VBM683
Machine Learning

Pinar Duygulu

Slides are adapted from
Dhruv Batra

Model Selection

- How do we pick the right model class?

- Similar questions
  - How do I pick magic hyper-parameters?
  - How do I do feature selection?

Errors

- Expected Loss/Error

- Training Loss/Error
- Validation Loss/Error
- Test Loss/Error

- Reporting Training Error (instead of Test) is CHEATING

- Optimizing parameters on Test Error is CHEATING

Cross-validation

- The improved holdout method: k-fold *cross-validation*
  - Partition data into k roughly equal parts;
  - Train on all but j-th part, test on j-th part

Cross-validation

- The improved holdout method: k-fold *cross-validation*
  - Partition data into k roughly equal parts;
  - Train on all but j-th part, test on j-th part

Cross-validation

- The improved holdout method: k-fold *cross-validation*
  - Partition data into k roughly equal parts;
  - Train on all but j-th part, test on j-th part

Cross-validation

- The improved holdout method: k-fold *cross-validation*
  - Partition data into k roughly equal parts;
  - Train on all but j-th part, test on j-th part

Cross-validation

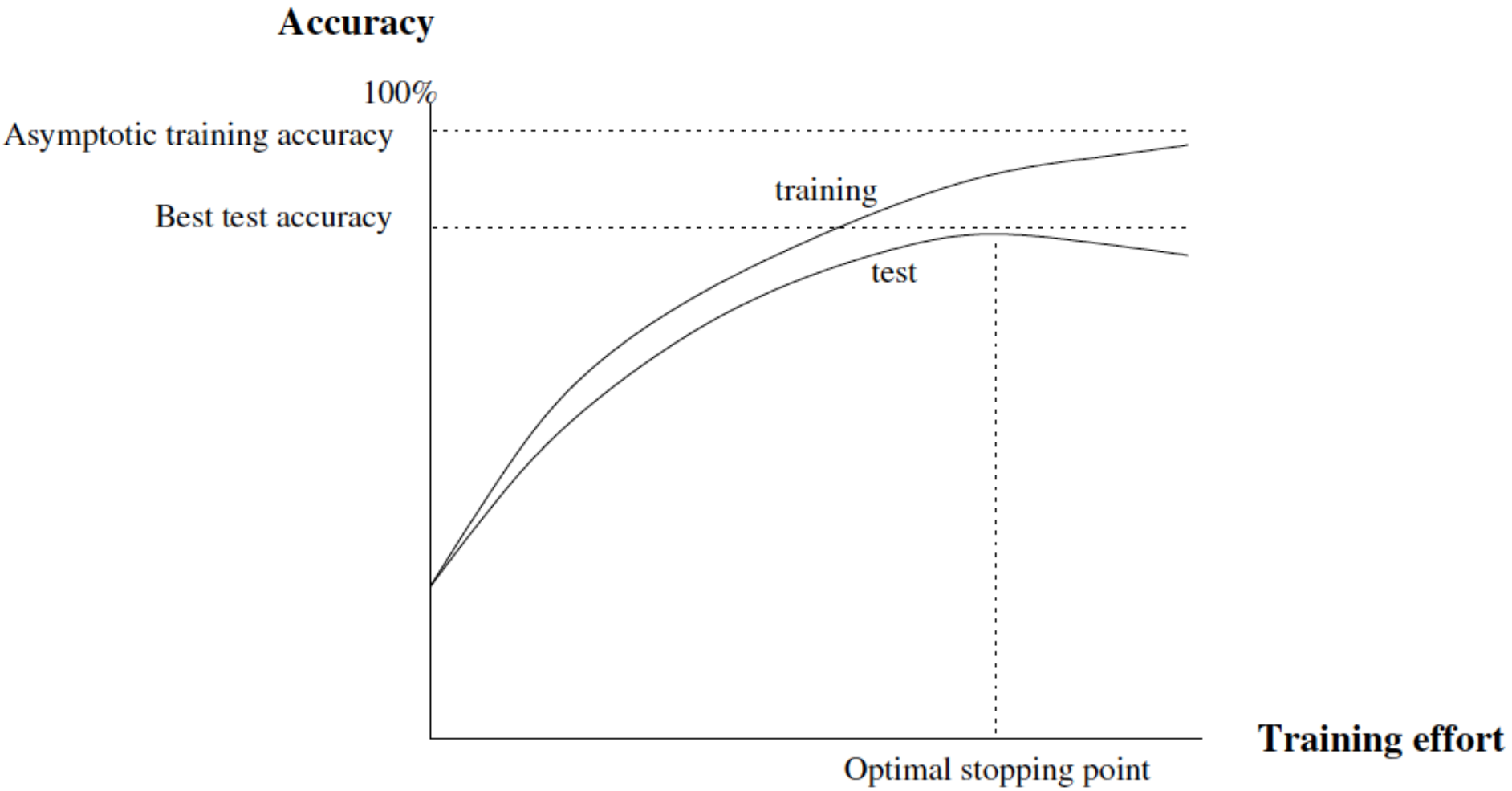
- The improved holdout method: k-fold *cross-validation*
  - Partition data into k roughly equal parts;
  - Train on all but j-th part, test on j-th part

- An extreme case: *leave-one-out* cross-validation

$$\hat{L}_{\text{cv}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(\mathbf{x}_i; \hat{\mathbf{w}}_{-i}))^2$$

where $\hat{\mathbf{w}}_{-i}$ is fit to all the data but the i-th example.

Typical Behavior

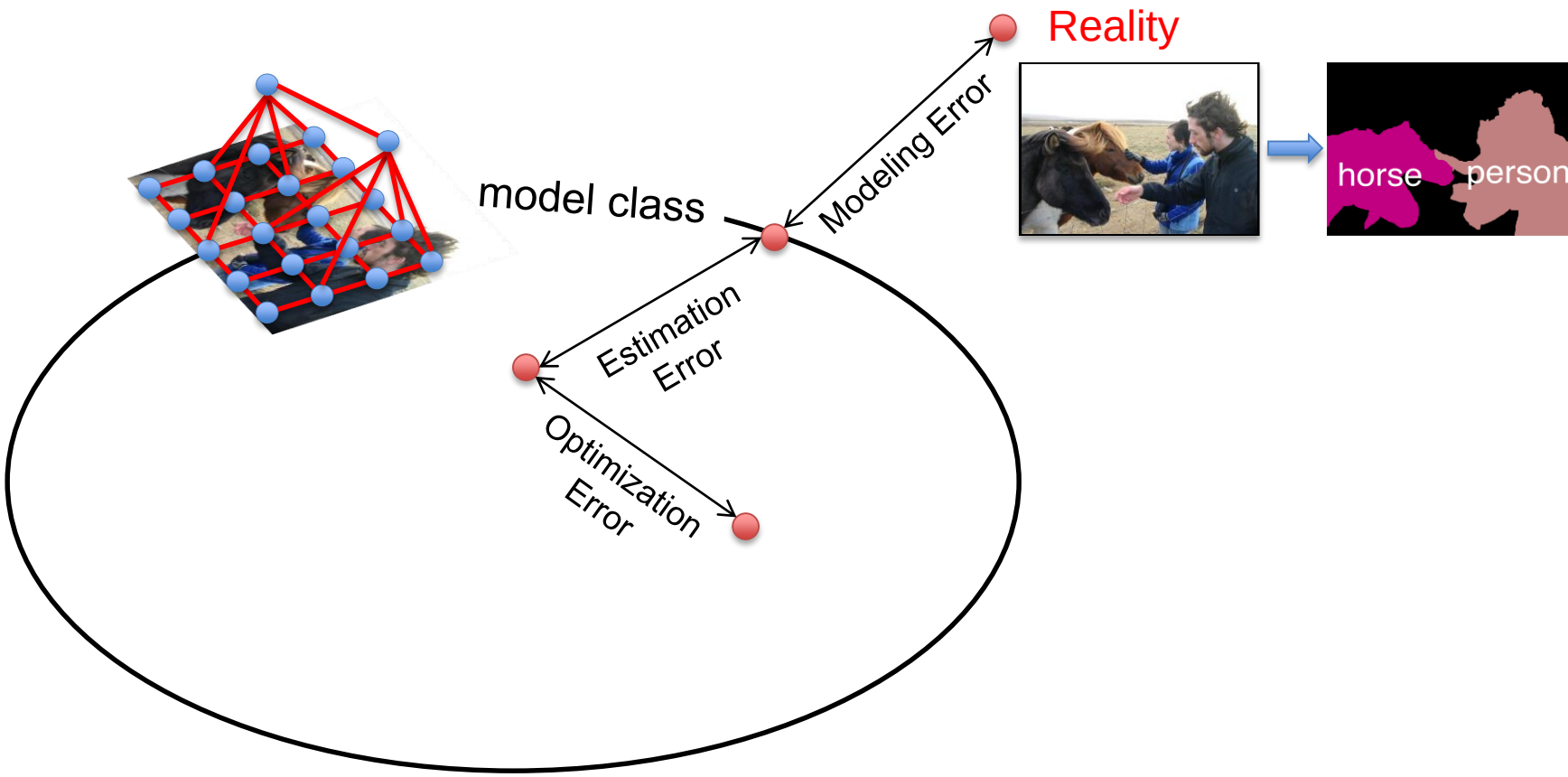


Overfitting

- **Overfitting:** a learning algorithm overfits the training data if it outputs a solution **w** when there exists another solution **w'** such that:

$$[error_{train}(\mathbf{w}) < error_{train}(\mathbf{w}')] \wedge [error_{true}(\mathbf{w}') < error_{true}(\mathbf{w})]$$

Error Decomposition



Error Decomposition

- Approximation/Modeling Error
  - You approximated reality with model
- Estimation Error
  - You tried to learn model with finite data
- Optimization Error
  - You were lazy and couldn't/didn't optimize to completion

Bias-Variance Tradeoff

- **Bias:** difference between what you expect to learn and truth
  - Measures how well you expect to represent true solution
  - Decreases with more complex model

- **Variance:** difference between what you expect to learn and what you learn from a from a particular dataset
  - Measures how sensitive learner is to specific dataset
  - Increases with more complex model

Debugging Machine Learning

- My algorithm doesn't work
  - High test error

- What should I do?
  - More training data
  - Smaller set of features
  - Larger set of features
  - Lower regularization
  - Higher regularization