Theory, Practice & Products

The three perspectives of machine learning

Approaches

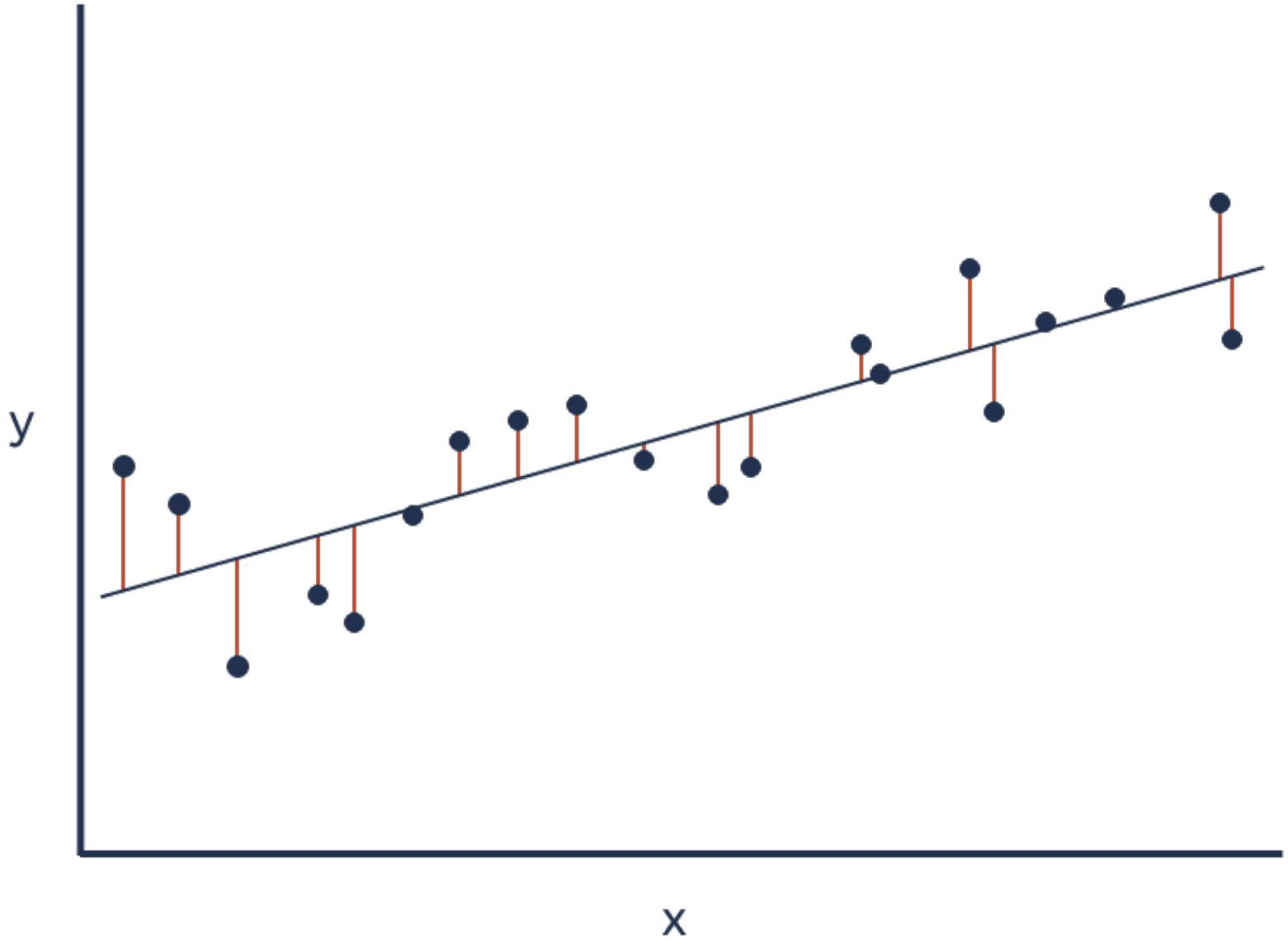
Roles

- Software Engineers
- Data Analysts
- Data Engineers
- Machine Learning Engineers
- Data Scientists
- Research Engineers

Theory

$$f(x) = w_1 x + w_0$$

$$f(x) = w_1 x + w_0 + \epsilon$$


y
x

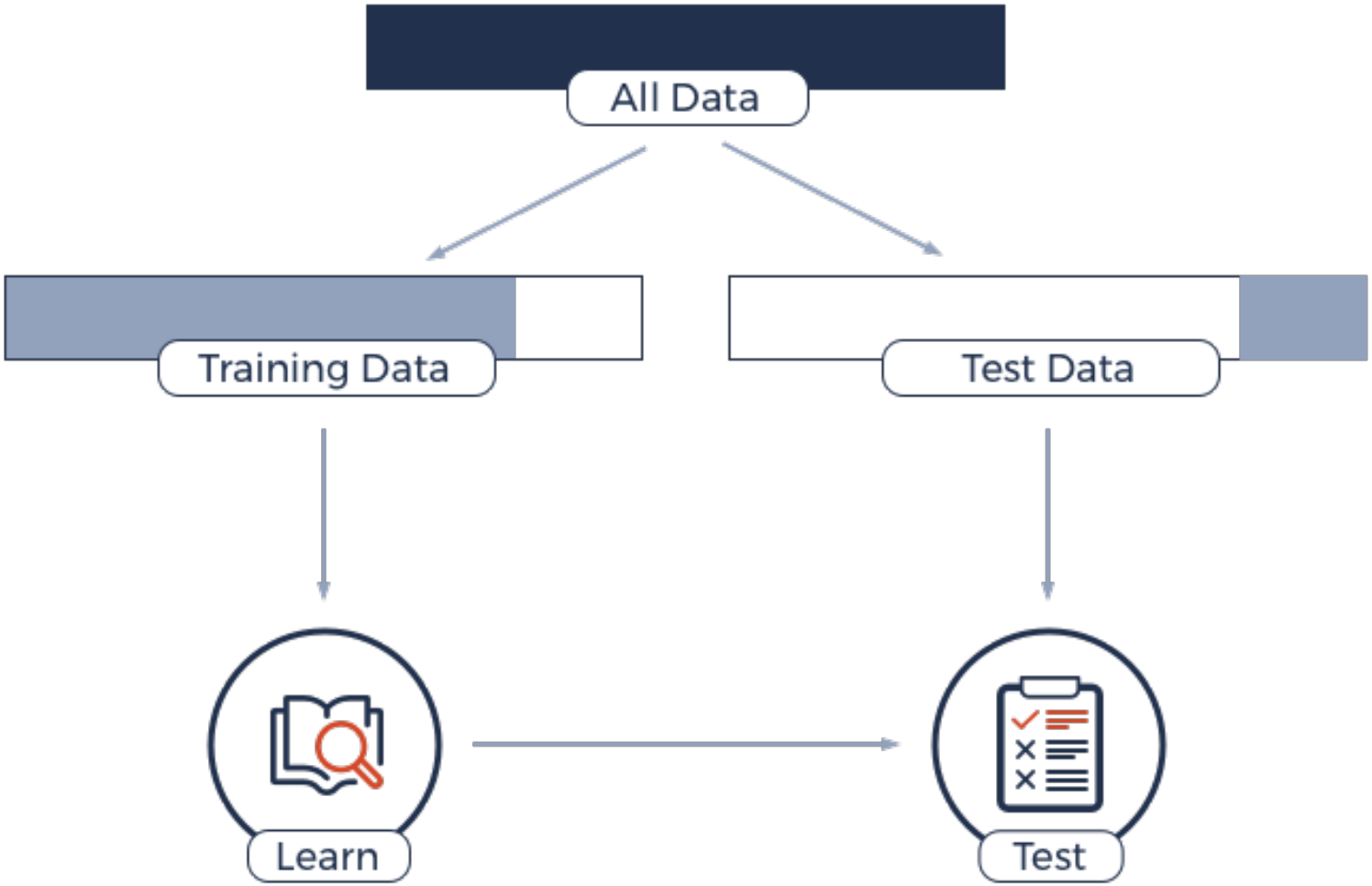
$$\min_{w_1, w_0} \sqrt{\sum_i^n (y_i - (w_1 x_i + w_0))^2}^2$$

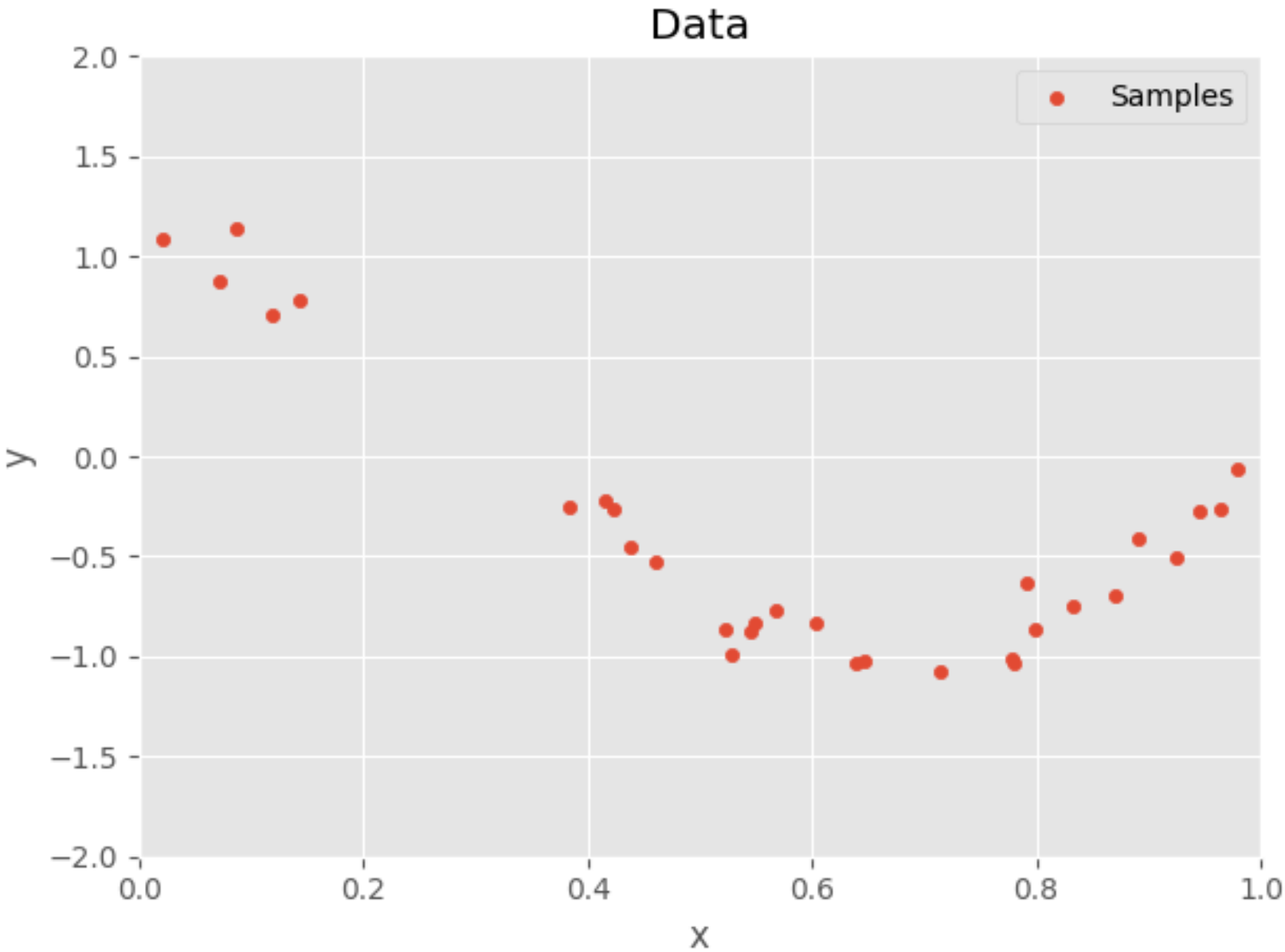
$$\min_{\mathbf{w}} \; ||\mathbf{y} - \mathbf{w}^T\mathbf{X}||_2^2$$

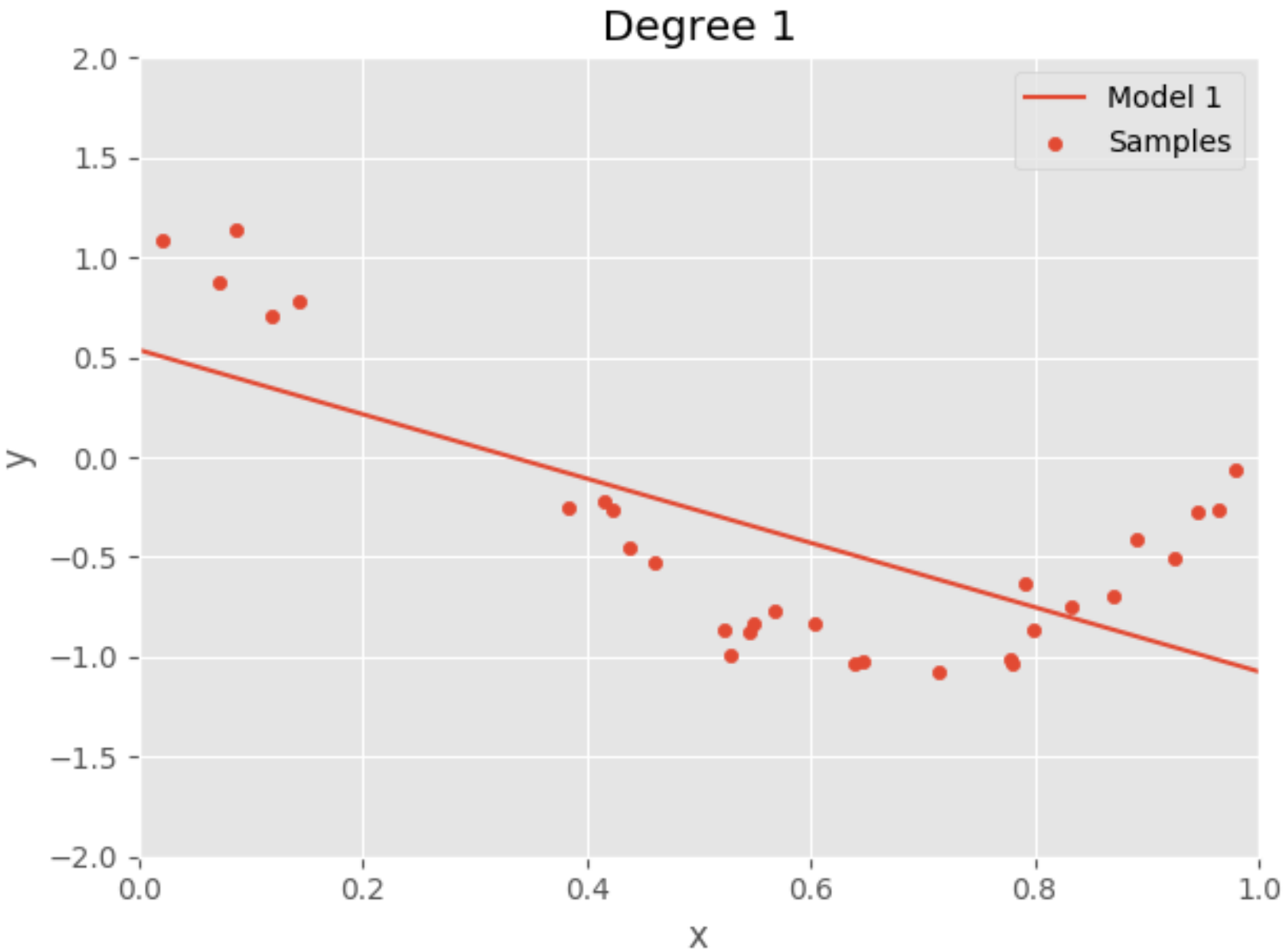
$$\min_{\mathbf{w}} \; ||\mathbf{y} - \mathbf{w}^T\mathbf{X}||_2^2$$

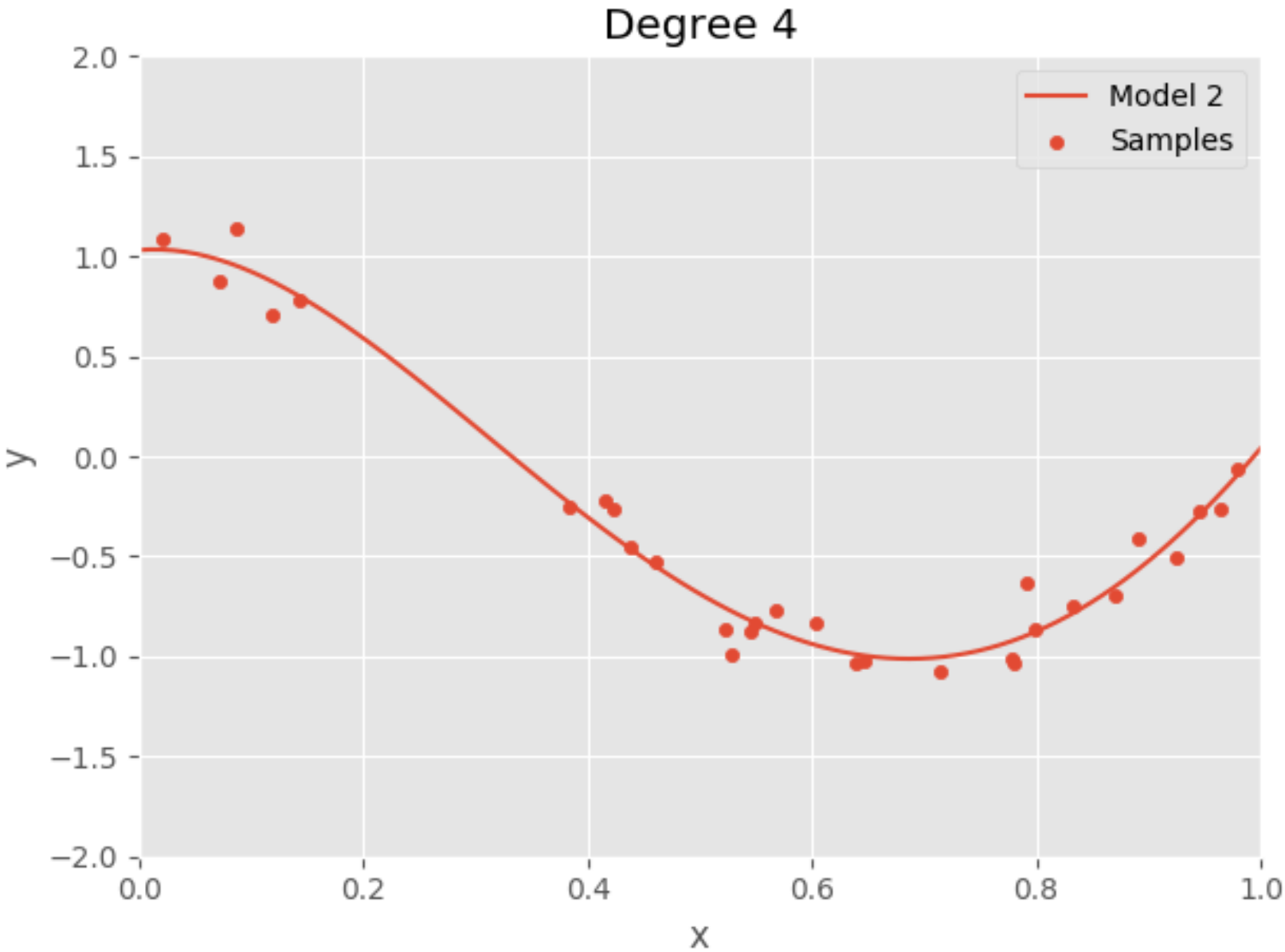
$$\min_{\mathbf{w}} \left\| \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \begin{bmatrix} x_1^0 & x_1^1 & x_1^2 & \dots & x_1^m \\ x_2^0 & x_2^1 & x_2^2 & \dots & x_2^m \\ x_3^0 & \ddots & & & \\ \vdots & & \ddots & & \\ x_n^0 & \dots & \dots & \dots & x_n^m \end{bmatrix} \right\|_{2,1}^2$$

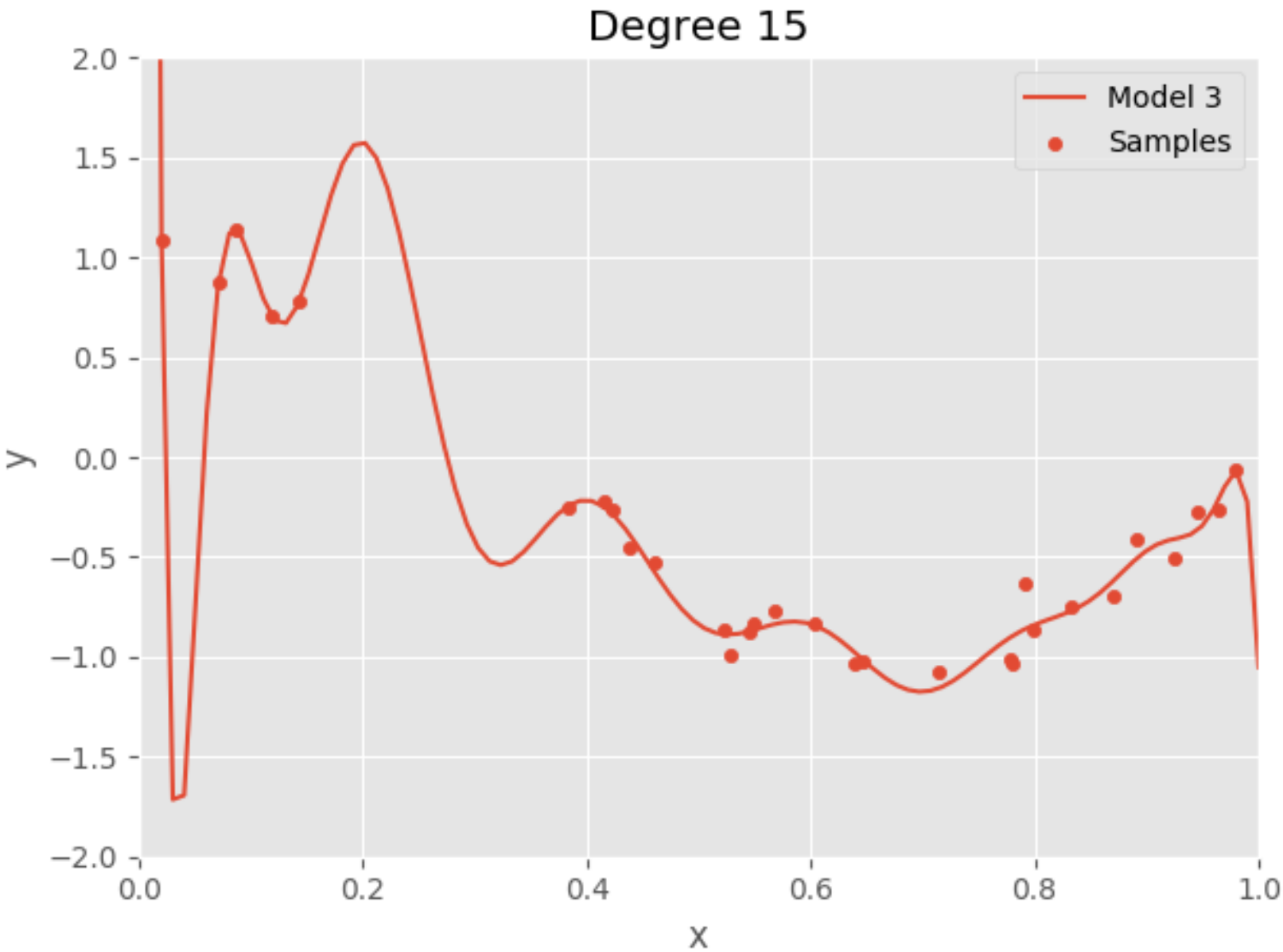
Practice

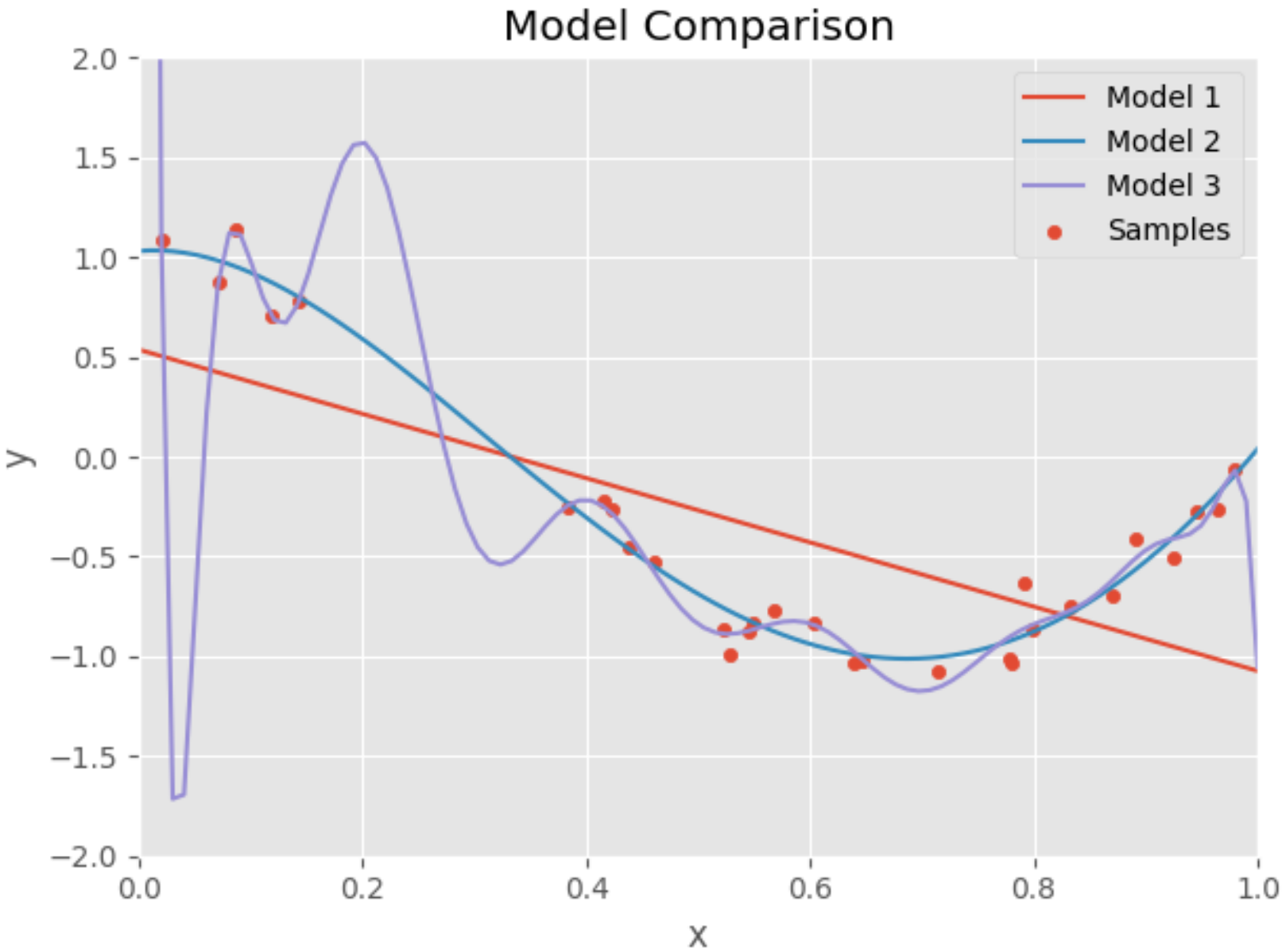

All Data
Training Data
Test Data
Learn
Test


Data
Samples
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x


Degree 1
Model 1
Samples
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x

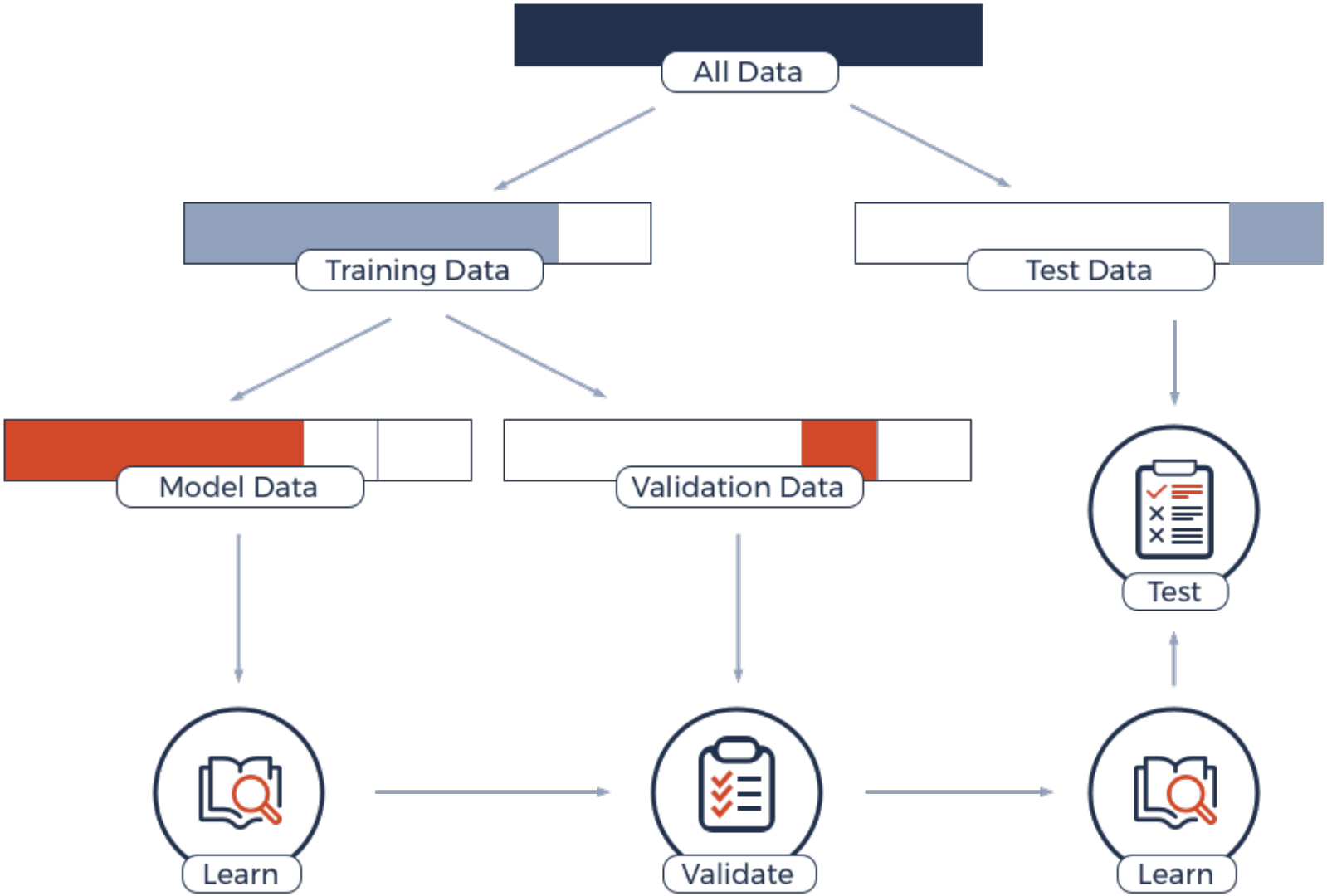

Degree 4
Model 2
Samples
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x

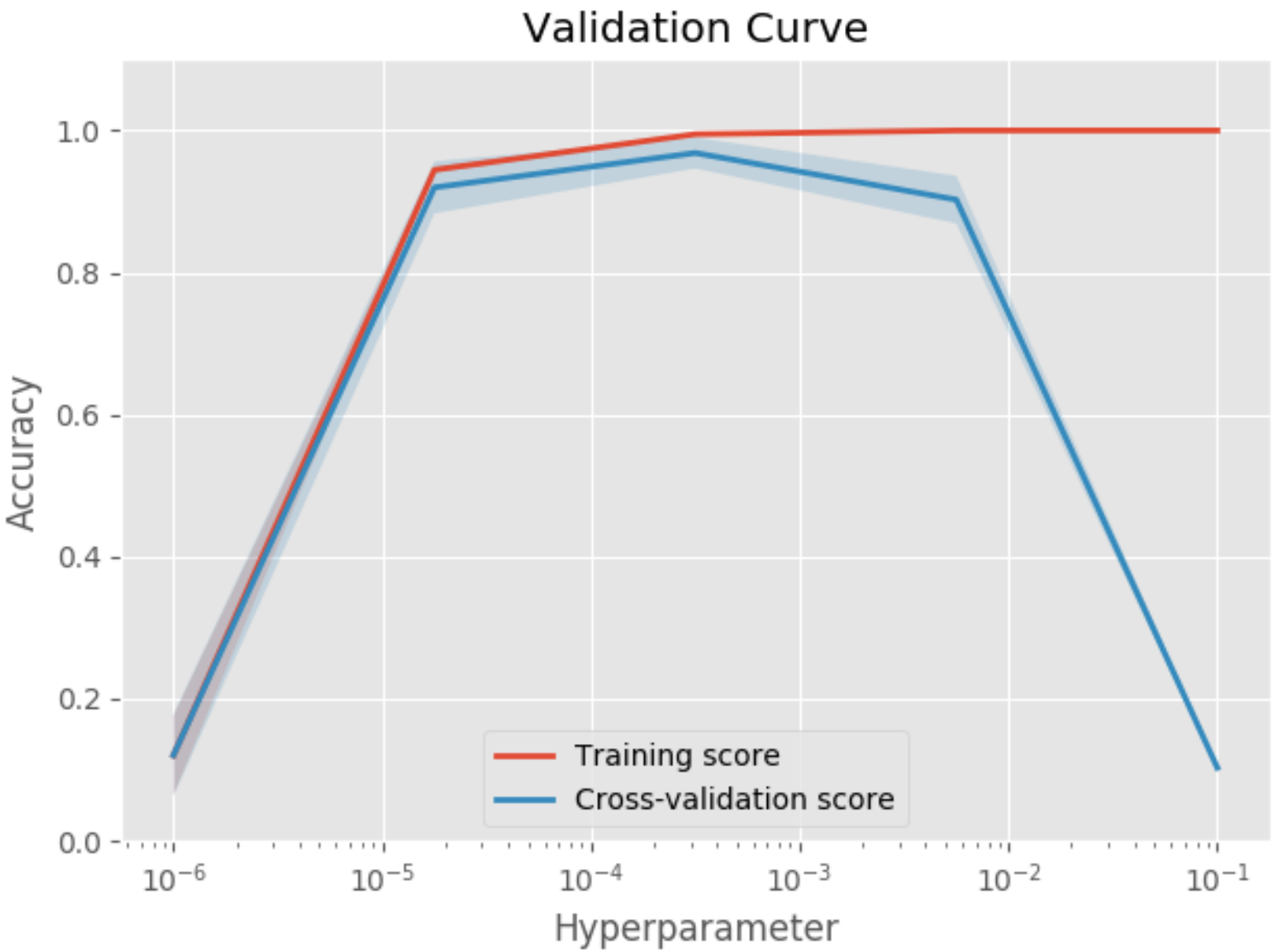

Degree 15
Model 3
Samples
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x

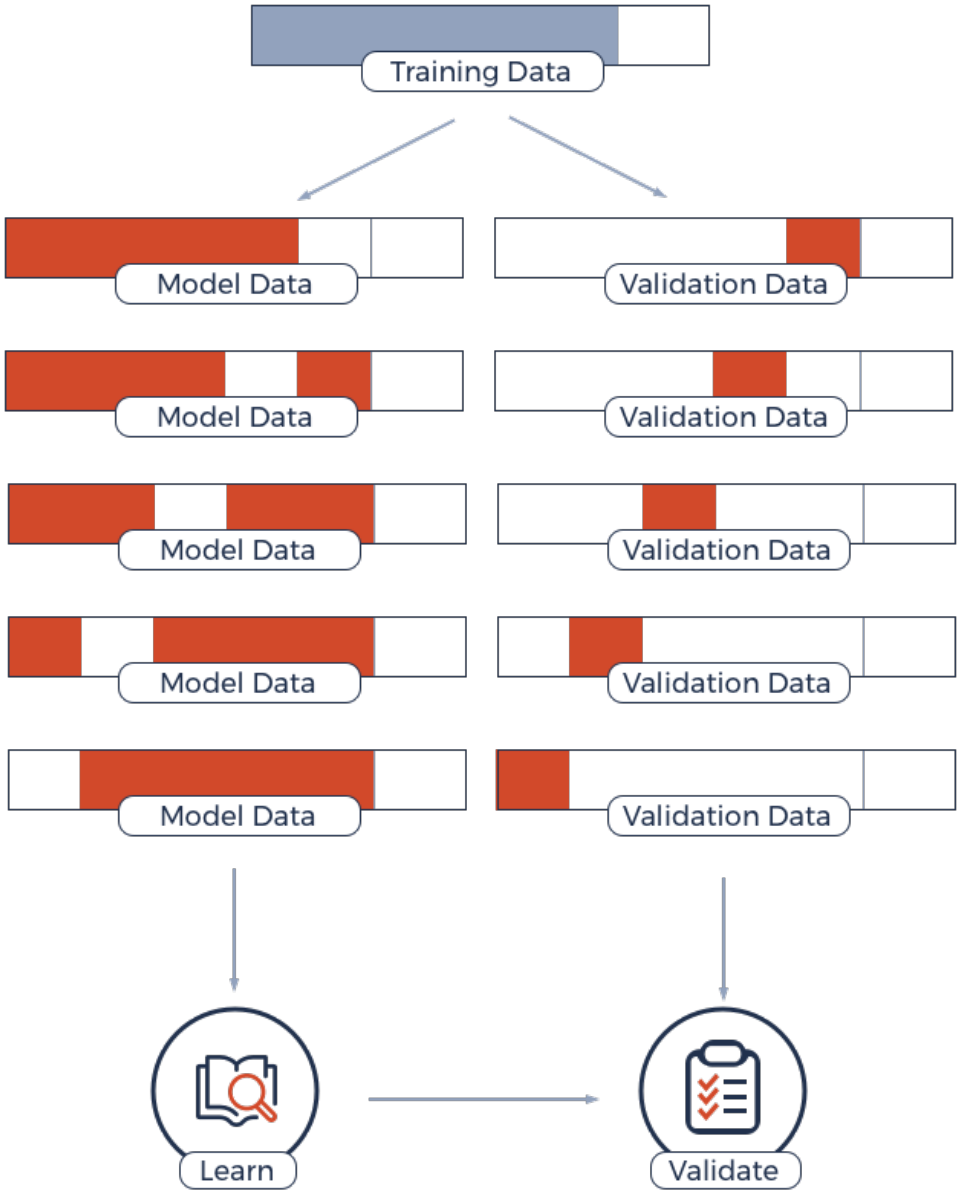

Model Comparison
Model 1
Model 2
Model 3
Samples
y
x
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0

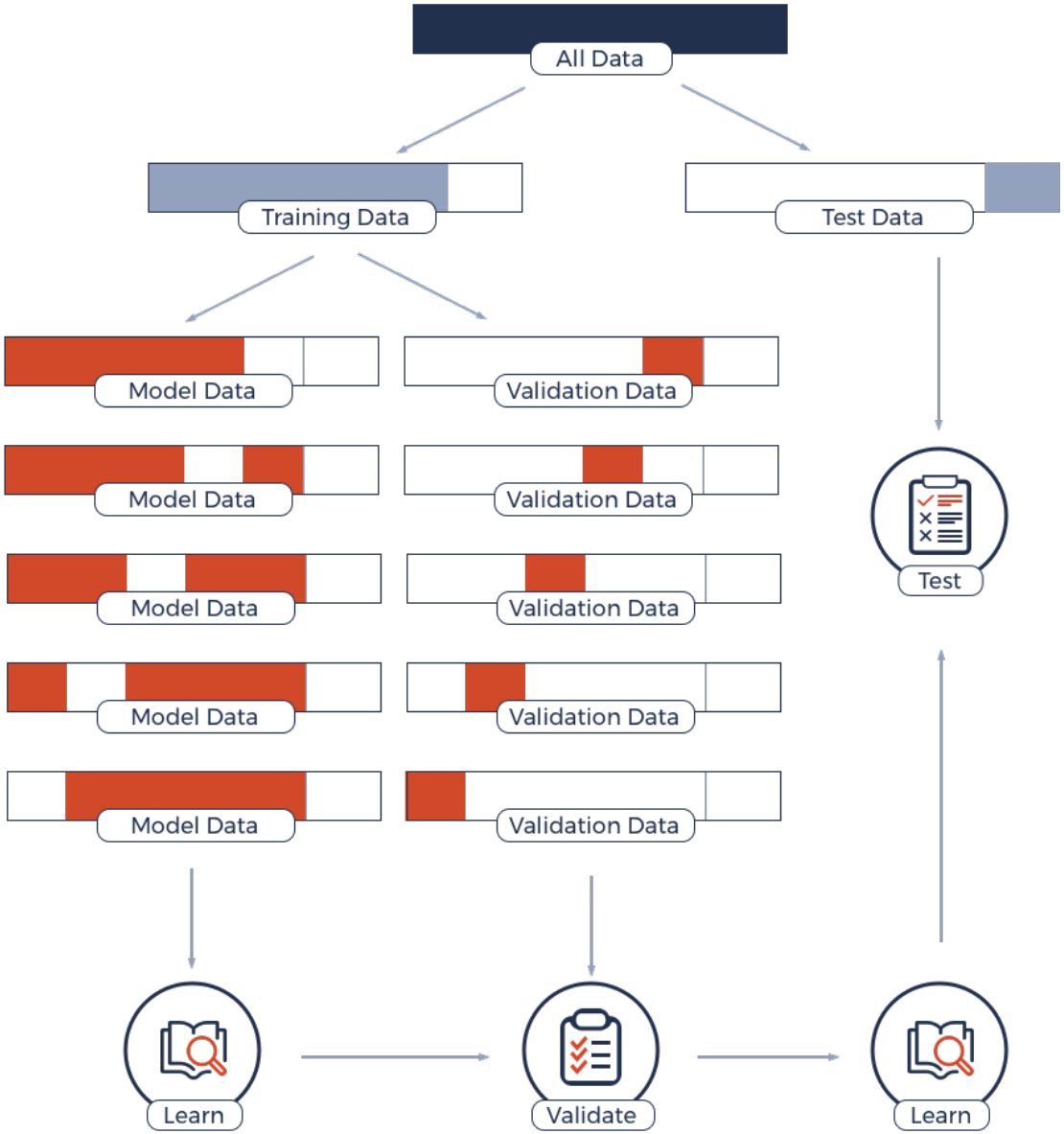
Model Fitting

- Underfitting
- Generalizing
- Overfitting


All Data
Training Data
Test Data
Model Data
Validation Data
Test
Learn
Validate
Learn


Validation Curve
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
Hyperparameter
Training score
Cross-validation score


Training Data
Model Data
Validation Data
Model Data
Validation Data
Model Data
Validation Data
Model Data
Validation Data
Model Data
Validation Data
Learn
Validate


All Data
Training Data
Test Data
Model Data
Validation Data
Model Data
Validation Data
Model Data
Validation Data
Model Data
Validation Data
Model Data
Validation Data
Test
Learn
Validate
Learn

Products

- Business
- Engineering

Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.