Review

Lecture 14

Termeh Shafie

Estimate $\hat{f}$ = Learn $\hat{f}$

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

sources of error:
irreducible error ϵ
reducible error $\hat{f}$

the squared error for a given estimate $\hat{f}$ is

$$E(\text{actual} - \text{predicted})^2 = E(Y - \hat{Y})^2$$

which factors as

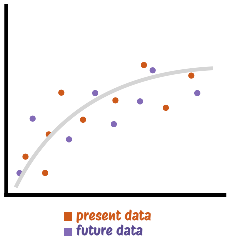
$$E[f(X) + \epsilon - \hat{f}(X)]^2$$

$$\underbrace{[f(X) - \hat{f}(X)^2]}_{\text{reducible}} + \underbrace{\text{Var}(\epsilon)}_{\text{irreducible}}$$

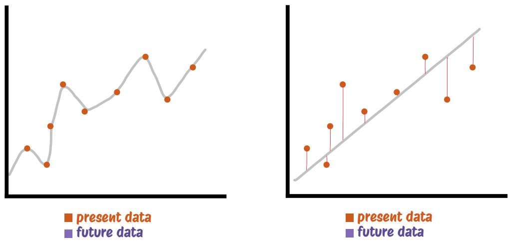
Training	Testing
training data set	testing data sets (unseen)
$\{(y_1, x_1,), \ldots, (y_n, x_n))\}$	(y_0, x_0)
used to find function q that minimizes **Training MSE**	used to compute **Test MSE**
$\hat{f} = \arg\min_q MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - q(x_i))^2$	$E[y_0 - \hat{f}(x_0)^2]$

`often not so closely related`

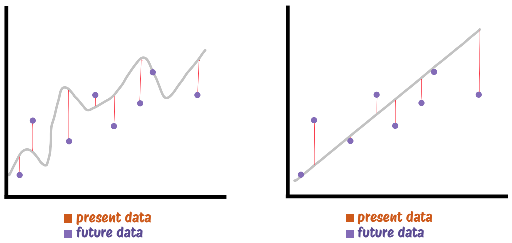
Training and Testing



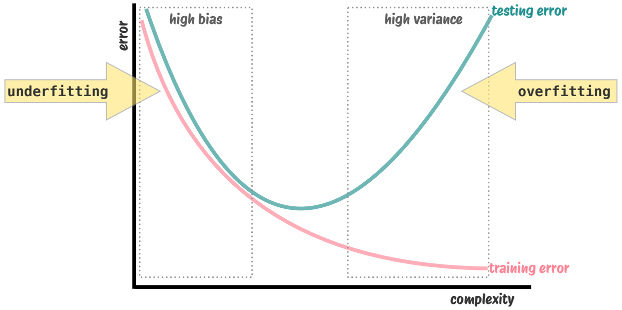
Training and Testing



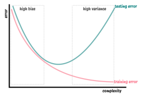
Training and Testing



Bias Variance Trade-Off



Formalizing Bias Variance Trade-Off

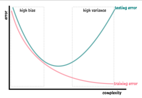


Expected test MSE

$$\underbrace{E\left(y_0 - \hat{f}(x_0)\right)^2}_{\text{expected MSE at } x_0 \text{ if we repeatedly estimated } f(x) \text{ with different training sets}} = \text{Var}(\hat{f}(x_0)) + \left[\text{bias}(\hat{f}(x_0))\right]^2 + \underbrace{\text{Var}(\epsilon)}_{\text{irreducible error}}$$

[try it out: https://floswald.shinyapps.io/bias_variance/]

Formalizing Bias Variance Trade-Off

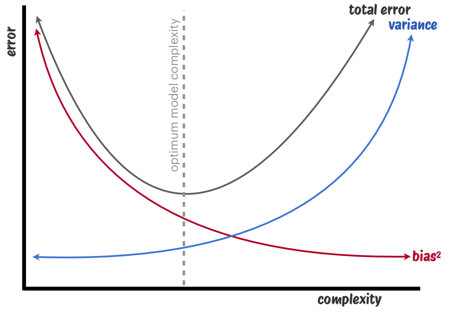


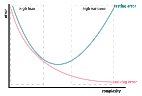
Expected test MSE

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \underbrace{\text{Var}(\hat{f}(x_0))}_{\text{variance increases with complexity}} + \underbrace{\left[\text{bias}(\hat{f}(x_0))\right]^2}_{\text{bias decreases with complexity}} + \text{Var}(\epsilon)$$

[try it out: https://floswald.shinyapps.io/bias_variance/]

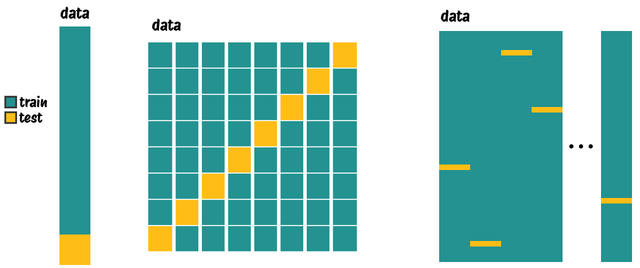
Bias Variance Trade-Off





Model Validation

cross validation **simulates multiple train-test-splits** on the training data



Loss Functions

Common loss functions (continuous):

- MSE
- MAE

⋮

Common loss functions (categorical):

- Log Loss/Binary Cross Entropy
- Hinge Loss

⋮

Loss Function
a **metric** for model **performance**, lower values are better

Classification Metrics

		Predicted	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Negative (TN)
	Negative	False Positive (FP)	True Negative (TN)

Hyperparameter Tuning

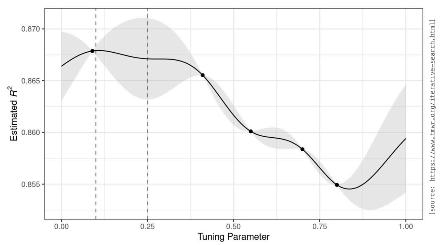
- Parameters: values in your model that your model chooses (e.g. coefficients)
- Hyperparameters: values in your model that your model does NOT choose (e.g. K in KNN, max_depth in DTs)

Two options:

1. Choose the hyperparameter based on domain knowledge
2. Choose the hyperparameter using hyperparameter tuning
   Train-Test-Validation: The validation set serves as a pretend-test-set for us to see how well different hyperparameter values do on unseen data without touching our actual test set

Hyperparameter Tuning

- Grid search
- Random search
- Bayesian optimization



Regression vs. Classification

Common Regression Models
- Linear Regression
- Polynomial Regression
- Regression Trees/Random Forests
- Neural Nets

⋮

Common Classification Models
- Logistic Regression
- Decision Trees
- KNN
- Support Vector Machines
- Naive Bayes
- Neural Nets

⋮

Supervised vs. Unsupervised

Supervised
- Has labeled data (we know the correct answers and use them to train the model)
- Goal: to accurately predict our target value
- e.g. classification or regression

Unsupervised
- Does not have labeled data (there are no correct answers)
- Goal: to create/recognize latent structure in the data
- e.g. PCA, clustering

we made it!