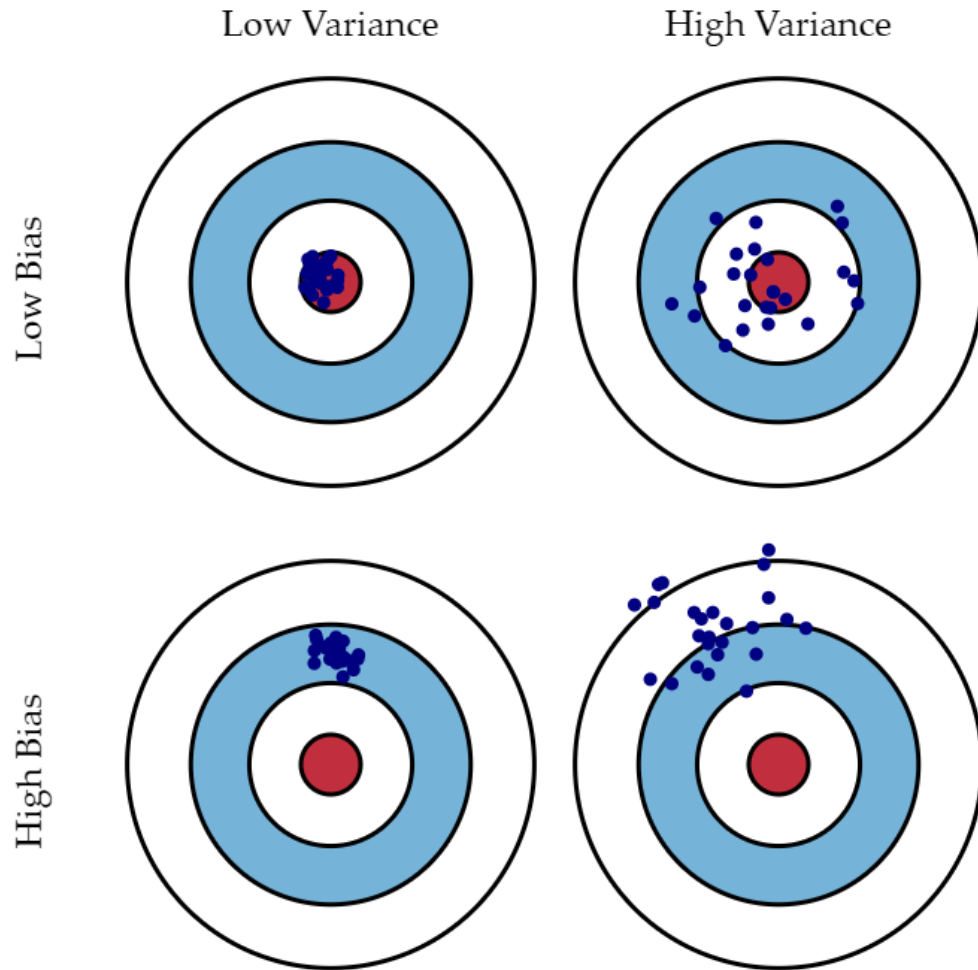
The hypothesis space made available by a particular classification method does **not** include **sufficient** hypotheses

Variance Problem

The hypothesis space made available is **too large** for the training data, and the selected hypothesis may not be accurate on unseen data



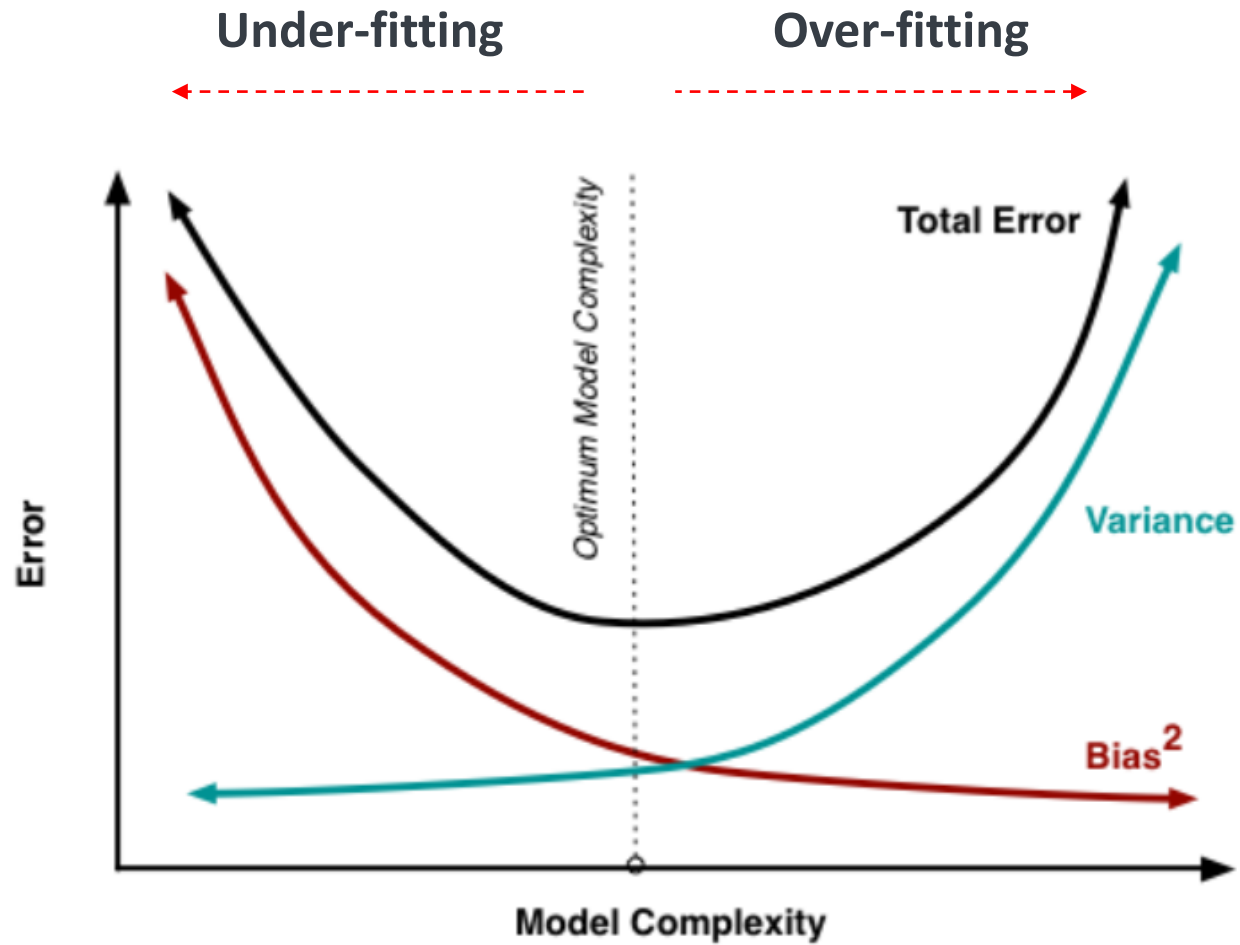
Prediction Errors: Bias-Variance (BV) Trade-off



Assuming the center (red circle) is the perfect model



Bias-Variance Trade-off



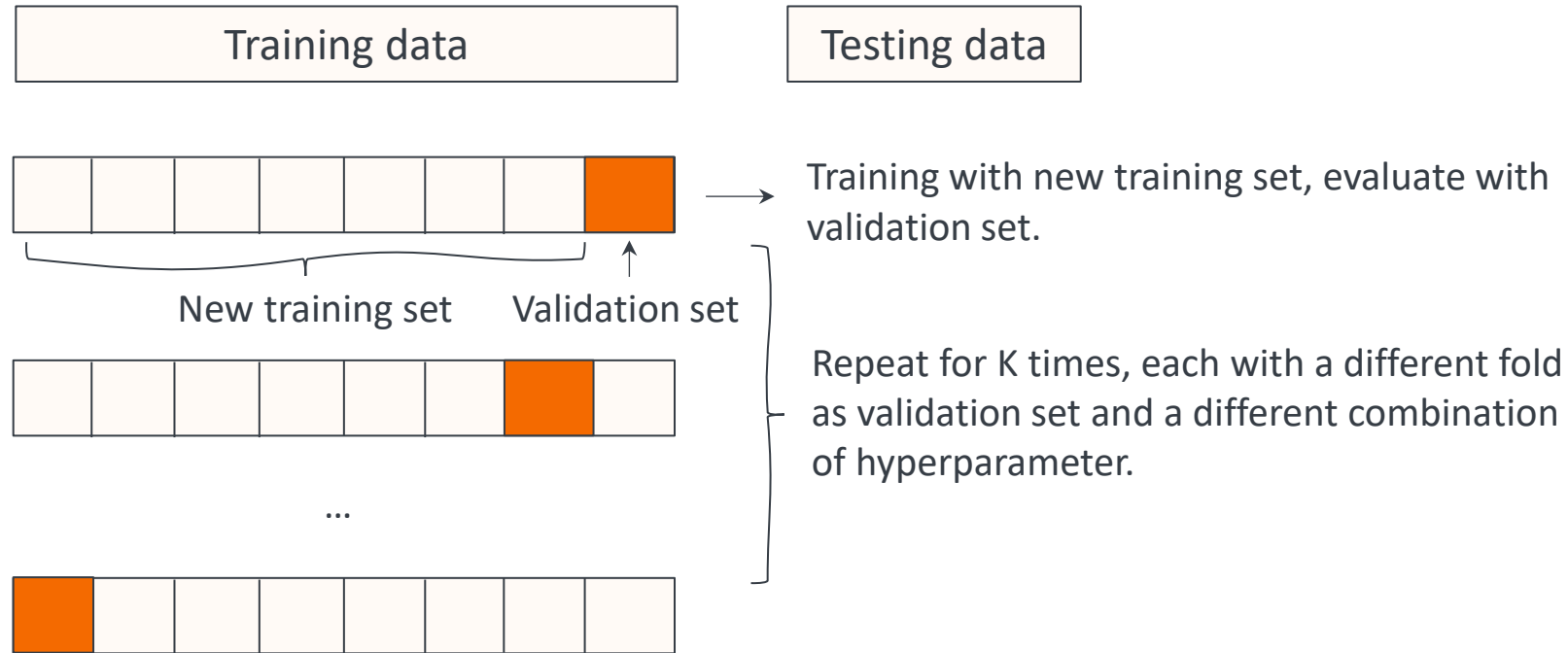
Prediction Errors: Bias-Variance (BV) Trade-off

- Practical techniques to reduce bias:
Increase hypothesis space (i.e., model complexity)
- Practical techniques to reduce variance:
 - resampling (e.g., random forest)
 - * bias of each tree is the same as full model but higher variance
 - * averaging many trees decreases variance without increasing bias
 - * theoretically the more trees, the less variance (if no computation limit)
- No analytical methods to find optimal bias-variance trade-off
 - * try different model complexity (needing accurate error measurement)



K-Fold Cross-Validation

To determine the best hyperparameters when training data, we can further divide training data into K folds, with K-1 folds as new training set and one as hold-out set (validation set):



At the end, train the model using all training data with the optimal hyperparameters, and test it using the testing data.





THANK YOU

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030