Regression

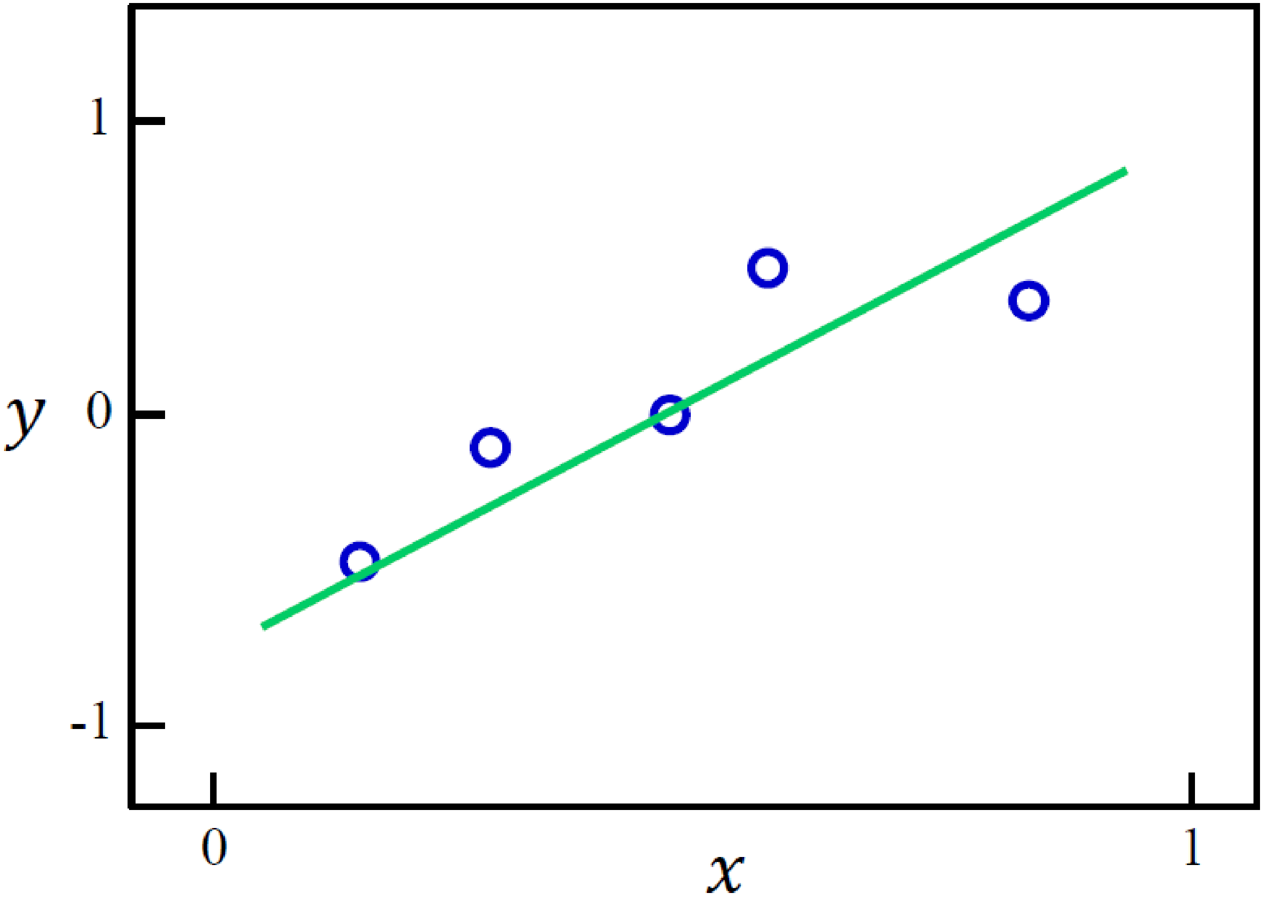
Adams Wai Kin Kong
School of Computer Science and Engineering
Nanyang Technological University, Singapore
adamskong@ieee.org

NANYANG TECHNOLOGICAL UNIVERSITY

Regression: Outline

- Brief review of background
- Linear regression
  - Estimation criterion
  - Least squares solution

Linear Fitting



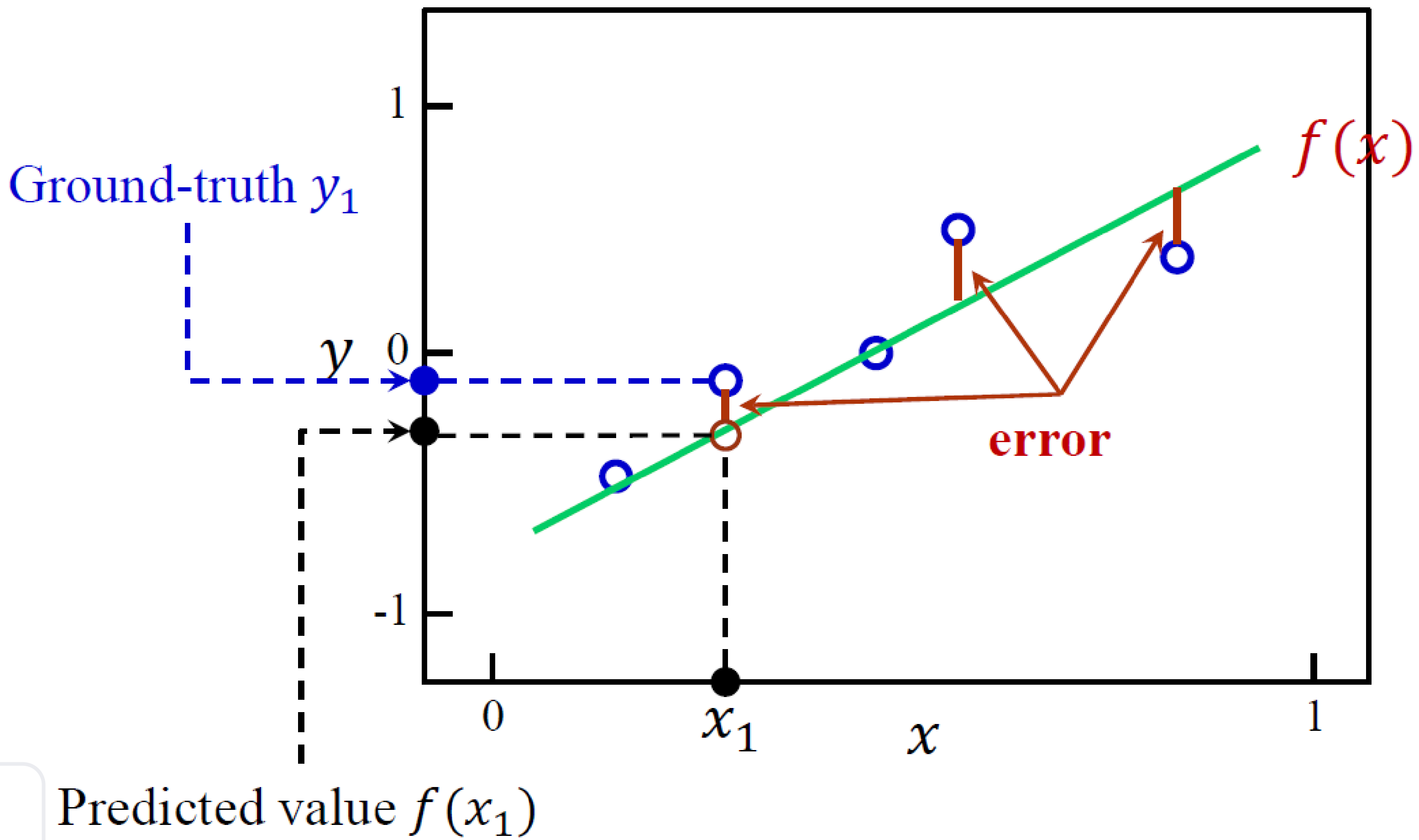
Linear Regression Model

- A special case, an instance is represented by one input feature
- To learn a linear function $f(x)$ in terms of w (drop bias term b for simplicity) from $\{x_i, y_i\}, i=1,\ldots,N$

$$f(x) = w \times x$$

- s.t. the difference (i.e., error) between the predicted values $f(x_i)'s$ and the ground-truth values y_i's is as small as possible

Linear Regression Model (cont.)



Linear Regression Model (cont.)

- Suppose sum-of-squares error is used

$$E(w) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i) - y_i)^2 = \frac{1}{2}\sum_{i=1}^{N}(w \times x_i - y_i)^2$$

- Learn the linear model in terms of w by minimizing the error

$$w^* = argmin_w E(w)$$

Linear Regression Model (cont.)

- To solve the unconstrained minimization problem, we can set the derivative of $E(w)$ w.r.t. w to zero

$$\frac{dE(w)}{dw} = \frac{d\frac{1}{2}\sum_{i=1}^{N}(w \times x_i - y_i)^2}{dw} = 0$$

$$\sum_{i=1}^{N}(w \times x_i - y_i) \times x_i = 0$$

Linear Regression Model (cont.)

$$\sum_{i=1}^{N}(w \times x_i - y_i) \times x_i = 0$$

$$\Downarrow$$

$$w\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} y_i x_i = 0$$

$$\Downarrow$$

$$w = \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2}$$

Linear Regression Model (cont.)

- To learn a linear function $f(\boldsymbol{x})$ in more general form in terms of $\boldsymbol{w}$ and b,

$$f(\boldsymbol{x})=\boldsymbol{w}\cdot\boldsymbol{x}+b$$

where $\boldsymbol{w} = [w_1, \cdots w_d]$ and $x = [x_1, \cdots x_d]$

- By defining $w_0 = b$ and $x_0 = 1$, $\boldsymbol{w}$ and $\boldsymbol{x}$ are of d+1 dimensions

$$f(\boldsymbol{x})=\boldsymbol{w}\cdot\boldsymbol{x}$$

where $\boldsymbol{w} = [w_0\ w_1, \cdots w_d]$ and $x = [1\ x_1, \cdots x_d]$

Linear Regression Model (cont.)

- Suppose sum-of-squares error is used

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i) - y_i)^2 = \frac{1}{2}\sum_{i=1}^{N}(w \cdot x_i - y_i)^2$$

- Learn the linear model in terms of $\boldsymbol{w}$ by minimizing the error

$$\boldsymbol{w}^* = argmin_{\boldsymbol{w}} E(\boldsymbol{w})$$

Linear Regression Model (cont.)

- To solve the unconstrained minimization problem, we can set the derivative of $E(\boldsymbol{w})$ w.r.t. $\boldsymbol{w}$ to **zero**

$$\frac{\partial E(\boldsymbol{w})}{\partial \boldsymbol{w}} = \frac{\partial \frac{1}{2} \sum_{i=1}^{N} (w \cdot x_i - y_i)^2}{\partial x} = \mathbf{0}$$

Linear Regression Model (cont.)

- We can obtain a closed-form solution

$$\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_N} \end{bmatrix} = \begin{bmatrix} x_{10} & \cdots & x_{1\boldsymbol{d}} \\ \vdots & \ddots & \vdots \\ x_{N0} & \cdots & x_{\boldsymbol{Nd}} \end{bmatrix}$$

The first d-dimensional training data point

The N d-dimensional training data point

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}.$$

Regression: Summary for N dimensional case

For an n-dimensional linear regression, $f(x, w) = w_0 + \sum_{i=1}^{n} w_i x_i$ the equation

$$\boldsymbol{w}^* = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

The corresponding $\boldsymbol{X}, \boldsymbol{y}$ and w are defined below.

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \\ y_m \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & \cdots & x_{md} \end{bmatrix} \text{ and } \boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ \vdots \\ w_n \end{bmatrix}$$

Example

Given the five data points in a 2D space, compute the optimal w_0 and w_1 in the linear regression model $y = w_0 + w_1 x$.

Five data points in 2D space					
(x, y)	(1, 1)	(0, 2)	(2,1)	(3,2)	(2,0)

$$y = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 0 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{w_0} \\ \boldsymbol{w_1} \end{bmatrix} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$w_0 = 1.3846, w_1 = -0.1154$$

Example

Compute the regression outputs for the four testing data points below and compute the mean square error.

Four testing data points				
Testing data	-1	5	-3	4
Testing data y	1	2	2	-1

Regression outputs from the model $y = 1.3846 - 0.1154x$

Regression y	1.5	0.8076	1.7308	0.923

Mean square error:

$$\frac{1}{4} * (1.5 - 1)^2 + (2 - 0.8076)^2 + (2 - 1.7308)^2 + (-1 - 0.923)^2$$

$$=1.360554$$

NANYANG TECHNOLOGICAL UNIVERSITY

Regression: Polynomial regression

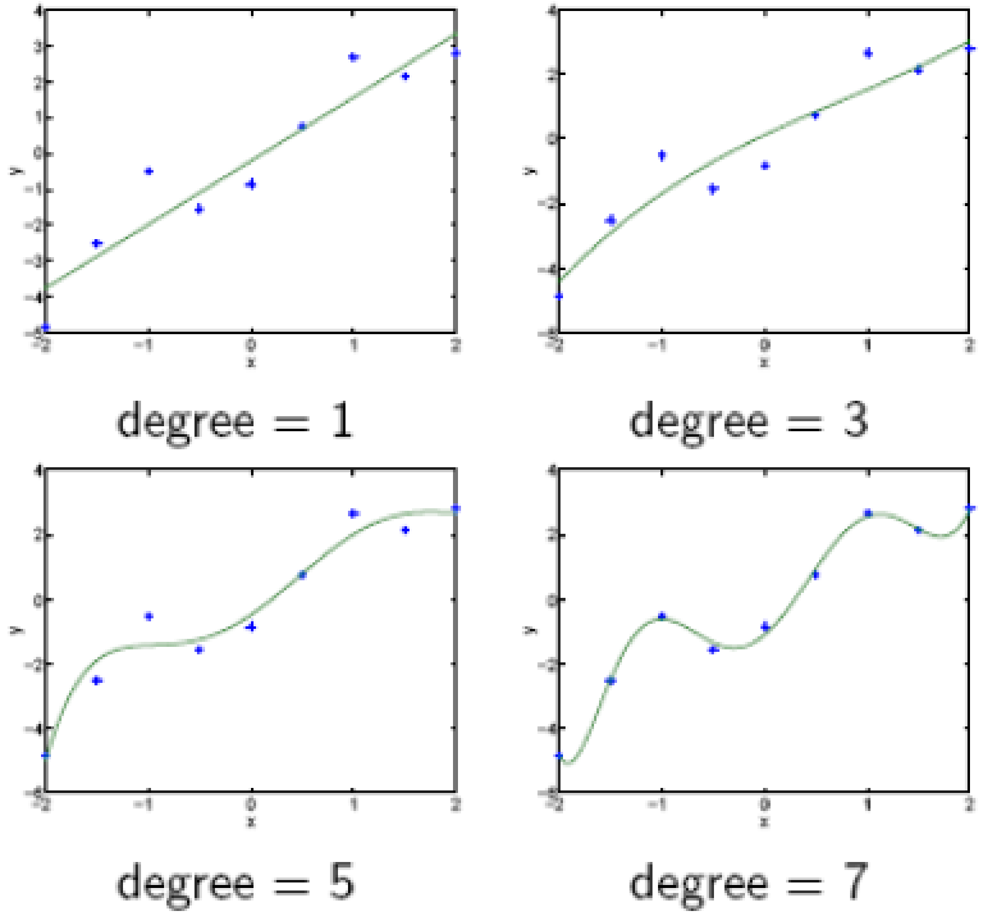
- For n^{th} order polynomial, $f(x, w) = \sum_{i=0}^{n} w_i x^i$, the equation

$$\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

The corresponding $\boldsymbol{X}$, $\boldsymbol{y}$ and $\boldsymbol{w}$ are defined below.

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \\ y_m \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & \cdots & x_m^n \end{bmatrix} \text{ and } \boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ \vdots \\ w_n \end{bmatrix}$$

Regression: Polynomial regression



Evaluation

- Root Mean Square Error (RMSE)

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2}$$

- Mean Absolute Error (MAE)

$$\frac{1}{N}\sum_{i=1}^{N}|f(x_i) - y_i|$$

Thank you