EE2211 Pre-Tutorial 7

Dr Feng LIN
feng_lin@nus.edu.sg

Agenda

- Recap
- Self-learning
- Tutorial 7

Today's Attendance

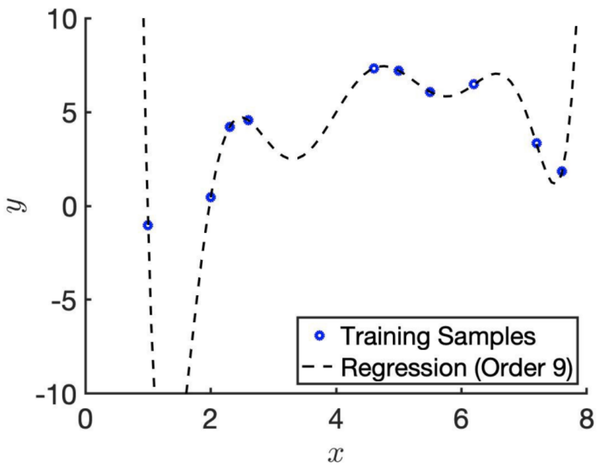
Recap

- Overfitting, underfitting & model complexity
  - Overfitting: low error in training set, high error in test set
  - Underfitting: high error in both training & test sets
  - Overly complex models can overfit; Overly simple models can underfit
- Feature selection
  - Extract useful features from training set
- Regularization (e.g., L2 regularization)
  - Solve “ill-posed” problem (e.g., more unknowns than data points)
  - Reduce overfitting
- Bias-Variance Decomposition Theorem
  - Test error = Bias Squared + Variance + Irreducible Noise
  - Can be interpreted as trading off bias & variance:
    - Overly complex models can have high variance, low bias
    - Overly simple models can have low variance, high bias

Overfitting

Training

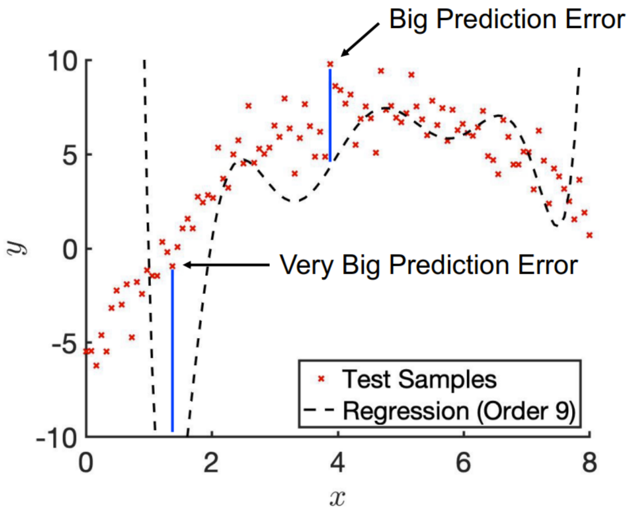
Overfitting Example



	Training Set Fit	
Order 9	Good	

Testing

Overfitting Example

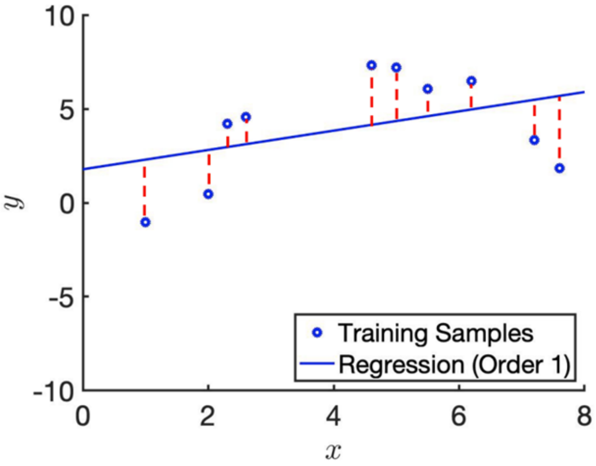


	Training Set Fit	Test Set Fit
Order 9	Good	Bad

Underfitting

Training

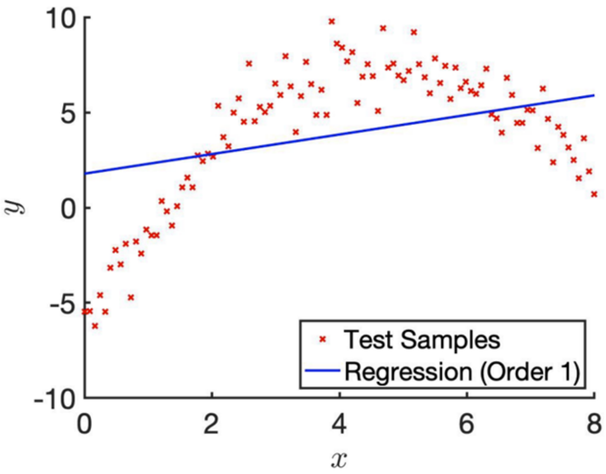
Underfitting Example



	Training Set Fit	Test Set Fit
Order 9	Good	Bad
Order 1	Bad	

Testing

Underfitting Example

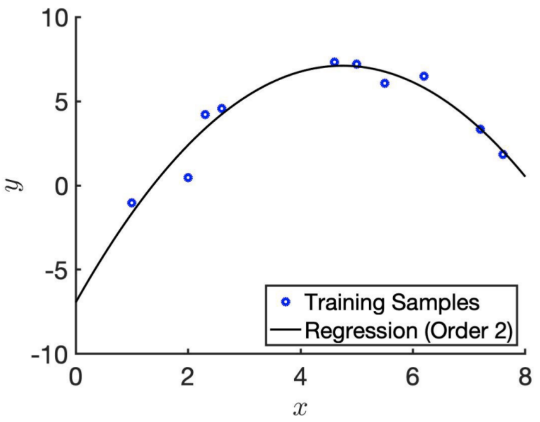


	Training Set Fit	Test Set Fit
Order 9	Good	Bad
Order 1	Bad	Bad

Perfect Fitting

Training

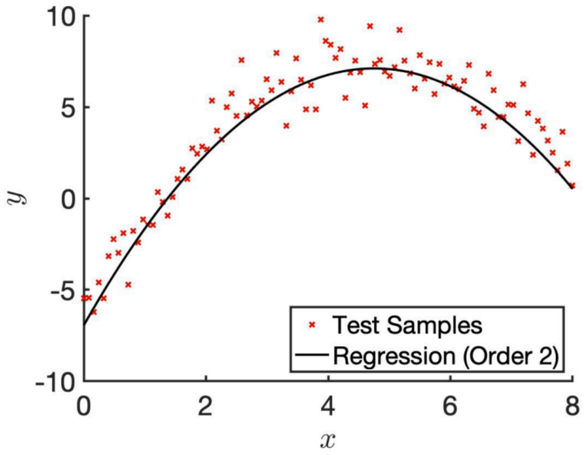
"Just Nice"



	Training Set Fit	Test Set Fit
Order 9	Good	Bad
Order 1	Bad	Bad
Order 2	Good	

Testing

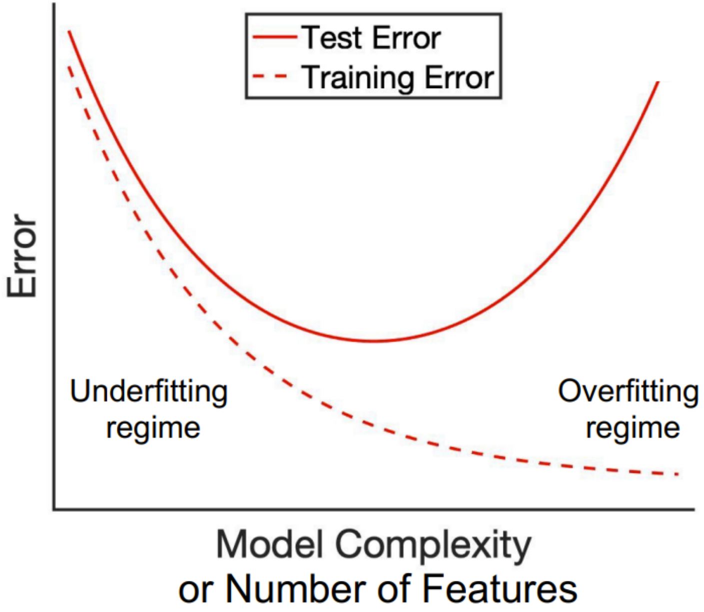
"Just Nice"



	Training Set Fit	Test Set Fit
Order 9	Good	Bad
Order 1	Bad	Bad
Order 2	Good	Good

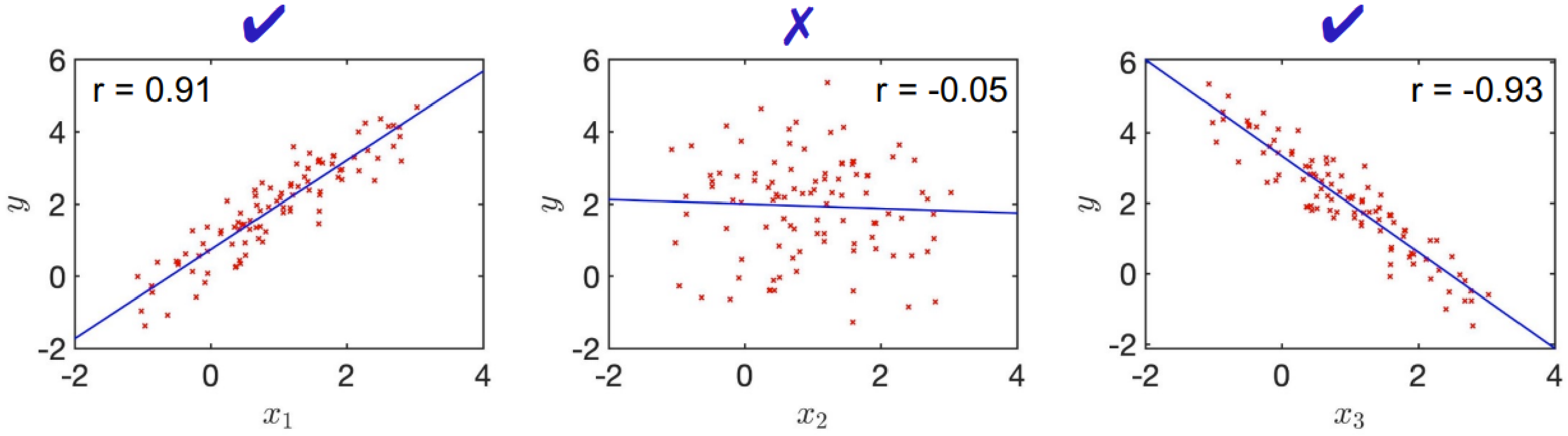
Fitting VS Model Complexity

Overfitting / Underfitting Schematic



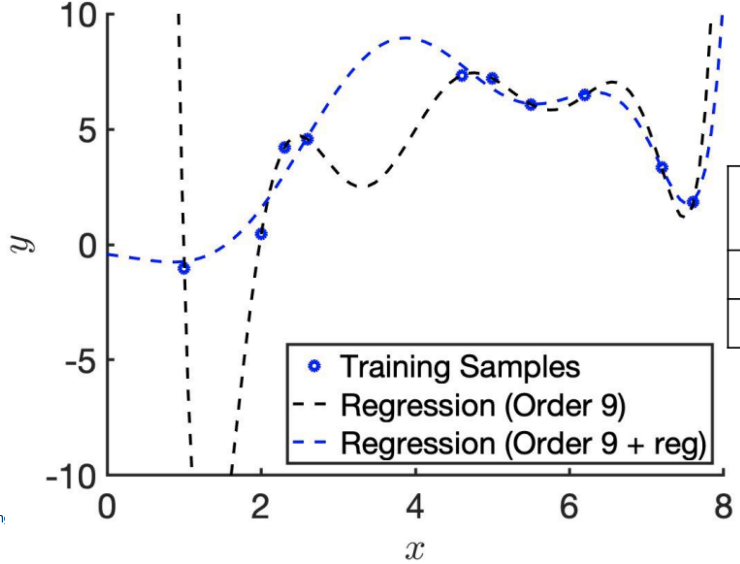
Pearson's R

– Pearson's correlation r measures linear relationship between two variables



Regularization

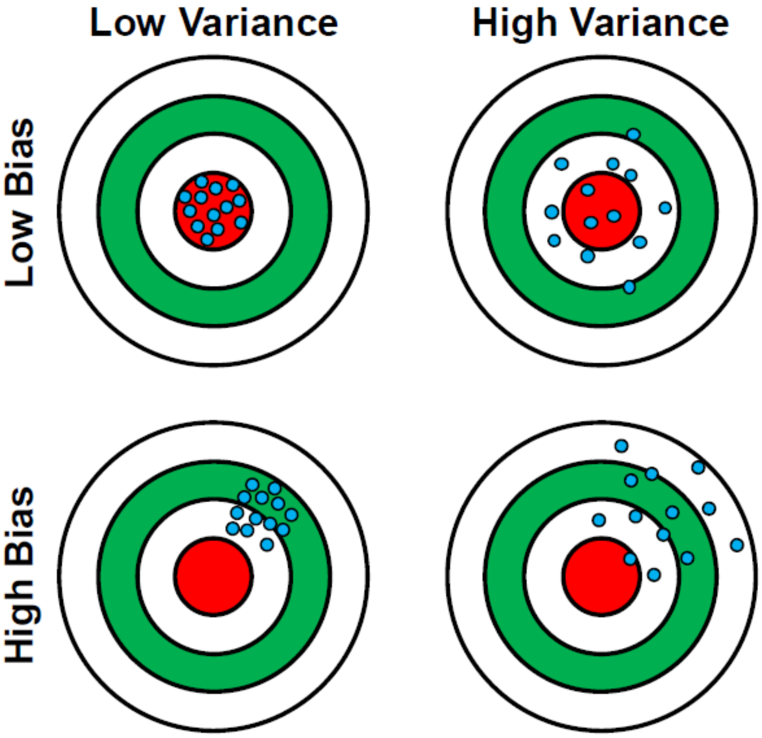
$$\underset{\mathbf{w}}{\operatorname{argmin}} \underbrace{(\mathbf{P}\mathbf{w}-\mathbf{y})^T(\mathbf{P}\mathbf{w}-\mathbf{y})}_{\text{Cost function quantifying data fitting error in training set}} + \underbrace{\lambda \mathbf{w}^T\mathbf{w}}_{\text{Regularization}}$$



	Training Set Fit	Test Set Fit
Order 9	Good	Bad
Order 9, $\lambda = 1$	Good	

Bias vs Variance

- Suppose we are trying to predict red target below:



Low Bias: blue predictions on average close to red target
Low Variance: low variability among blue predictions

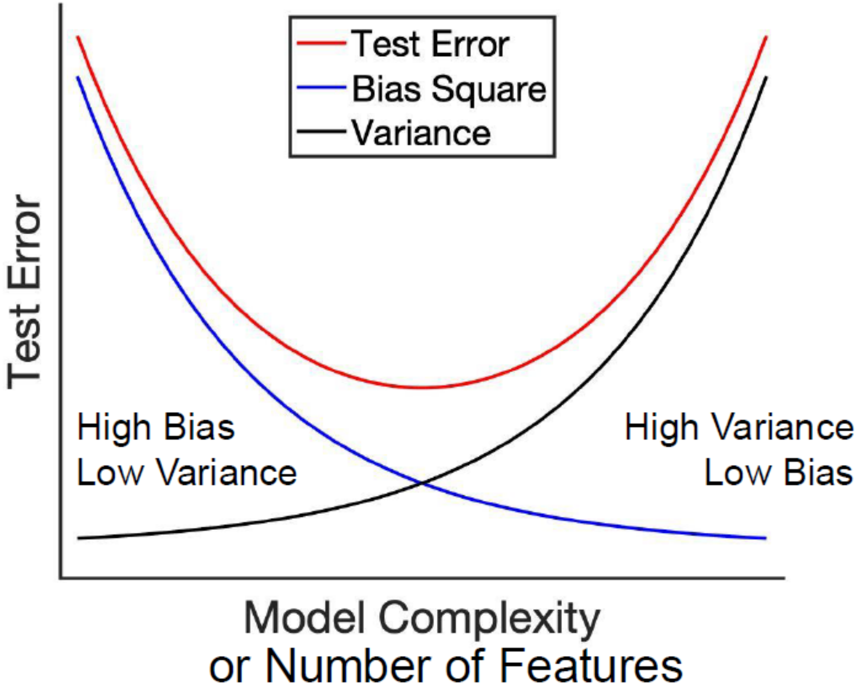
Low Bias: blue predictions on average close to red target
High Variance: large variability among blue predictions

High Bias: blue predictions on average not close to red target
Low Variance: Low variability among blue predictions

High Bias: blue predictions on average not close to red target
High Variance: high variability among blue predictions

Bias + Variance Trade Off

- Test error = Bias Squared + Variance + Irreducible Noise



Bias-Variance Decomposition Theorem

- **Test error** = Bias Squared + Variance + Irreducible Noise
  - Mathematical details in optional uploaded material (won't be tested)
- "**Variance**" refers to variability of prediction models across different training sets
  - In previous example, every time the training set of 10 samples changes, the trained model changes
  - "Variance" quantifies variability across trained models
- "**Bias**" refers to how well an average prediction model will perform
  - In previous example, every time the training set of 10 samples changes, the trained model changes
  - If we average the trained models, how well will this average trained model perform?
- "**Irreducible Noise**" reflects the fact that even if we are perfect modelers, it might not be possible to predict target y with 100% accuracy from feature(s) x

THANK YOU