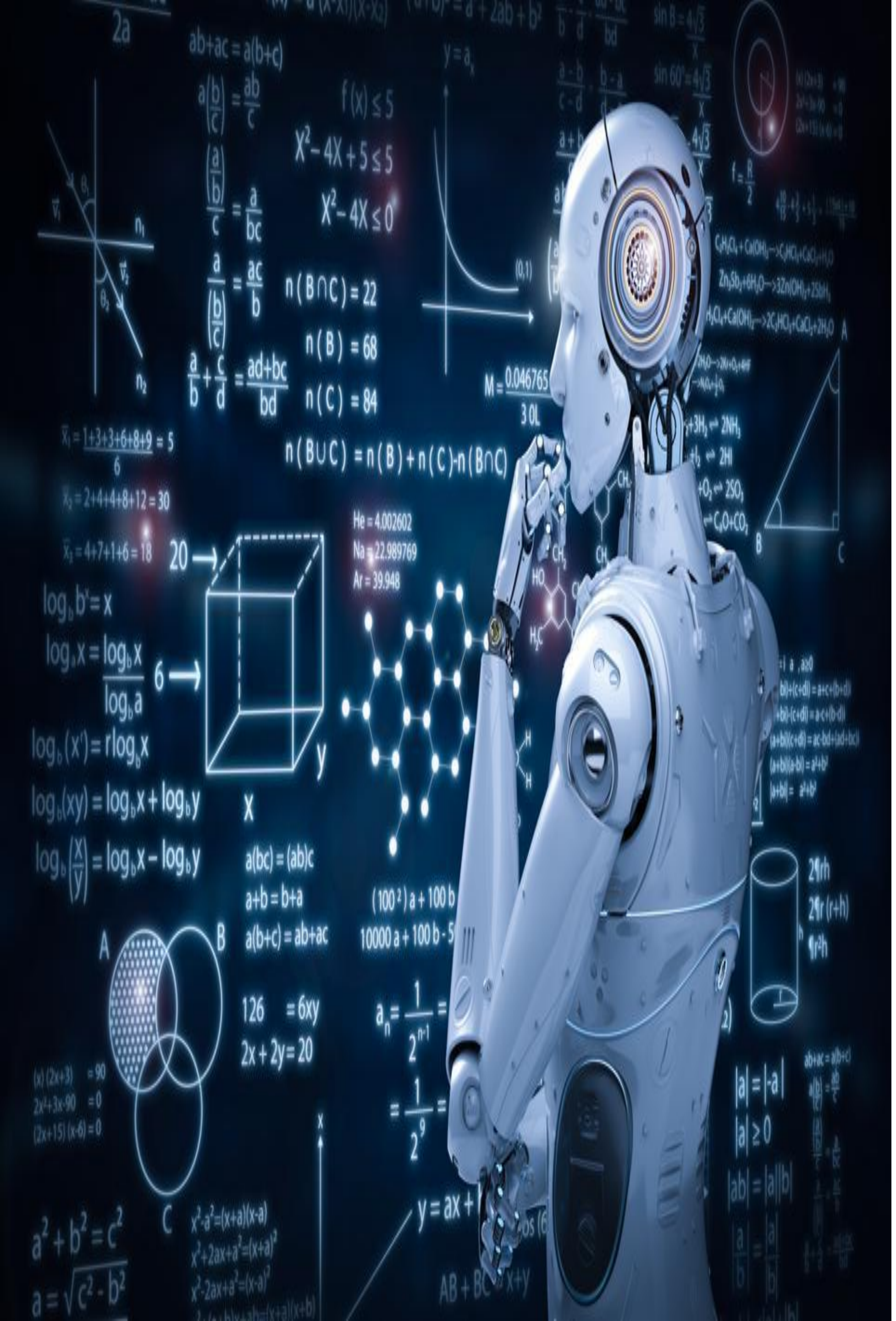


# Practical Machine Learning

## Day 6: Mar24 DBDA

Kiran Waghmare

# Agenda

- Regression
- Types of Regression

# Iris dataset

- Many exploratory data techniques are nicely illustrated with the iris dataset.
  - Dataset created by famous statistician Ronald Fisher
  - 150 samples of three species in genus *Iris* (50 each)
    - *Iris setosa*
    - *Iris versicolor*
    - *Iris virginica*
  - Four attributes
    - sepal width
    - sepal length
    - petal width
    - petal length
  - Species is class label

*Iris virginica*. Robert H. Mohlenbrock. USDA NRCS. 1995. Northeast wetland flora: Field office guide to plant species. Northeast National Technical Center, Chester, PA. Courtesy of USDA NRCS Wetland Science Institute.

# Types of Linear Regression

- Linear regression can be further divided into two types of the algorithm:
- **Simple Linear Regression:**

  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.
- **Multiple Linear regression:**

  If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

# What is multiple linear regression (MLR)? Visual model

## Linear Regression

Single predictor X → Y

## Multiple Linear Regression

Multiple predictors

$X_1$
$X_2$
$X_3$
$X_4$
$X_5$

Y

Simple Linear Regression

$$y = b_0 + b_1 x_1$$

Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

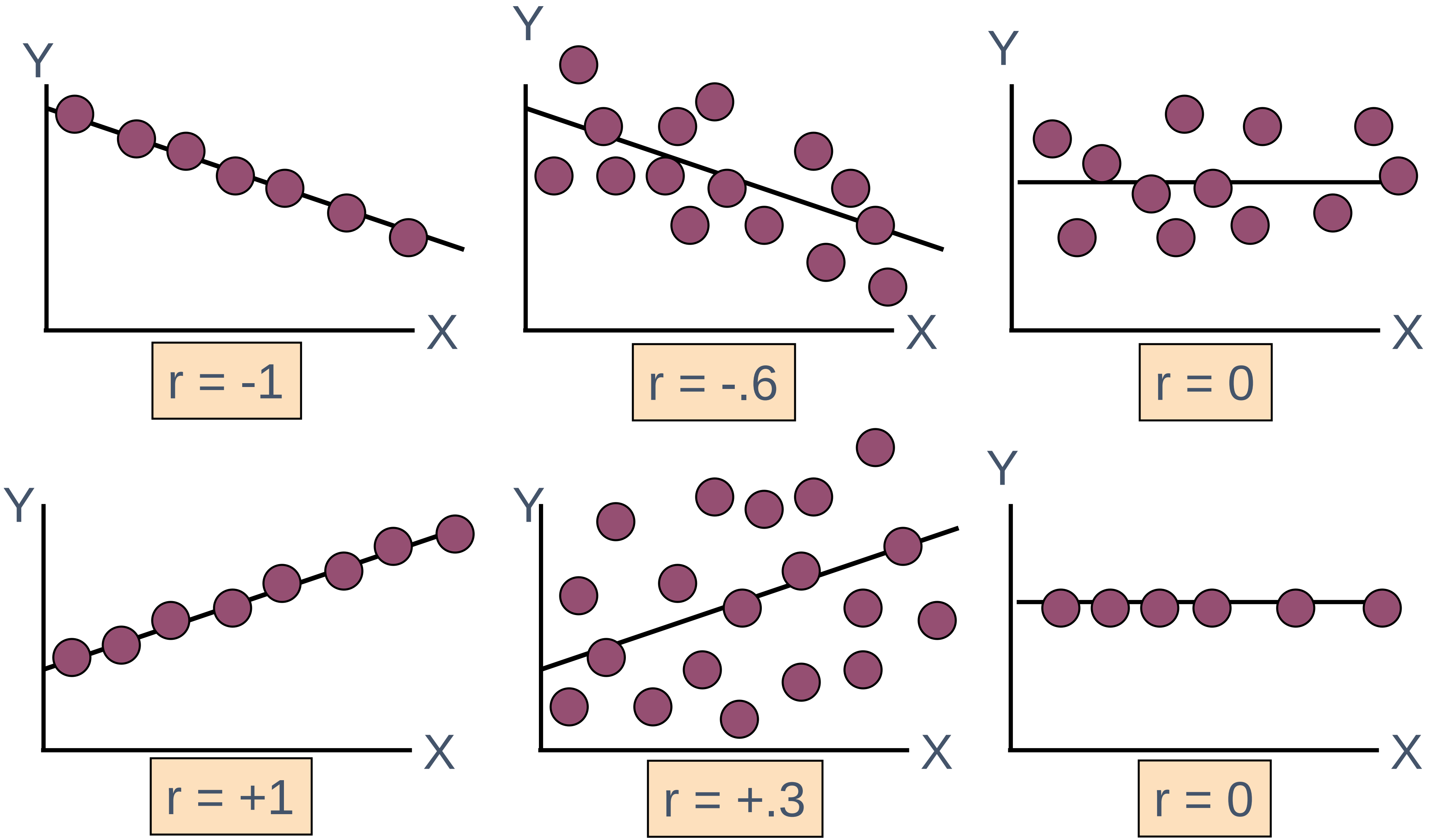
Polynomial Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \dots + b_n x_1^n$$

# Correlation

- Measures the relative strength of the *linear* relationship between two variables
- **Unit-less**
- Ranges between **–1 and 1**
- The closer to –1, the stronger the **negative linear** relationship
- The closer to 1, the stronger the **positive linear** relationship
- The closer to 0, the **weaker** any positive linear relationship

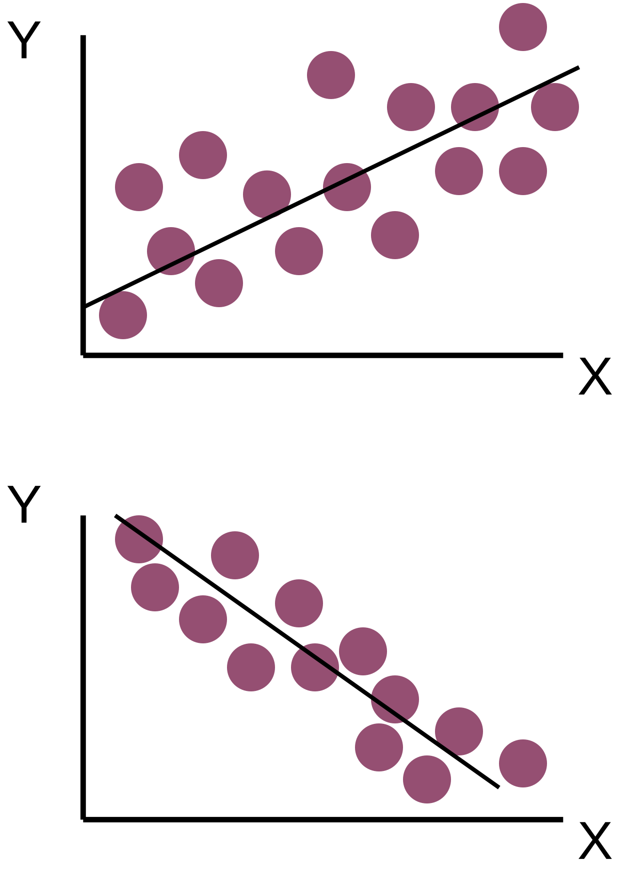
# Scatter Plots of Data with Various Correlation Coefficients



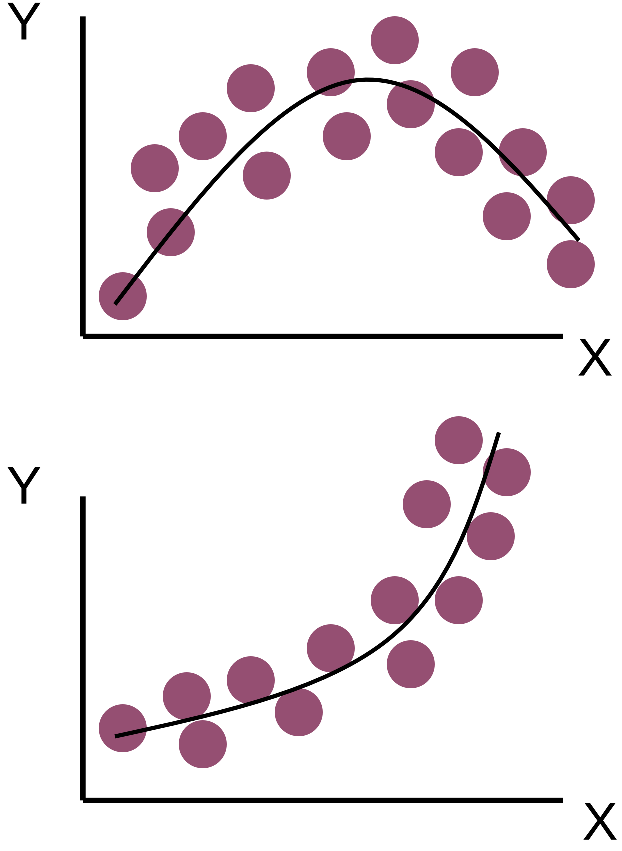

Slide from: Statistics for Managers Using Microsoft® Excel 4th Edition, 2004 Prentice-Hall

# Linear Correlation

Linear relationships

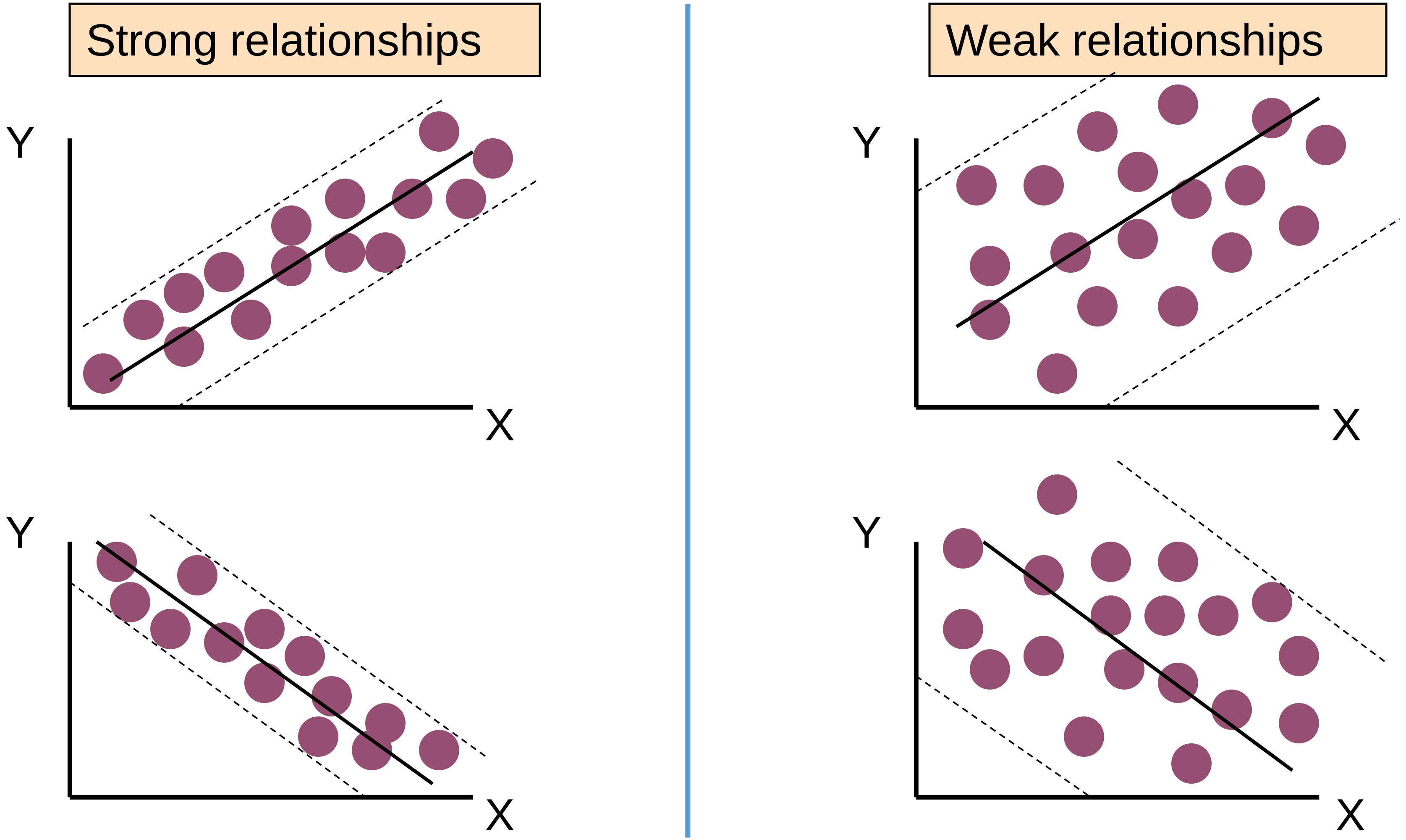



Curvilinear relationships



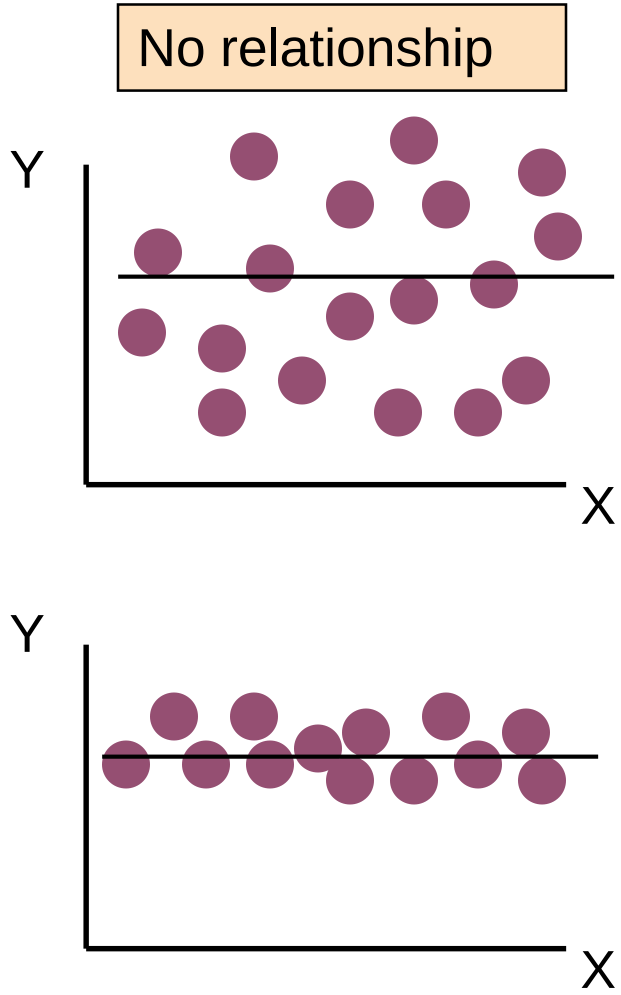

▪Slide from: Statistics for Managers Using Microsoft® Excel 4th Edition, 2004 Prentice-Hall

# Linear Correlation

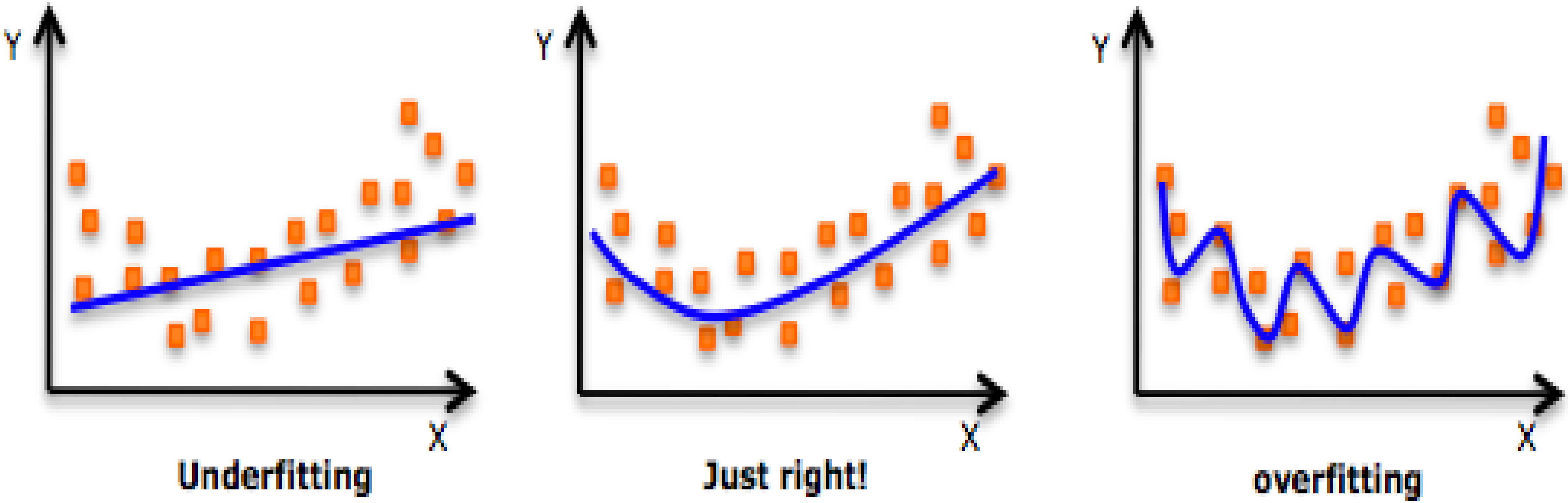



▪Slide from: Statistics for Managers Using Microsoft® Excel  4th Edition, 2004 Prentice-Hall

# Linear Correlation




■Slide from: Statistics for Managers Using Microsoft® Excel 4th Edition, 2004 Prentice-Hall


Y
X
Underfitting
Y
X
Just right!
Y
X
overfitting

Simple Linear Regression

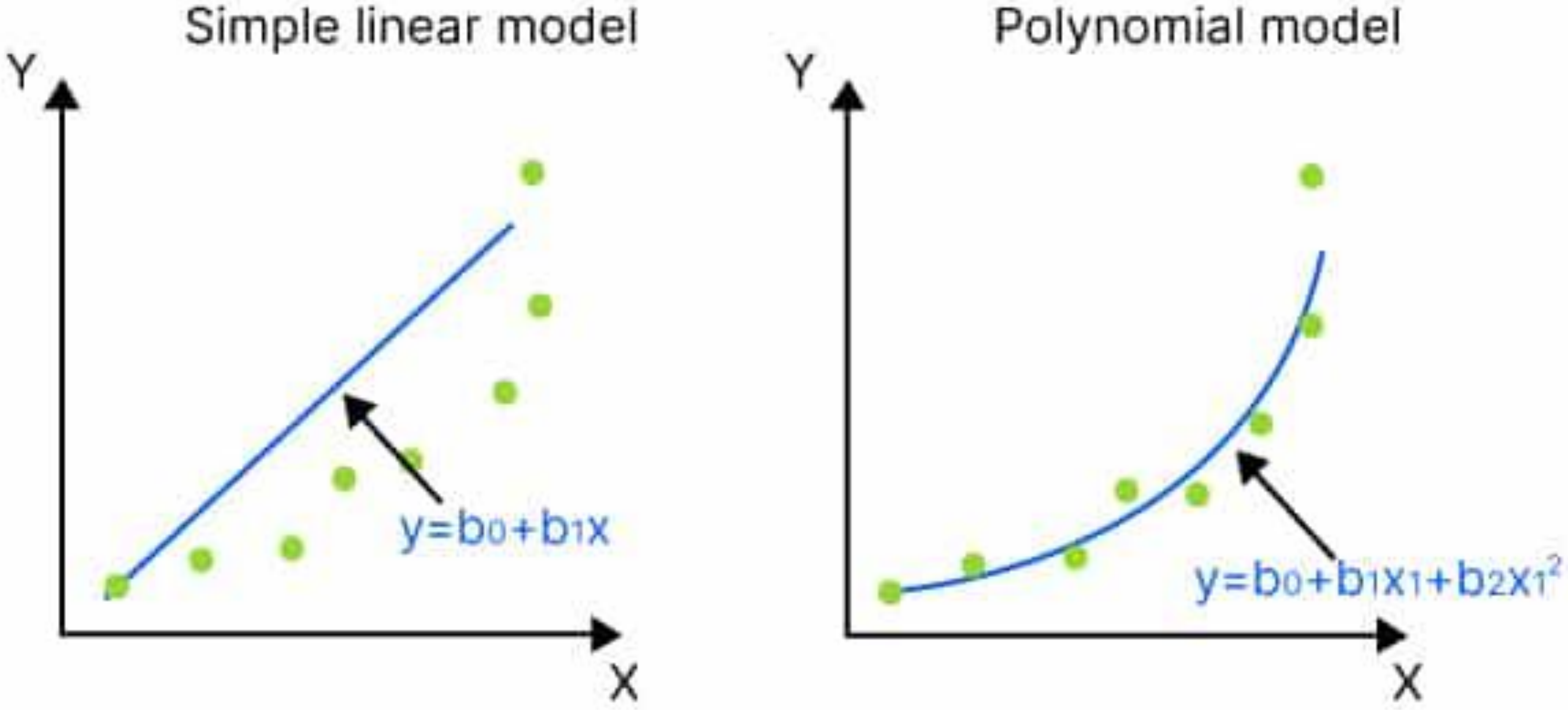
$$y = b_0 + b_1 x_1$$

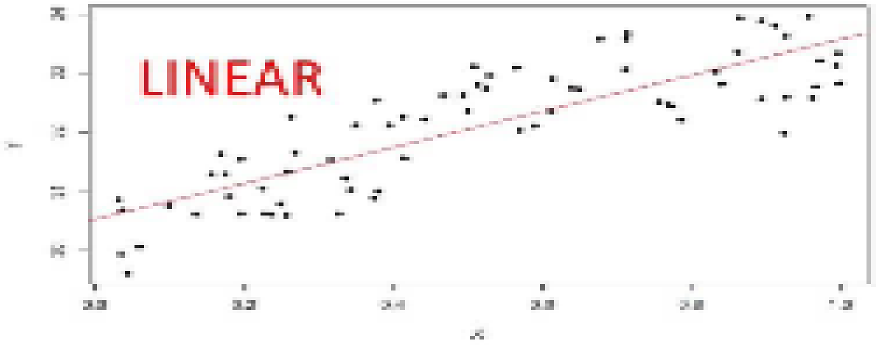
Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Polynomial Linear Regression

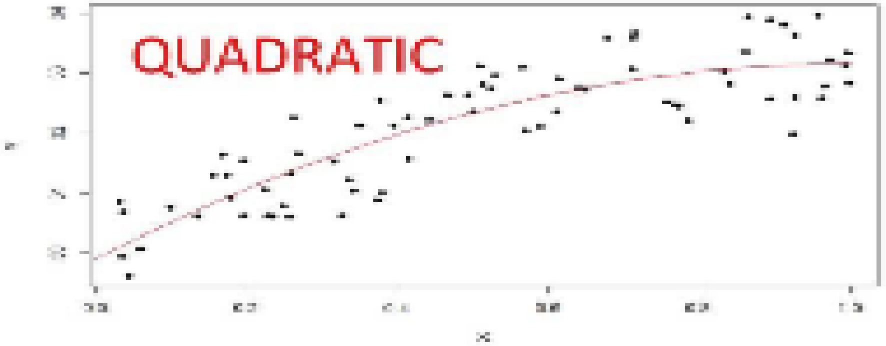
$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \dots + b_n x_1^n$$


Simple linear model
Y
y=b0+b1x
X
Polynomial model
Y
y=b0+b1x1+b2x1²
X



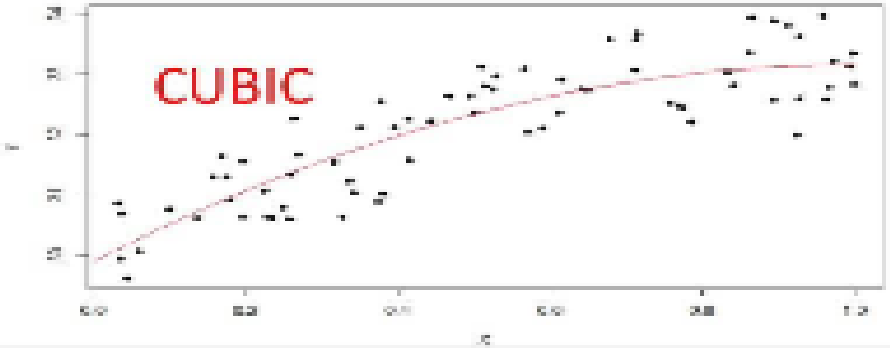

Multiple R-squared: 0.7044

$Y = 30.53 + 3.05*X$



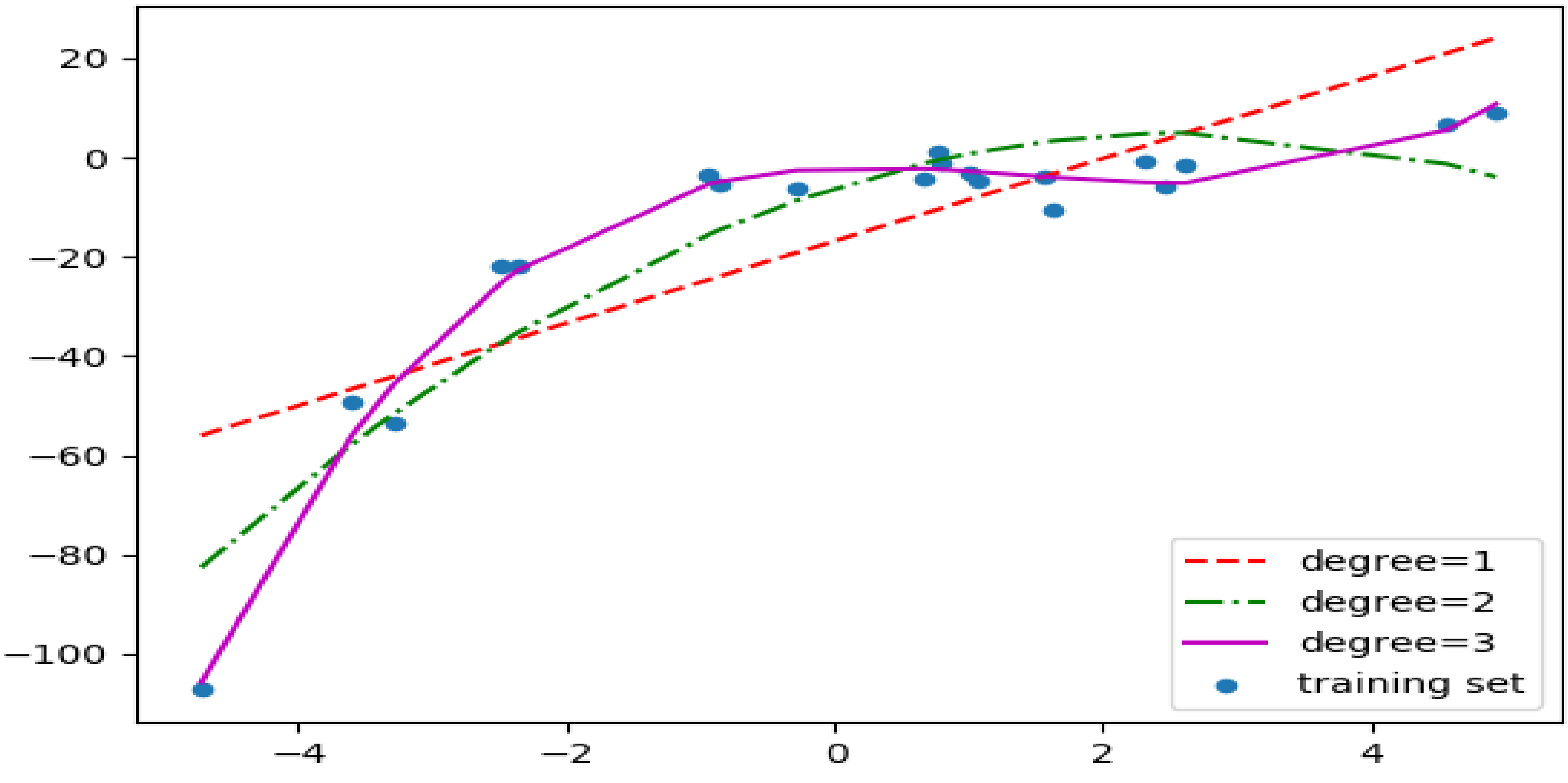

Multiple R-squared: 0.7559

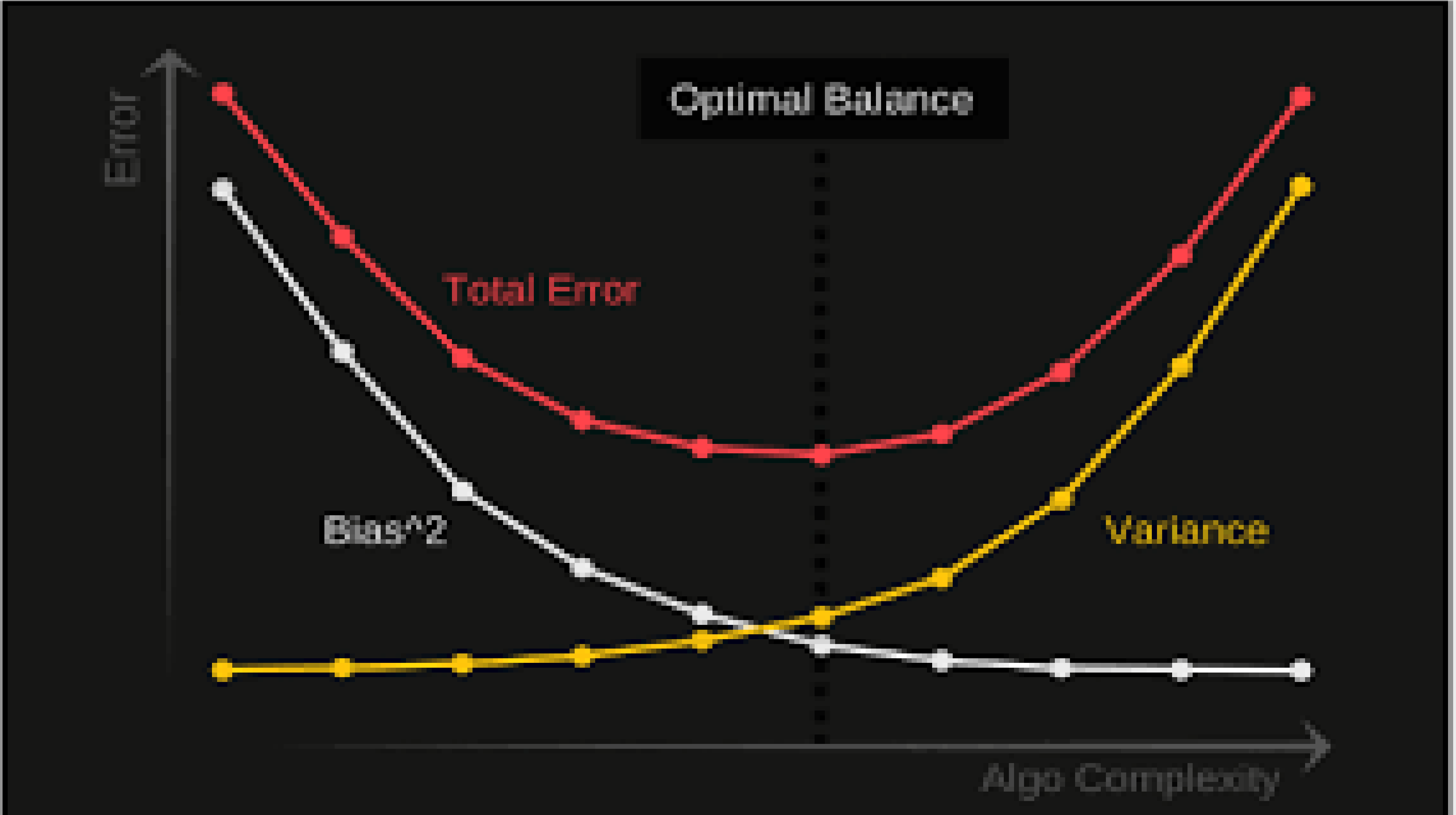
$Y = 29.90 + 6.48*X - 3.22*X^2$

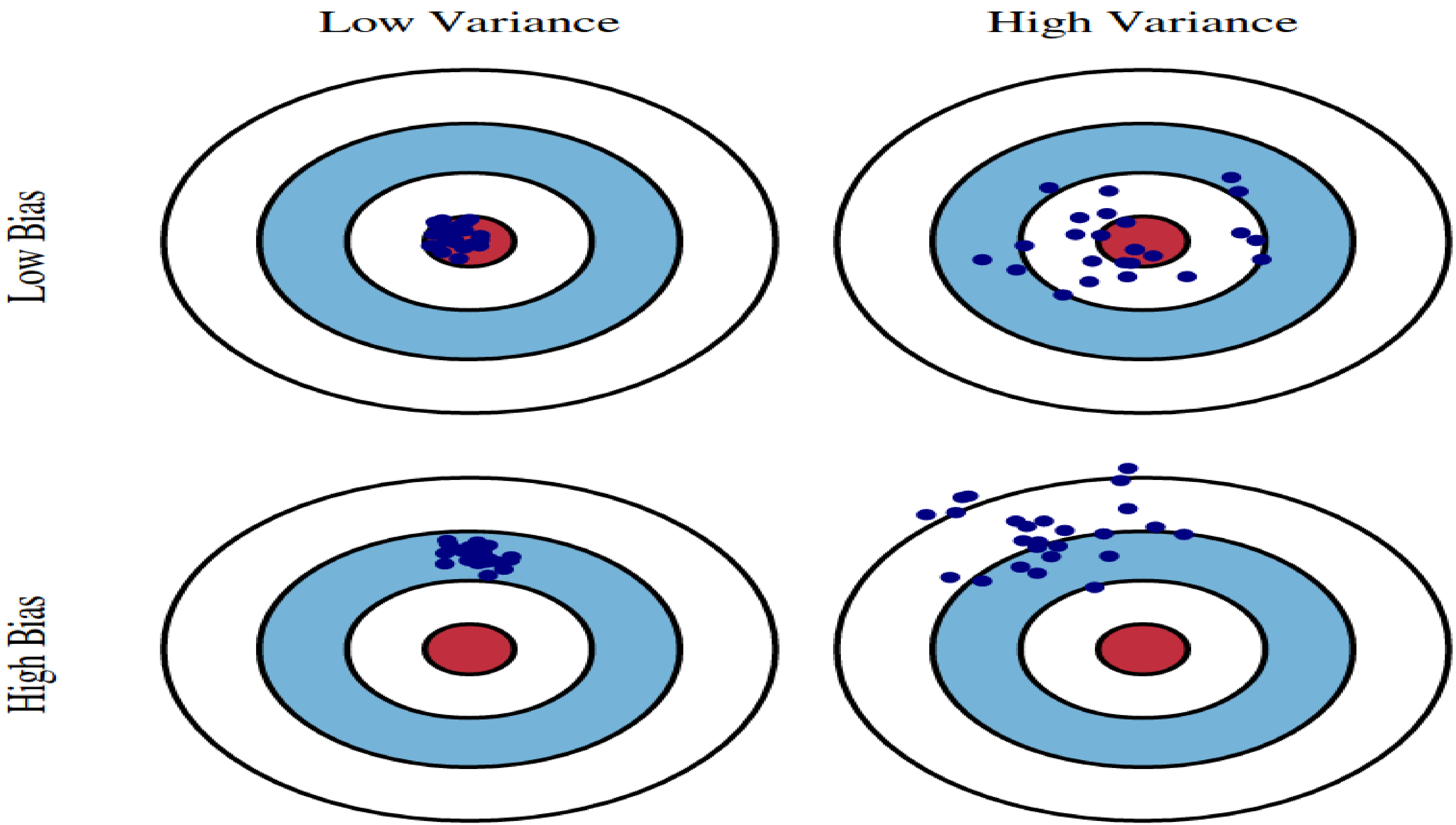



Multiple R-squared: 0.7623

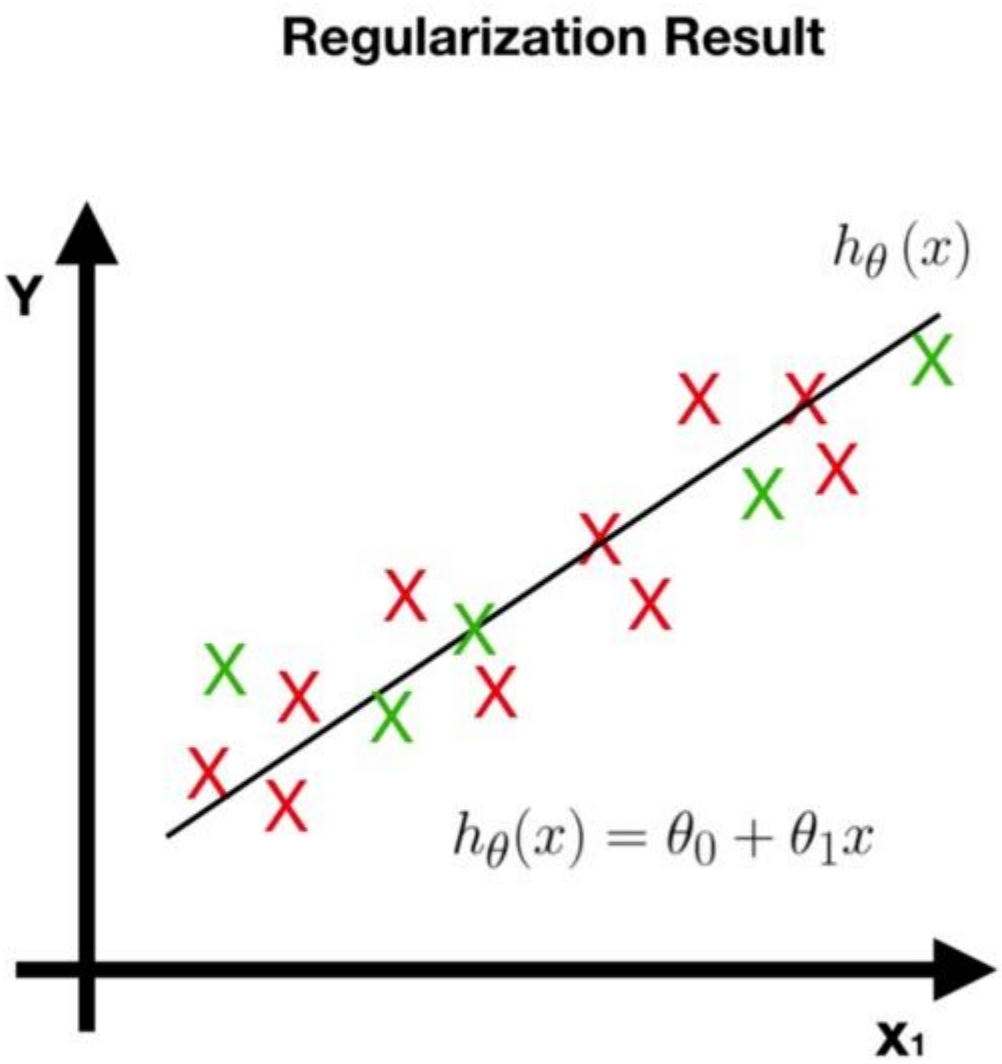
$Y = 30.17 + 3.61*X + 3.71*X^2 - 4.48*X^3$

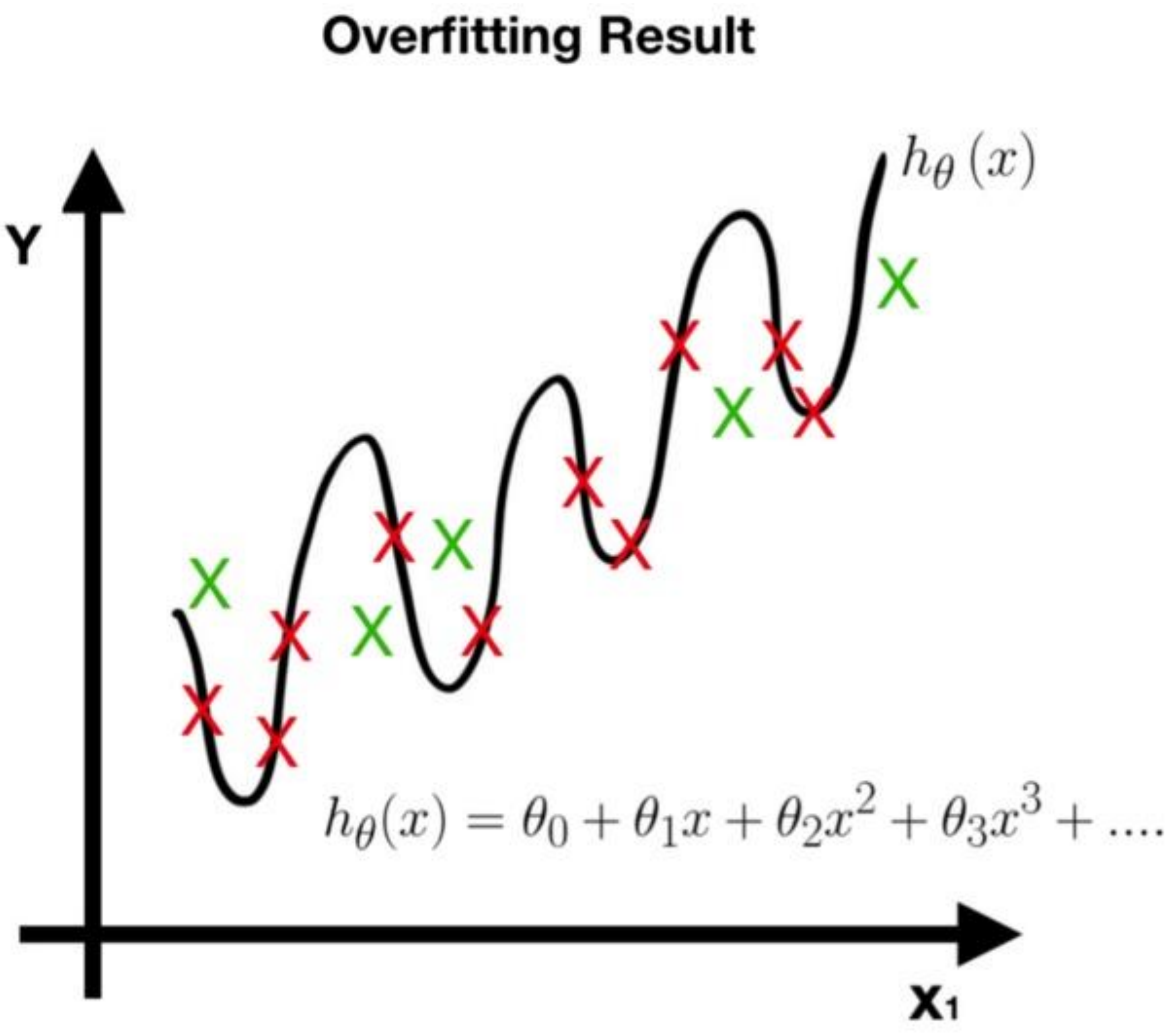

20
0
−20
−40
−60
−80
−100
−4
−2
0
2
4
degree=1
degree=2
degree=3
training set


Optimal Balance
Error
Total Error
Bias^2
Variance
Algo Complexity



Low Variance
High Variance
Low Bias
High Bias


Regularization Result
Y
$h_\theta(x)$
$h_\theta(x) = \theta_0 + \theta_1 x$
$X_1$


Overfitting Result
Y
$h_\theta(x)$
$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + ....$
$X_1$

# Regularization

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \overbrace{\boxed{\underbrace{\boxed{\lambda}}_{\text{Regularization Parameter}}\sum_{j=1}^{n}\theta_j^{\,2}}}^{\text{Regularization Term}}$$

start at $\theta_1$

# Ridge Regression

Ridge regression uses the mean squared error loss function and applies L2 Regularization. Its cost function *J(θ)* is given as

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2 + \lambda \sum_{j=1}^{n} w_j^2$$

where,

$\frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2$ is the Mean Squared error (loss function)

$\lambda \sum_{j=1}^{n} w_j^2$ is the penalty (L2 Regularization)

Now, substitute $\hat{y}$ as $wx_i + b$.

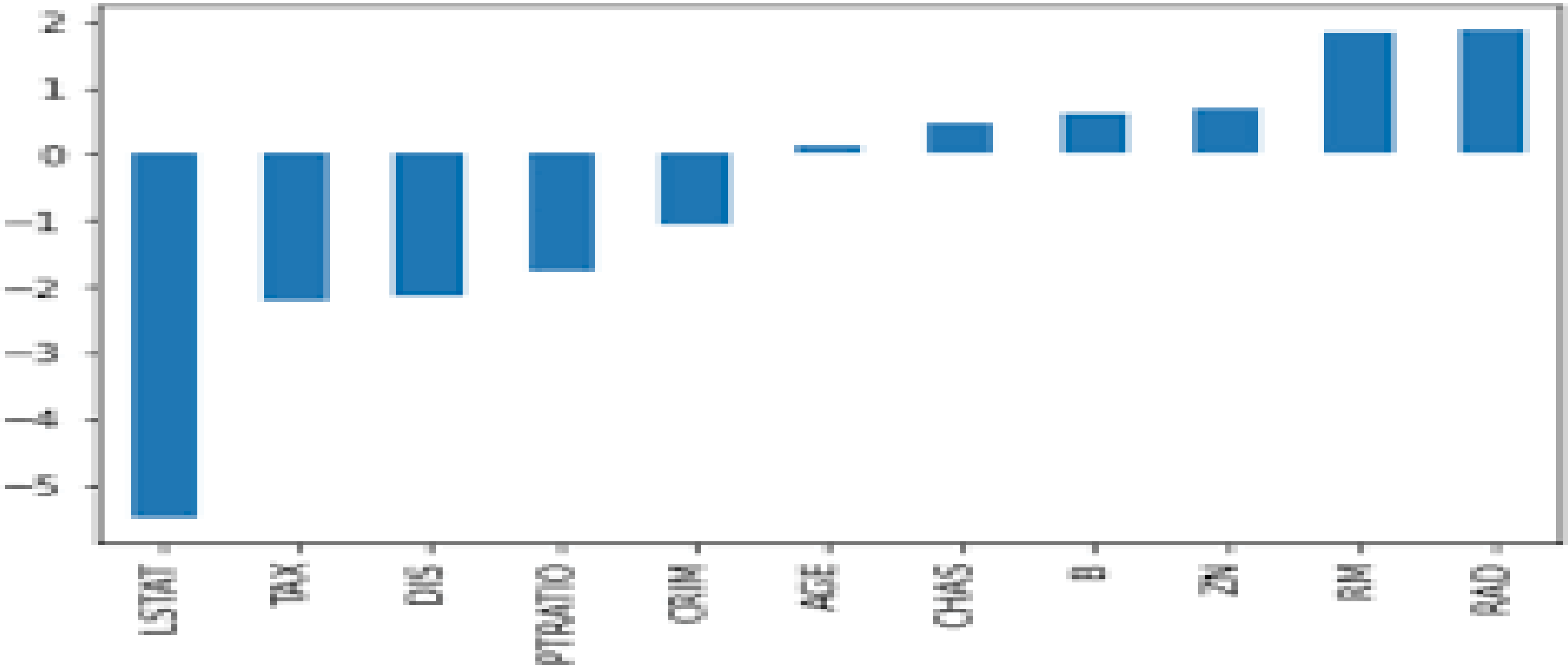
# Lasso Regression

Lasso regression uses the same mean squared error loss function and this applies L1 Regularization and will repeat the same steps as Ridge. The cost function of Lasso Regression *J(θ)* is given as

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2 + \lambda \sum_{j=1}^{n}|w_j|$$

where

$\lambda \sum_{j=1}^{n}|w_j|$ is the penalty (L1 Regularization).

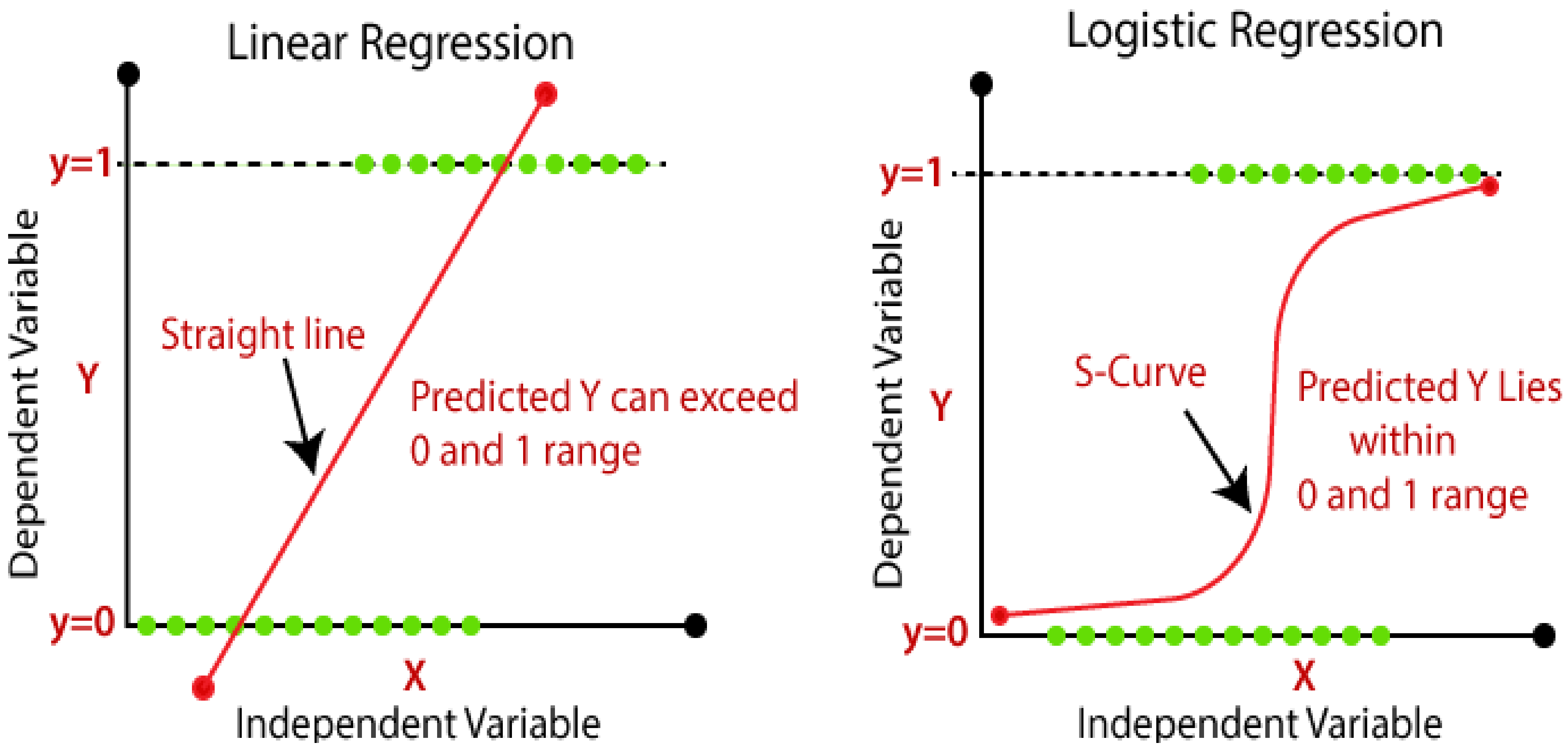

2
1
0
−1
−2
−3
−4
−5
LSTAT
TAX
DIS
PTRATIO
CRIM
AGE
CHAS
B
ZN
RM
RAD

# Ridge Regression

$$Regularization\ (L2) = Loss\ Function + \lambda \sum_{i=1}^{m} w_i^2$$

$$Loss\ Function = \frac{1}{m}\sum_{i=1}^{n}(y - \hat{y})^2$$

$$Regularization\ (L1) = \frac{1}{m}\sum_{i=1}^{n}(y - \hat{y})^2 + \lambda \sum_{j=1}^{m}|w_i|$$


Linear Regression
y=1
y=0
Dependent Variable
Y
Straight line
Predicted Y can exceed
0 and 1 range
X
Independent Variable
Logistic Regression
y=1
y=0
Dependent Variable
Y
S-Curve
Predicted Y Lies
within
0 and 1 range
X
Independent Variable

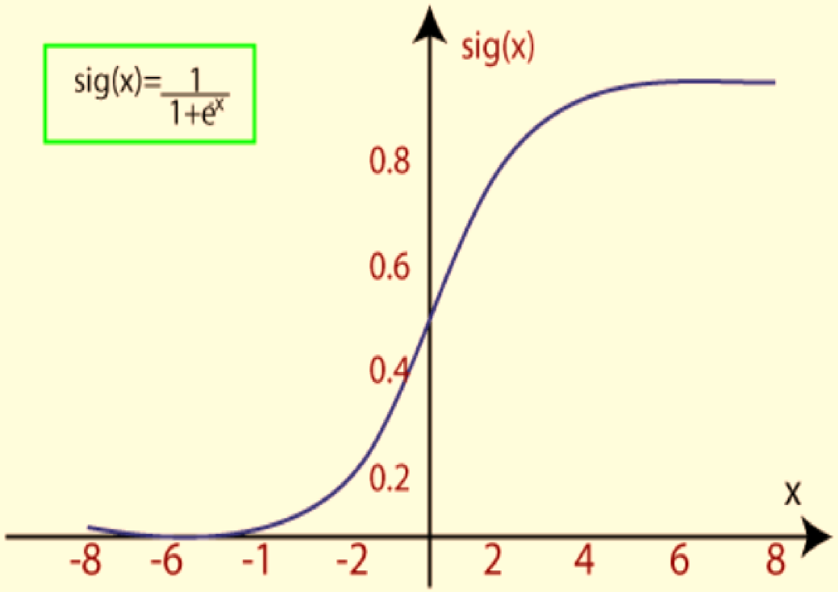
# Logistic Regression:

- Logistic regression is another supervised learning algorithm which is used to solve the classification problems. In **classification problems**, we have dependent variables in a binary or discrete format such as 0 or 1.
- Logistic regression algorithm works with the categorical variable such as 0 or 1, Yes or No, True or False, Spam or not spam, etc.
- It is a predictive analysis algorithm which works on the concept of probability.
- Logistic regression is a type of regression, but it is different from the linear regression algorithm in the term how they are used.
- Logistic regression uses **sigmoid function** or logistic function which is a complex cost function. This sigmoid function is used to model the data in logistic regression. The function can be represented as:

$$f(x) = \frac{1}{1+e^{-x}}$$

- f(x)= Output between the 0 and 1 value.
- x= input to the function
- e= base of natural logarithm.

When we provide the input values (data) to the function, it gives the S-curve as follows




- It uses the concept of threshold levels, values above the threshold level are rounded up to 1, and values below the threshold level are rounded up to 0.

There are three types of logistic regression:

- **Binary(0/1, pass/fail)**
- **Multi(cats, dogs, lions)**
- **Ordinal(low, medium, high)**