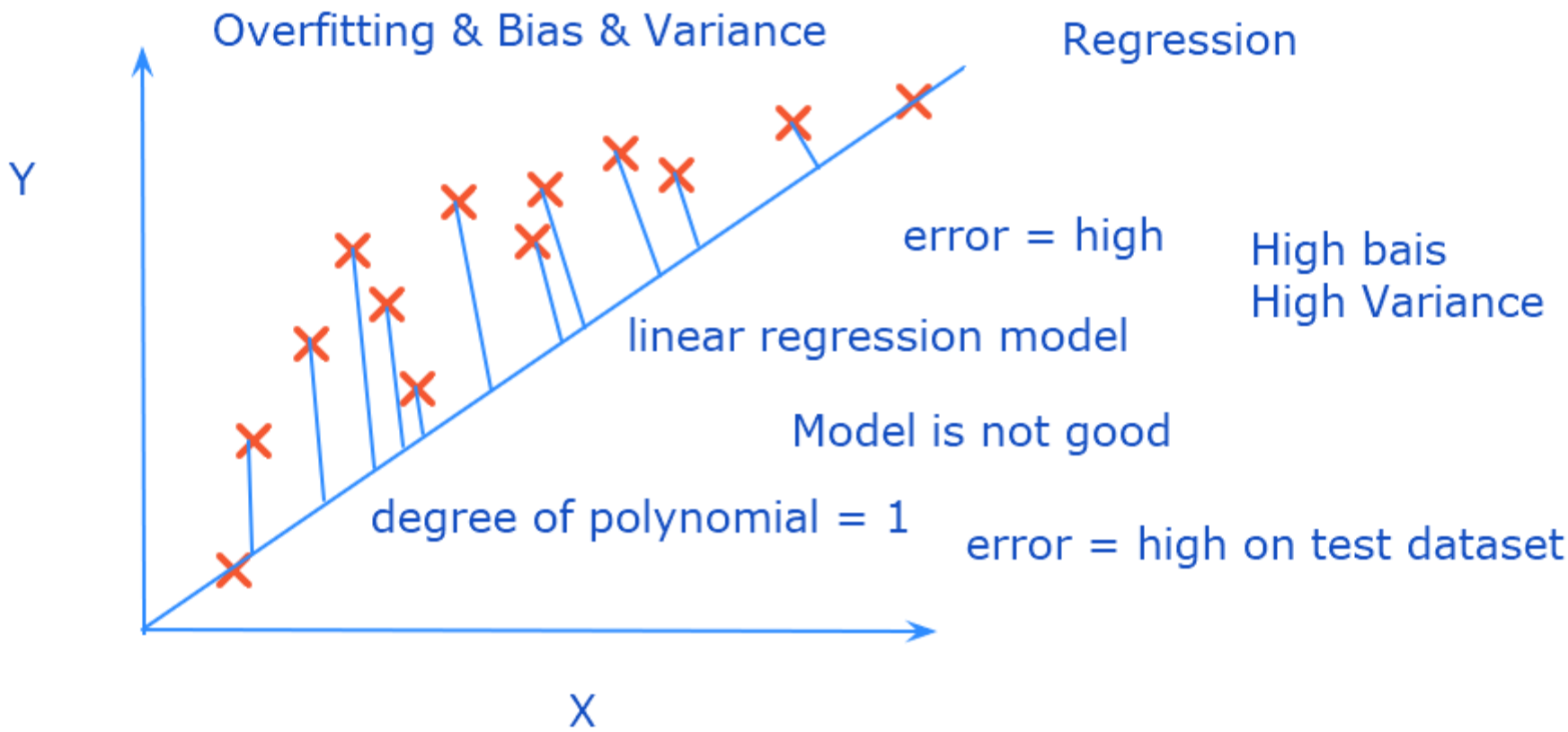
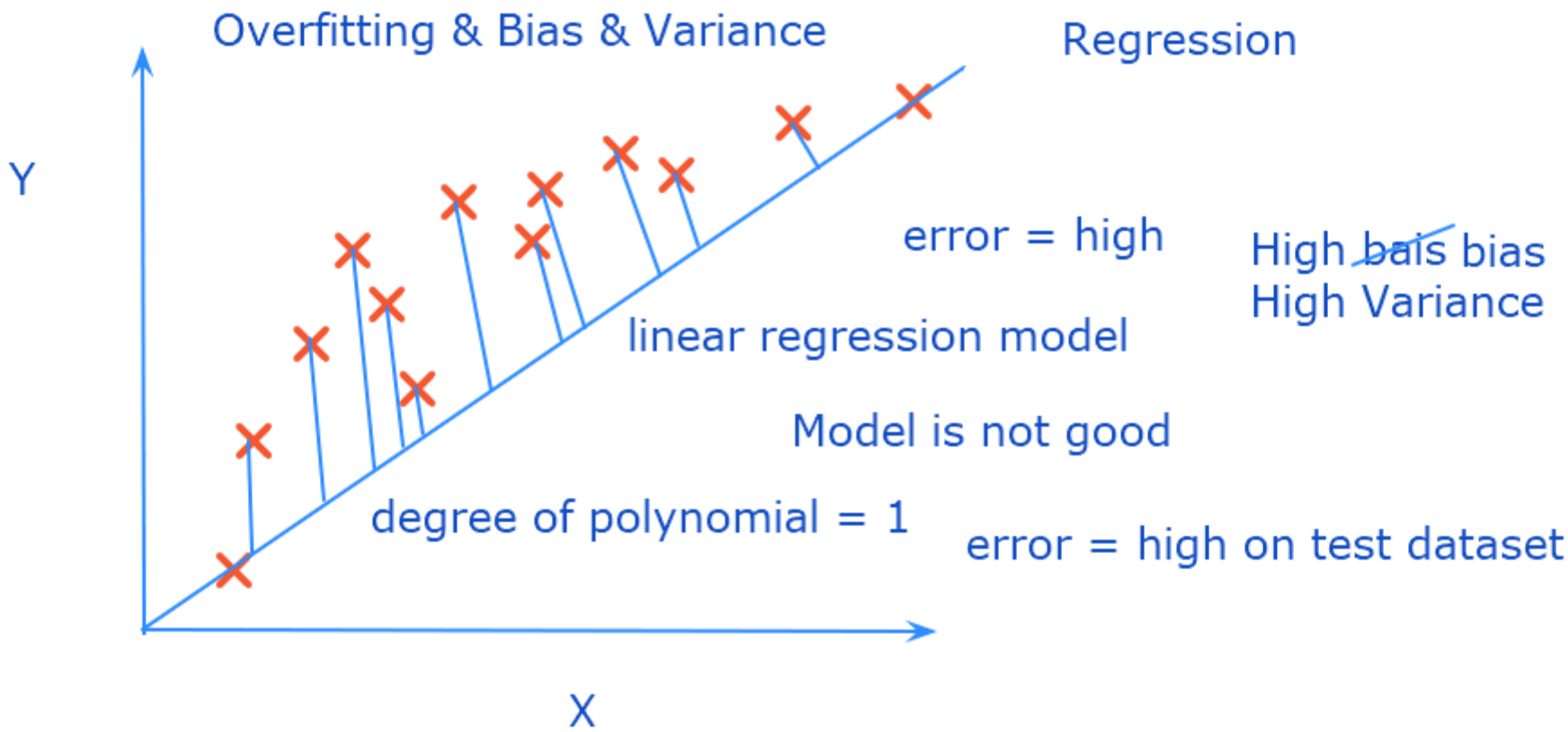
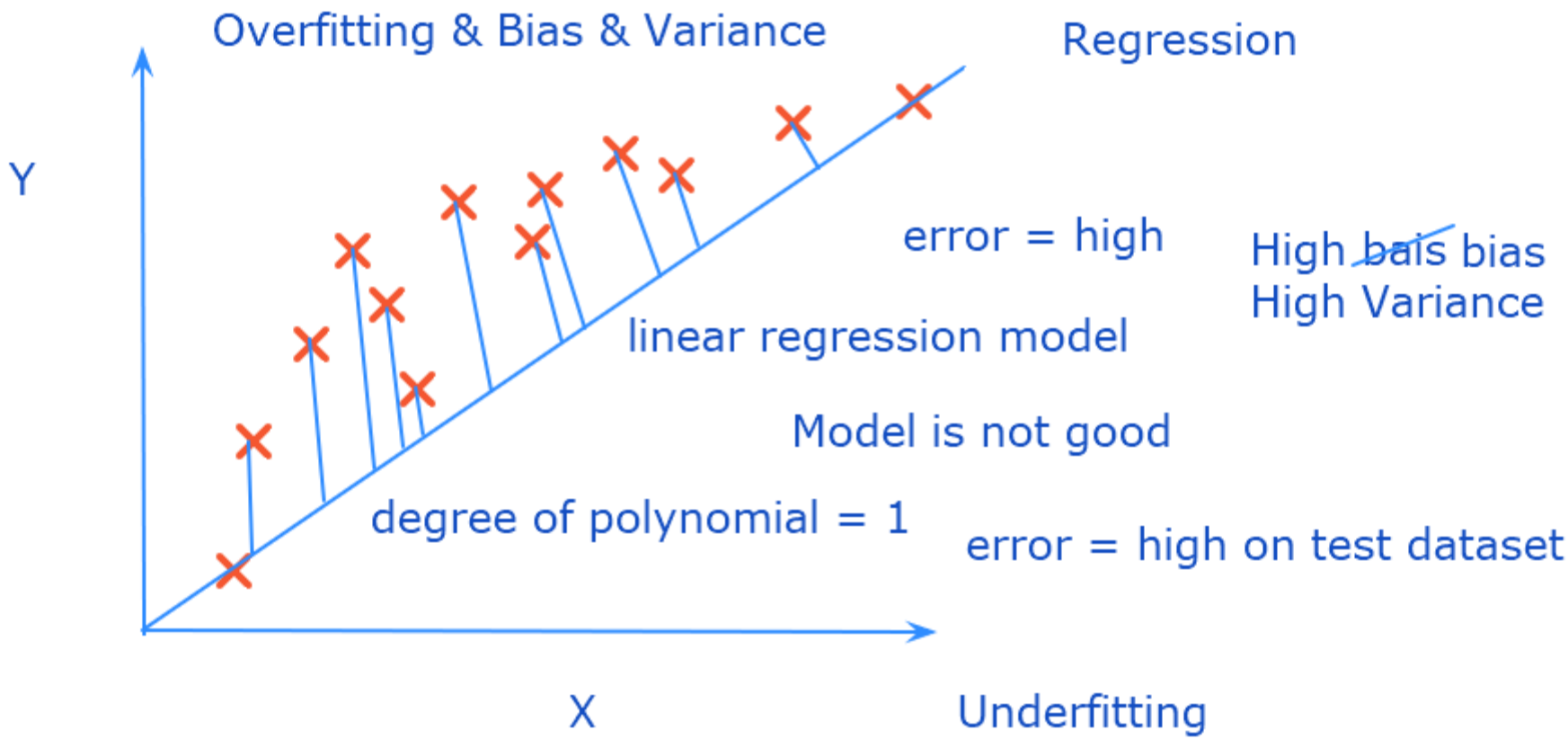
Overfitting & Regularization



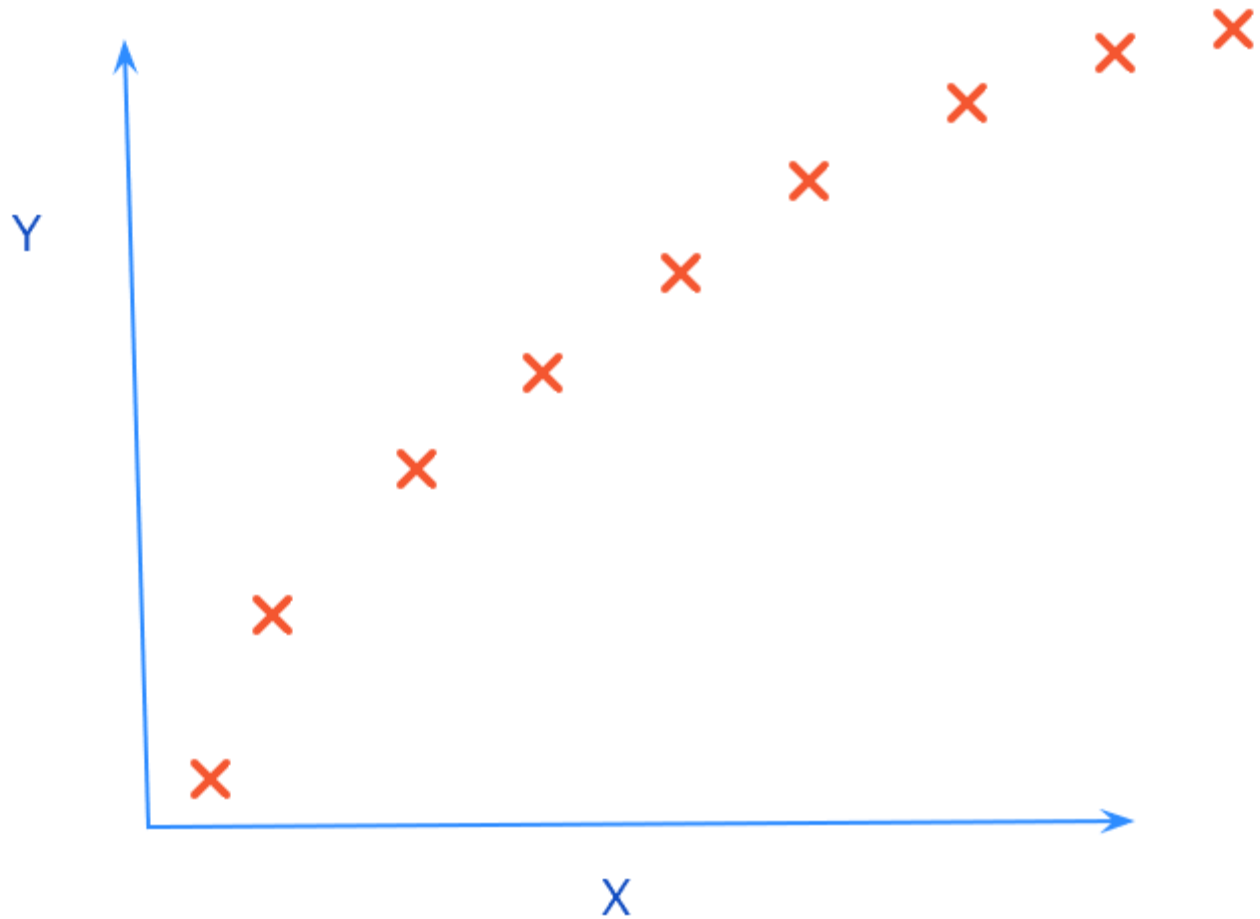
Bias = error in the training dataset
Variance = error in the test dataset

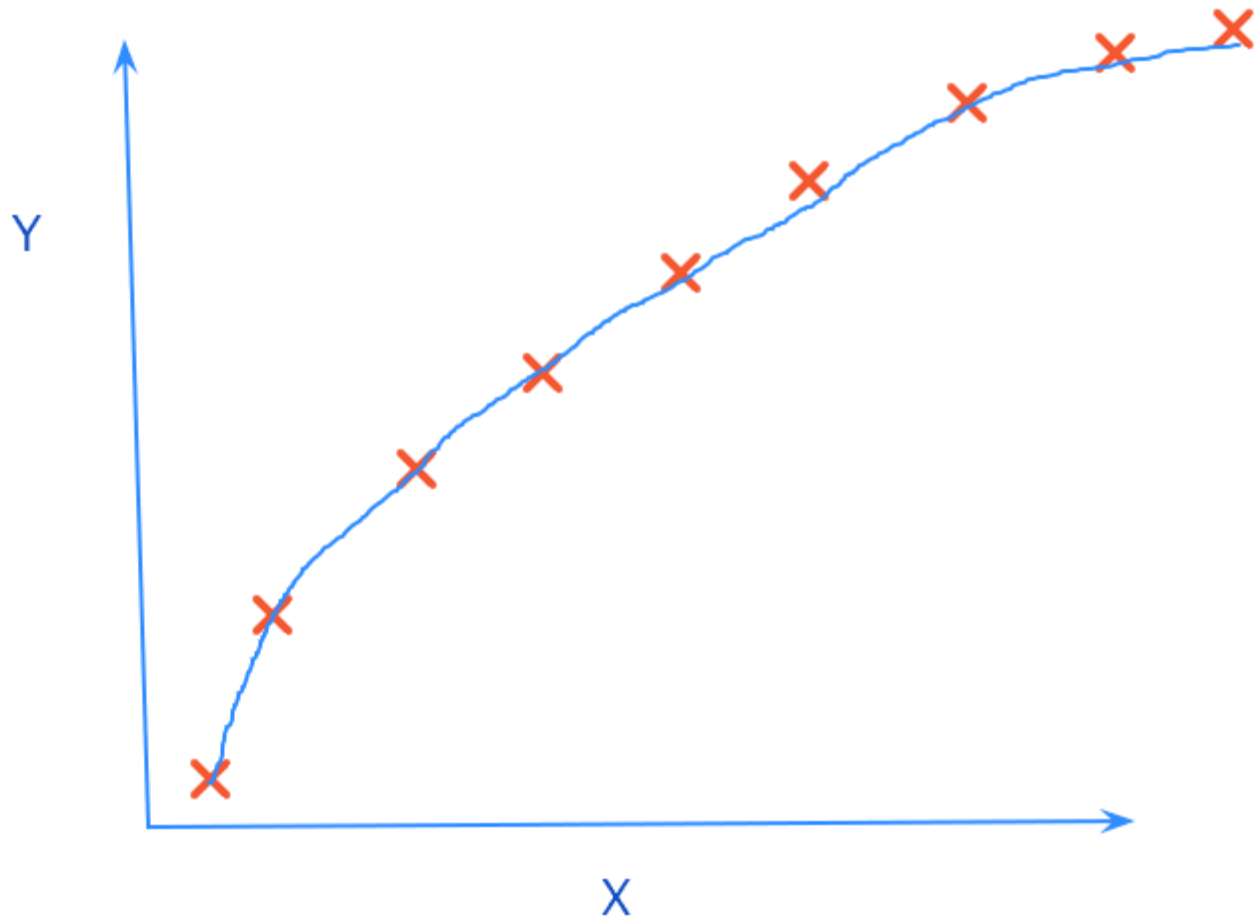


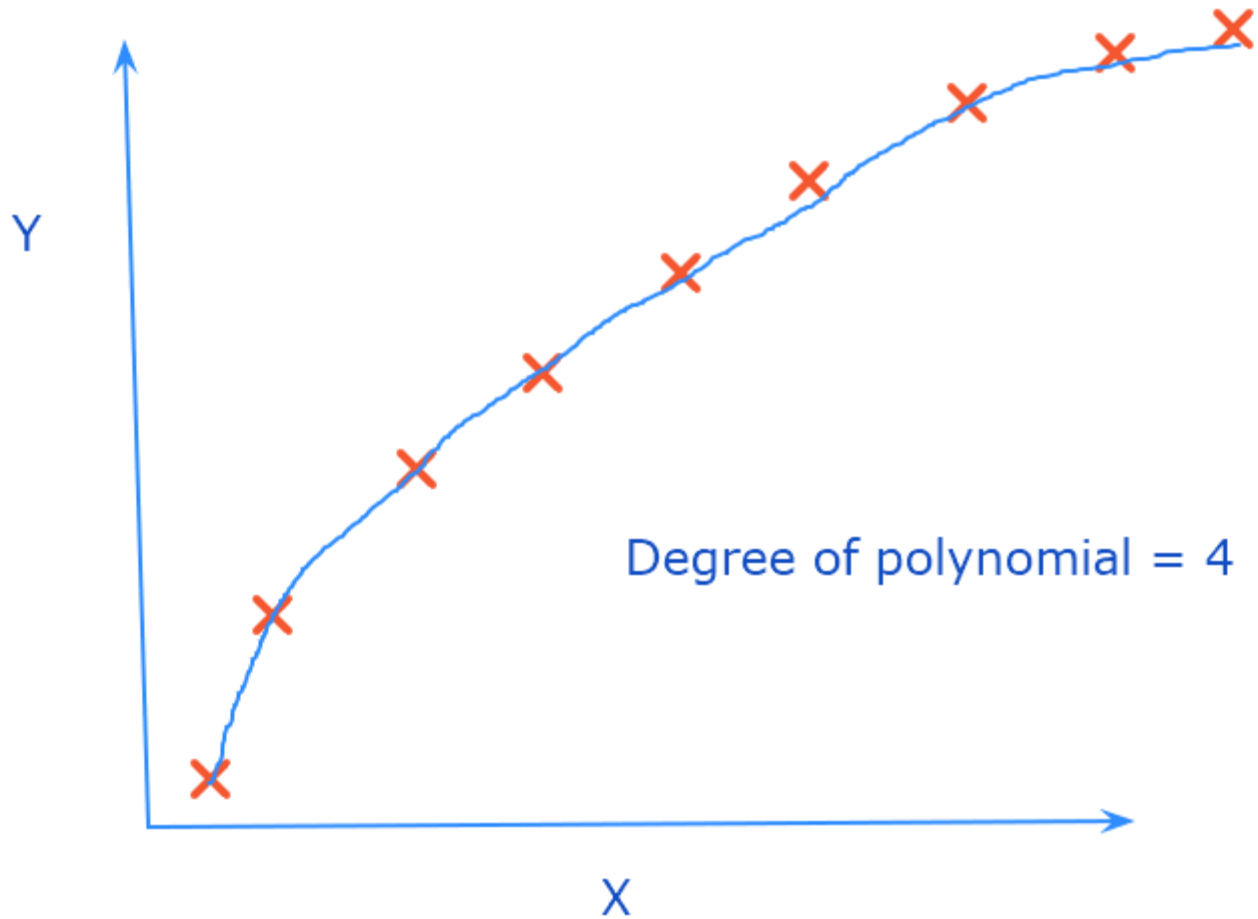
Bias = error in the training dataset
Variance = error in the test dataset

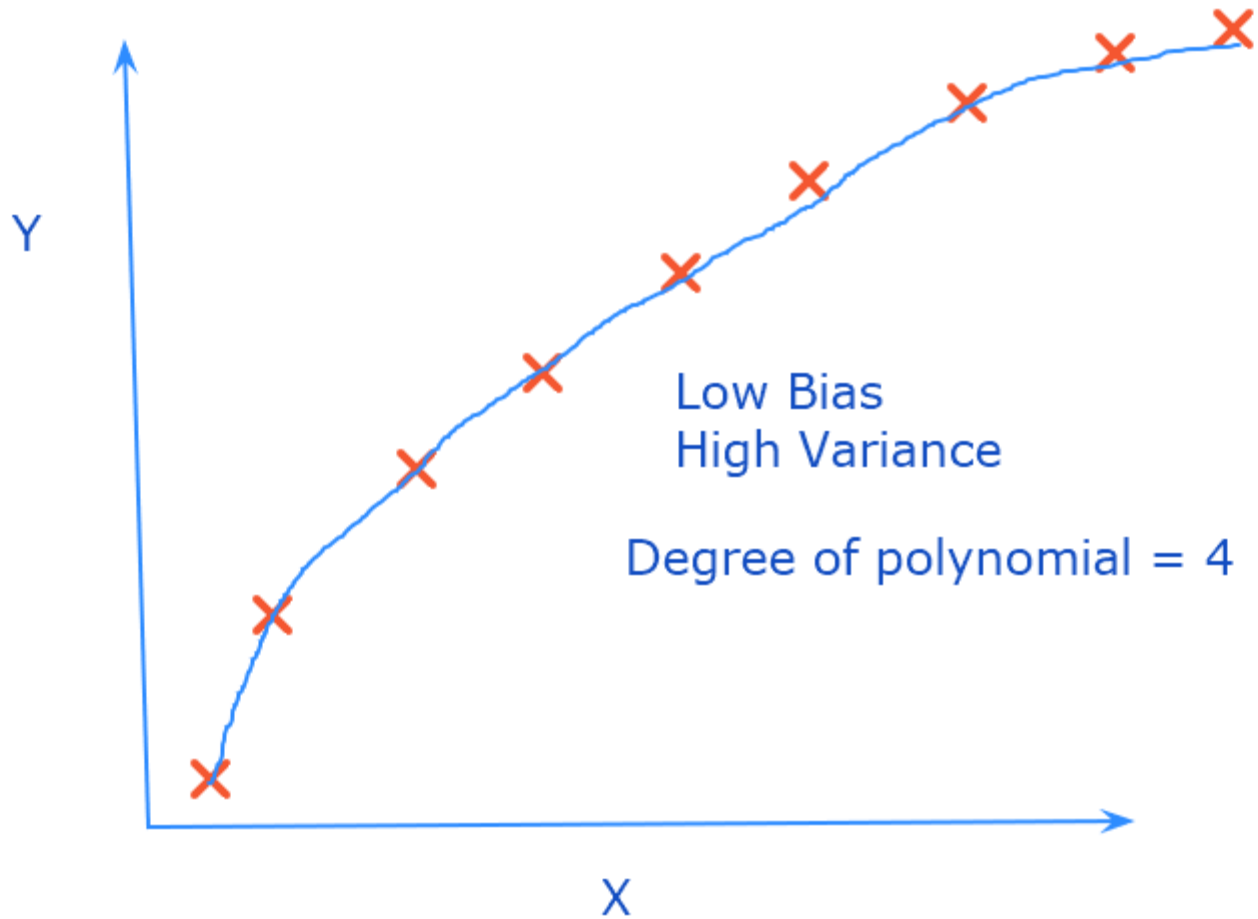


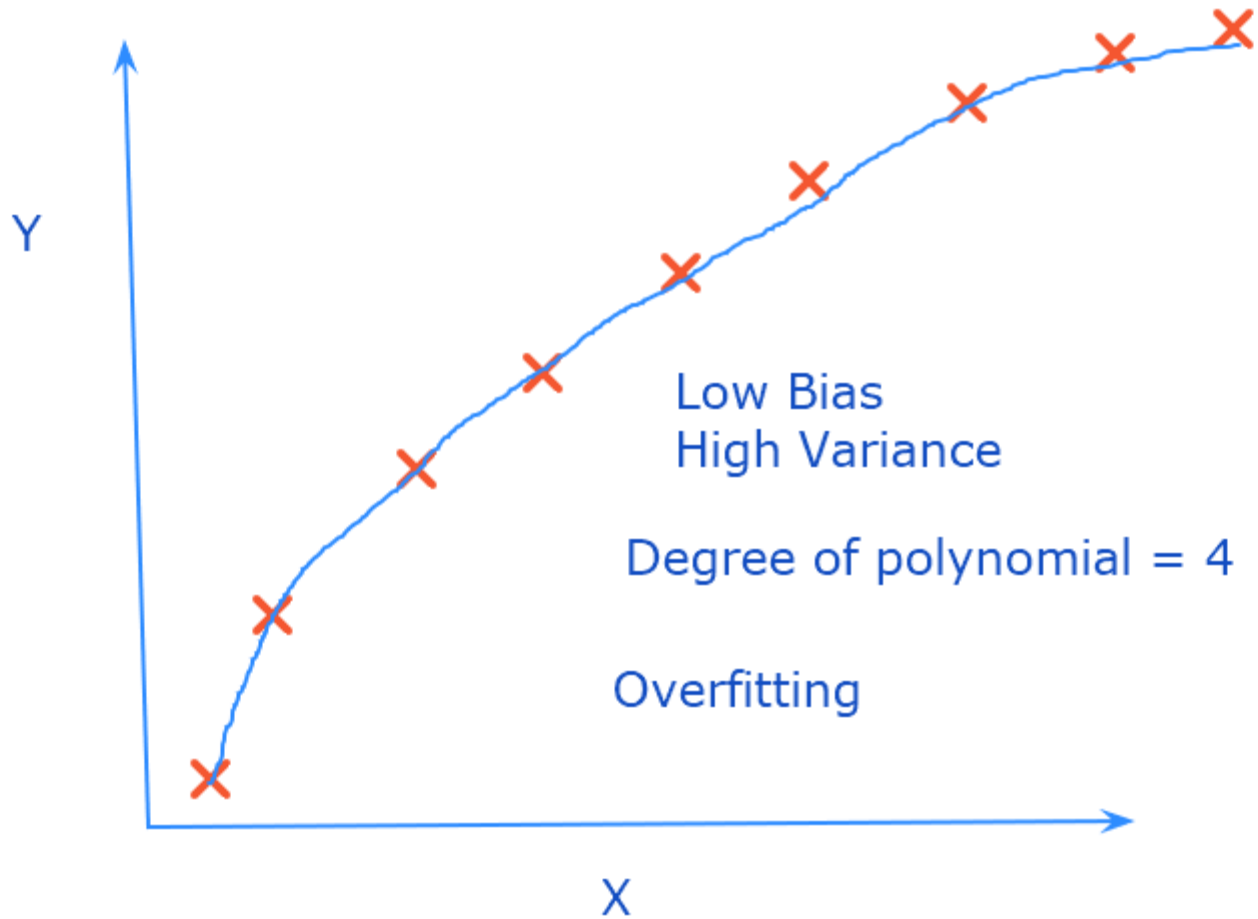
Bias = error in the training dataset
Variance = error in the test dataset

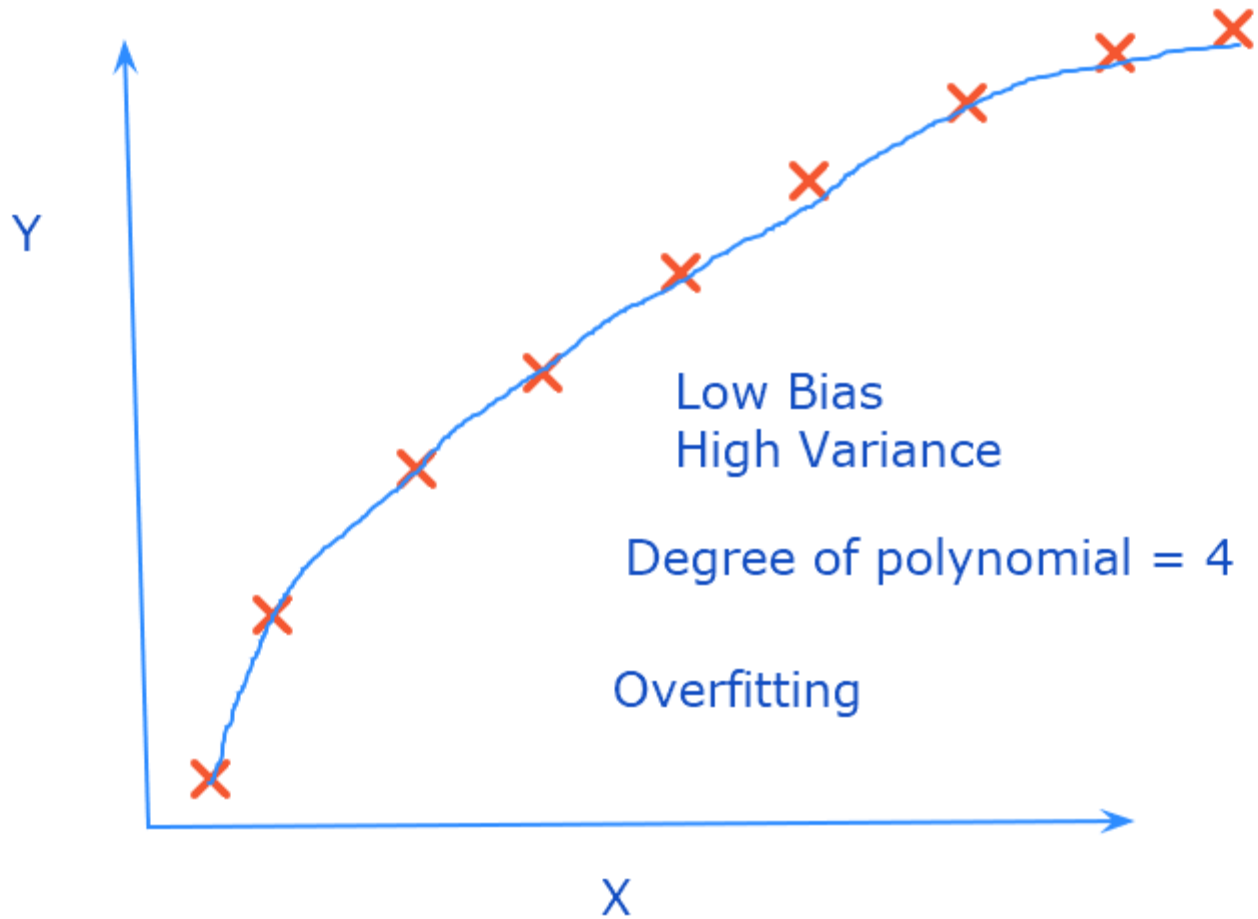

Y
X


Y
X

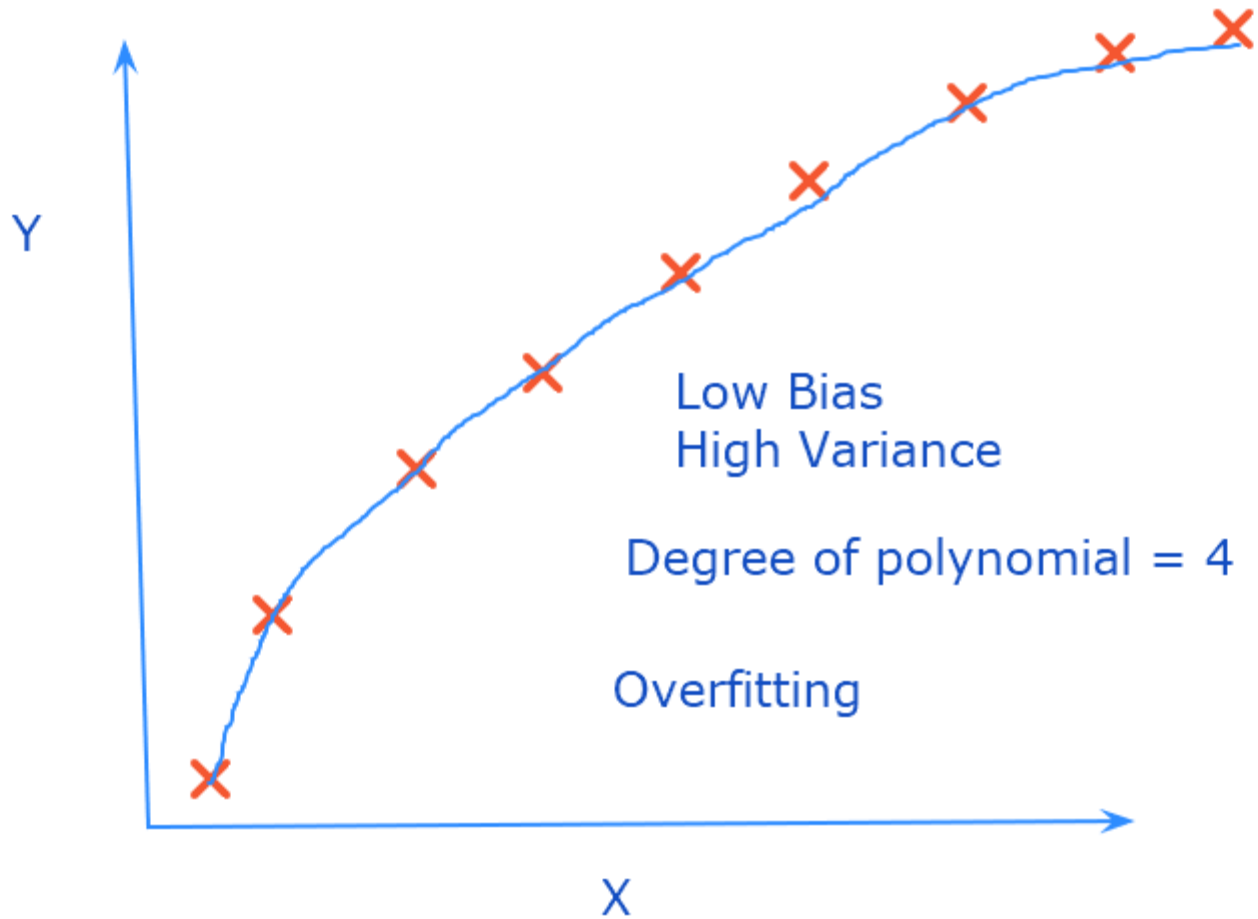

Y
X
Degree of polynomial = 4


Y
X
Low Bias
High Variance
Degree of polynomial = 4


Y
X
Low Bias
High Variance
Degree of polynomial = 4
Overfitting

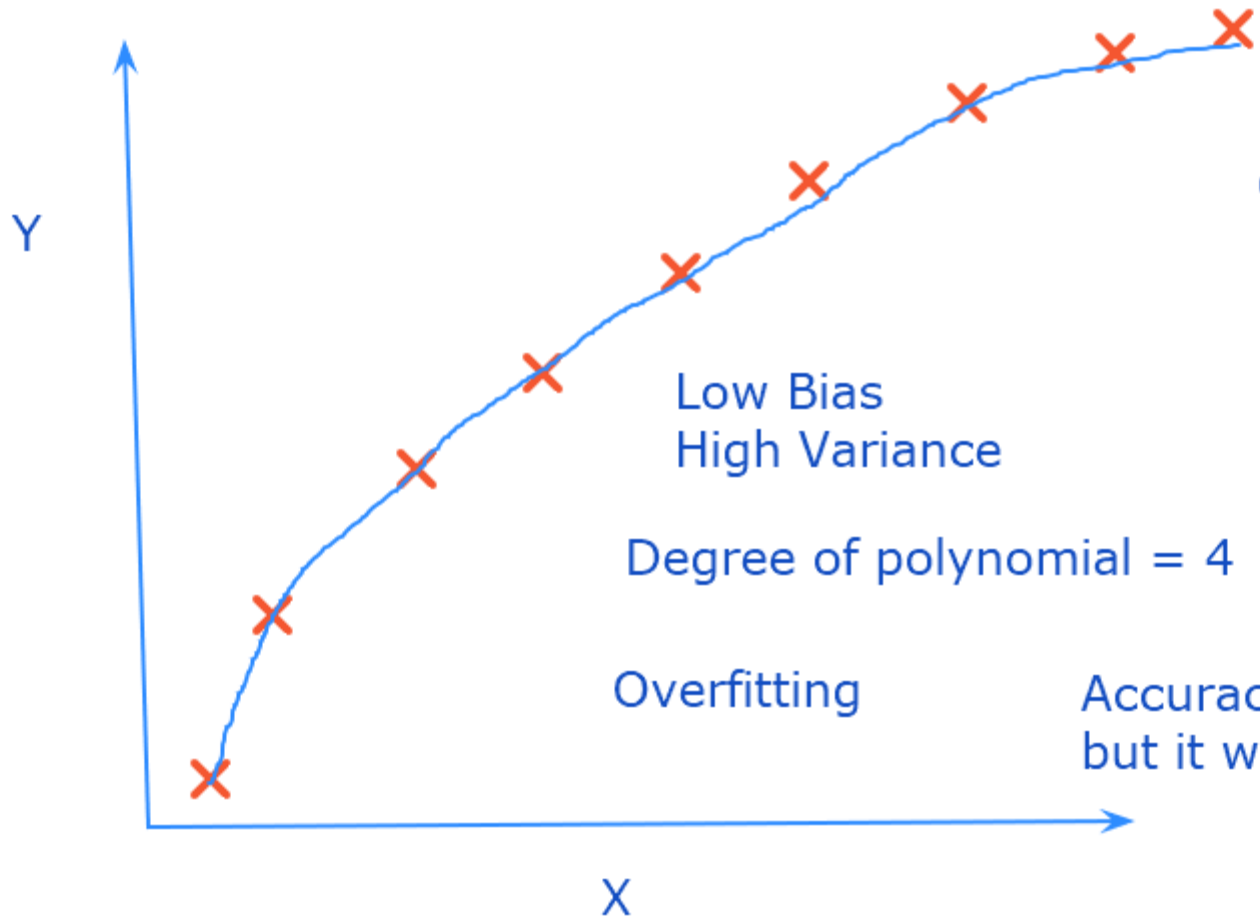


Goal: low bias & low variance



Goal: low bias & low variance

generalized model



Goal: low bias & low variance

generalized model

Low Bias
High Variance

Degree of polynomial = 4

Overfitting

Accuracy is high on training dataset
but it will be low on test dataset

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Model 3: training error <= 9%
Test error <= 9%

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Model 3: training error <= 9%
Test error <= 9%

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Underfitting case: High bias & High Variance

Model 3: training error <= 9%
Test error <= 9%

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Underfitting case: High bias & High Variance

Model 3: training error <= 9%
Test error <= 9%

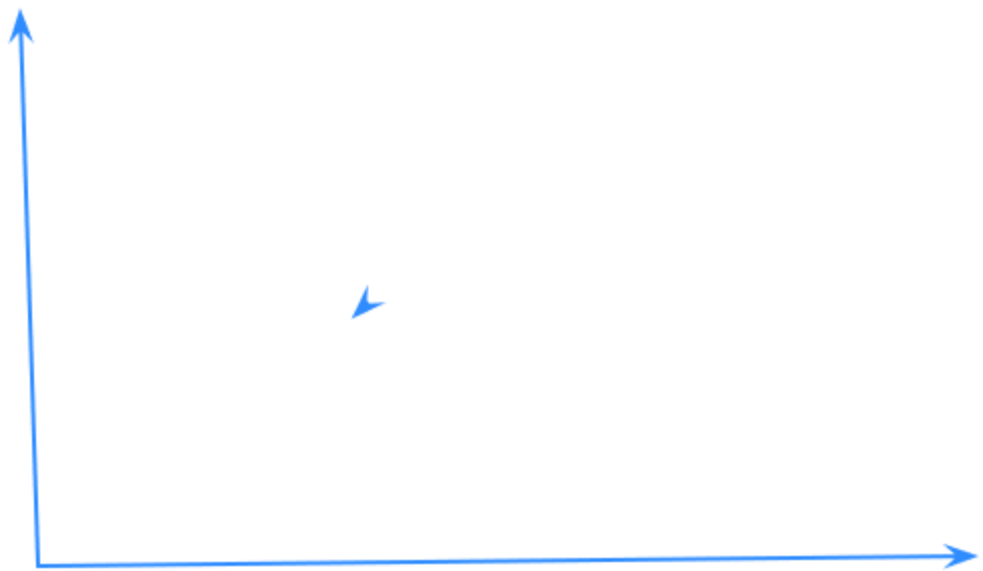
Generalized Model: Low Bias & Low Variance

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Underfitting case: High bias & High Variance



Model 3: training error <= 9%
Test error <= 9%

Generalized Model: Low Bias & Low Variance

Model 1: training error = 3%
Test error = 30%

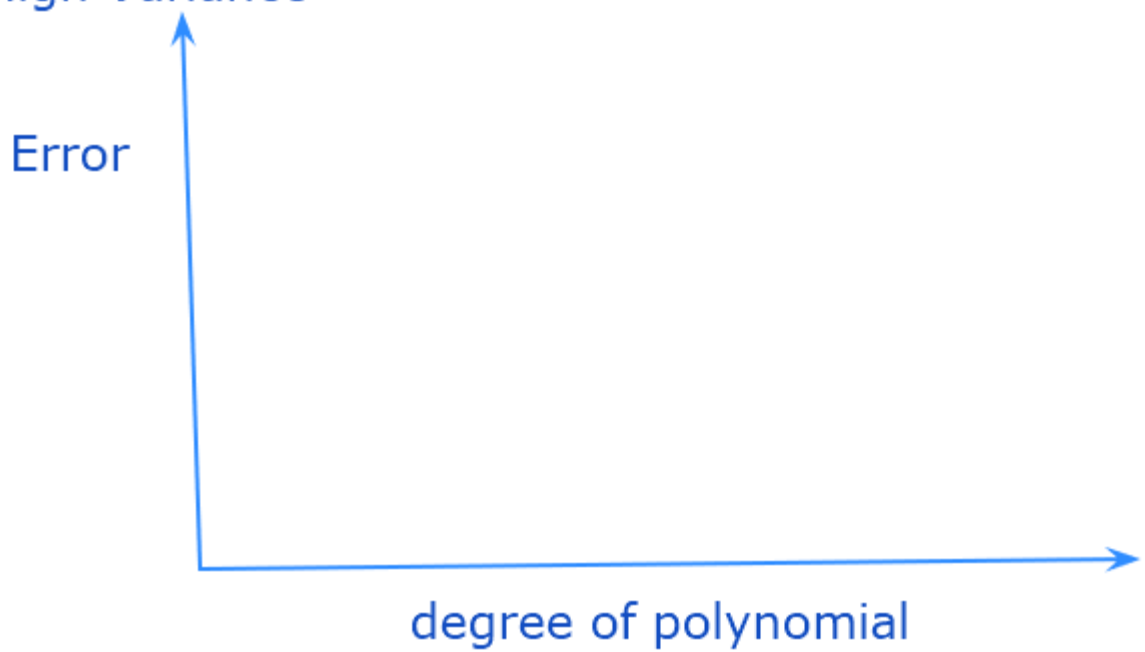
Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Underfitting case: High bias & High Variance

Model 3: training error <= 9%
Test error <= 9%

Generalized Model: Low Bias & Low Variance

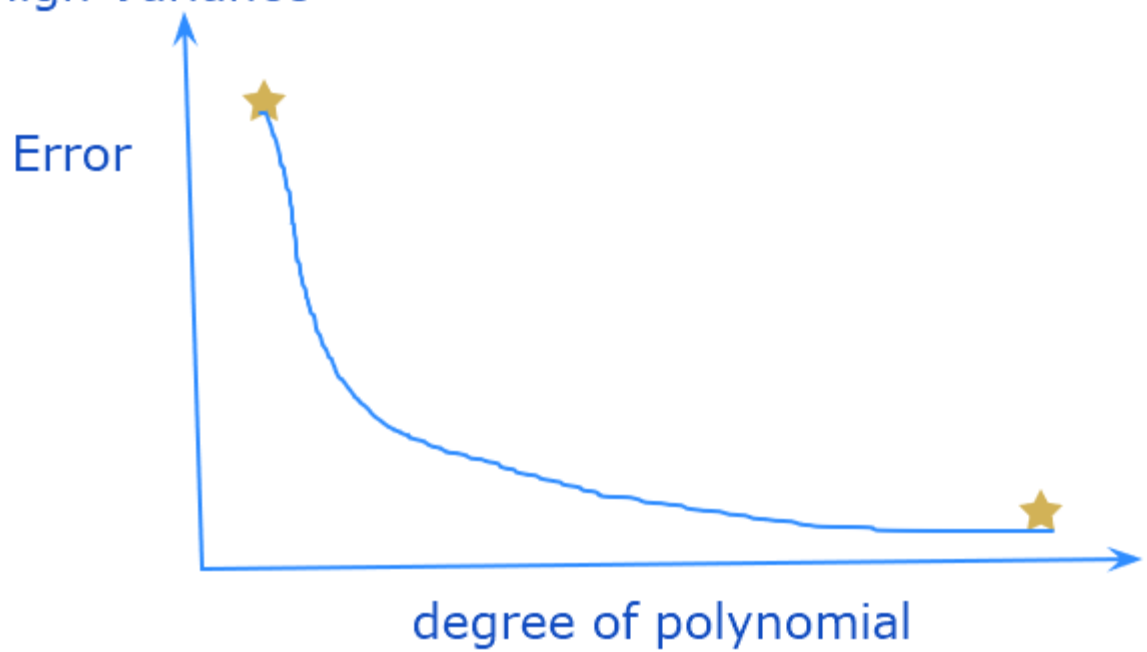


Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Underfitting case: High bias & High Variance



Model 3: training error <= 9%
Test error <= 9%

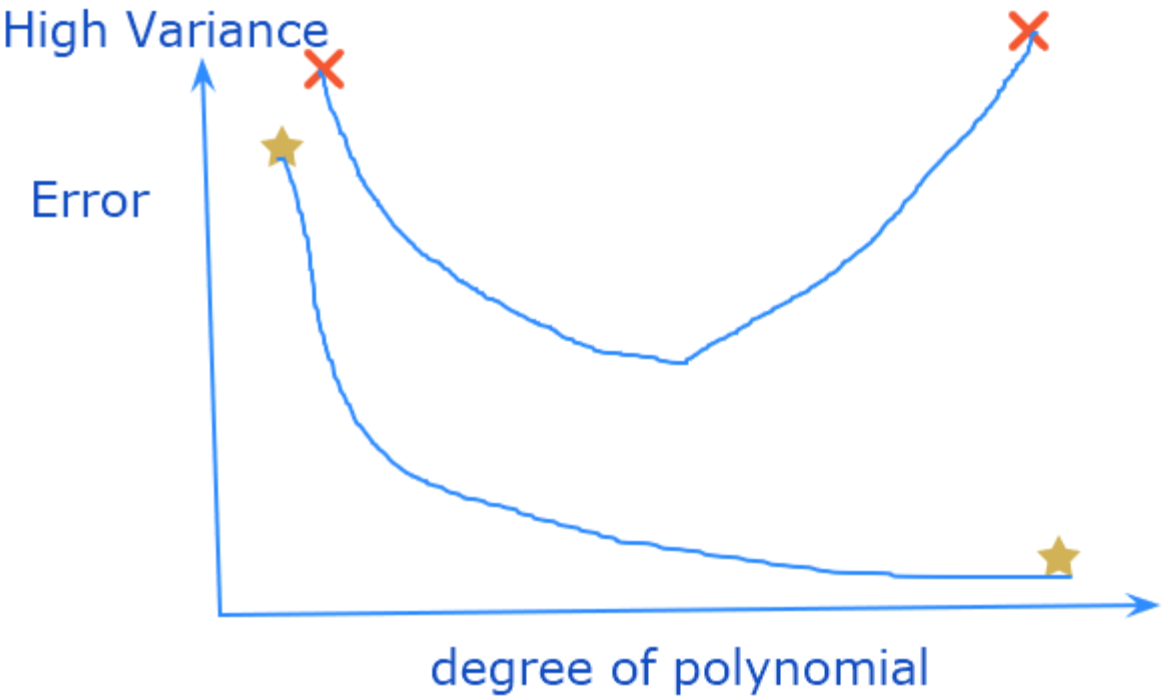
Generalized Model: Low Bias & Low Variance

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

Underfitting case: High bias & High Variance



Model 3: training error <= 9%
Test error <= 9%

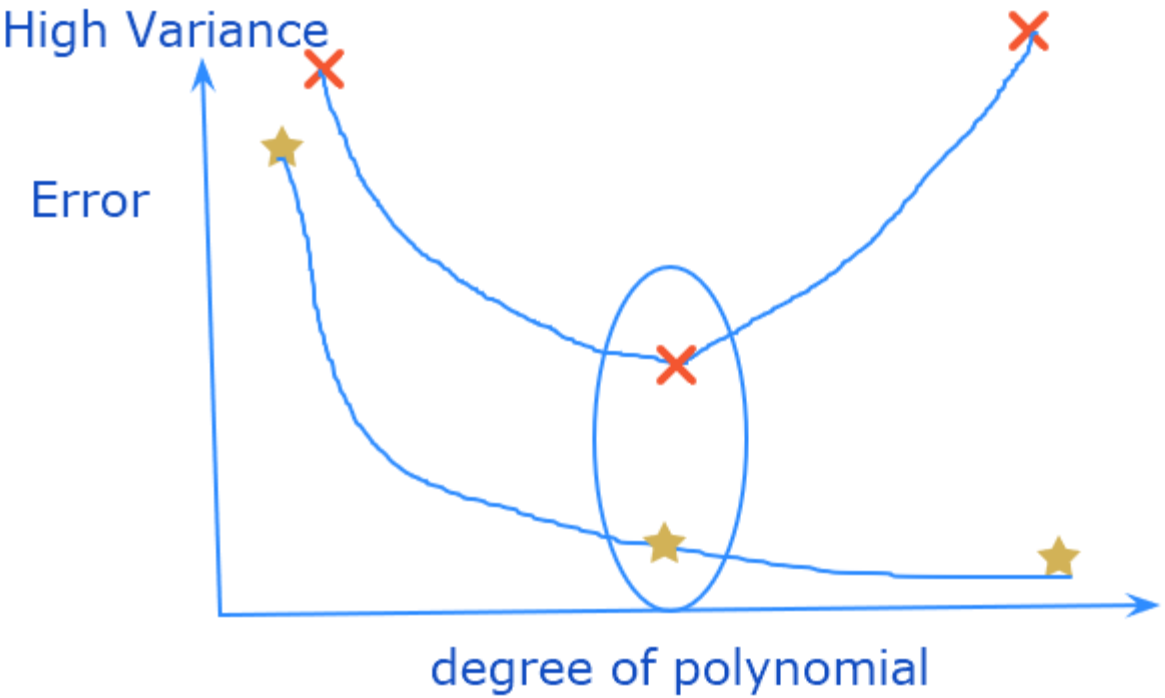
Generalized Model: Low Bias & Low Variance

Model 1: training error = 3%
Test error = 30%

Model 2: training error = 25%
Test error = 26%

Overfitting case: Low bias & high variance

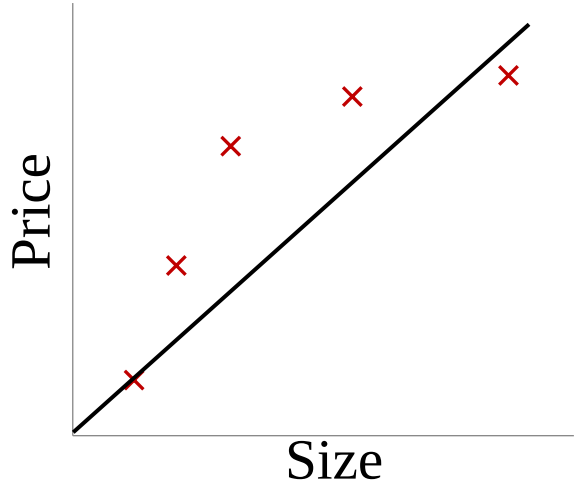
Underfitting case: High bias & High Variance



Model 3: training error <= 9%
Test error <= 9%

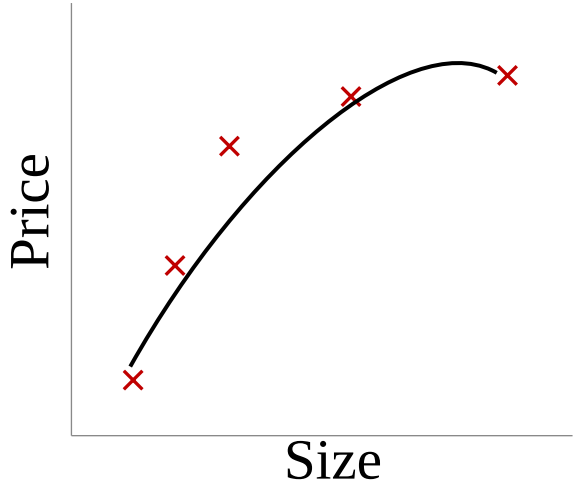
Generalized Model: Low Bias & Low Variance

Example: Linear regression (housing prices)



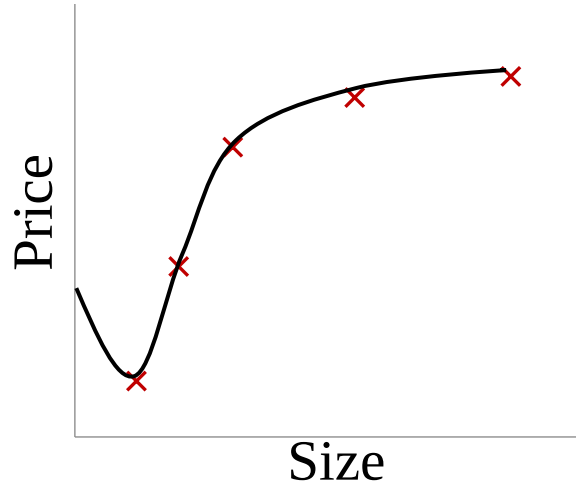
$\theta_0 + \theta_1 x$

Underfitting: “High Bias”



$\theta_0 + \theta_1 x + \theta_2 x^2$

Correct Fit



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Overfitting: “High Variance”

Overfitting: If we have too many features, the learned hypothesis may fit the training set very well ($J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \approx 0$), but fail to generalize to new examples (predict prices on new examples).