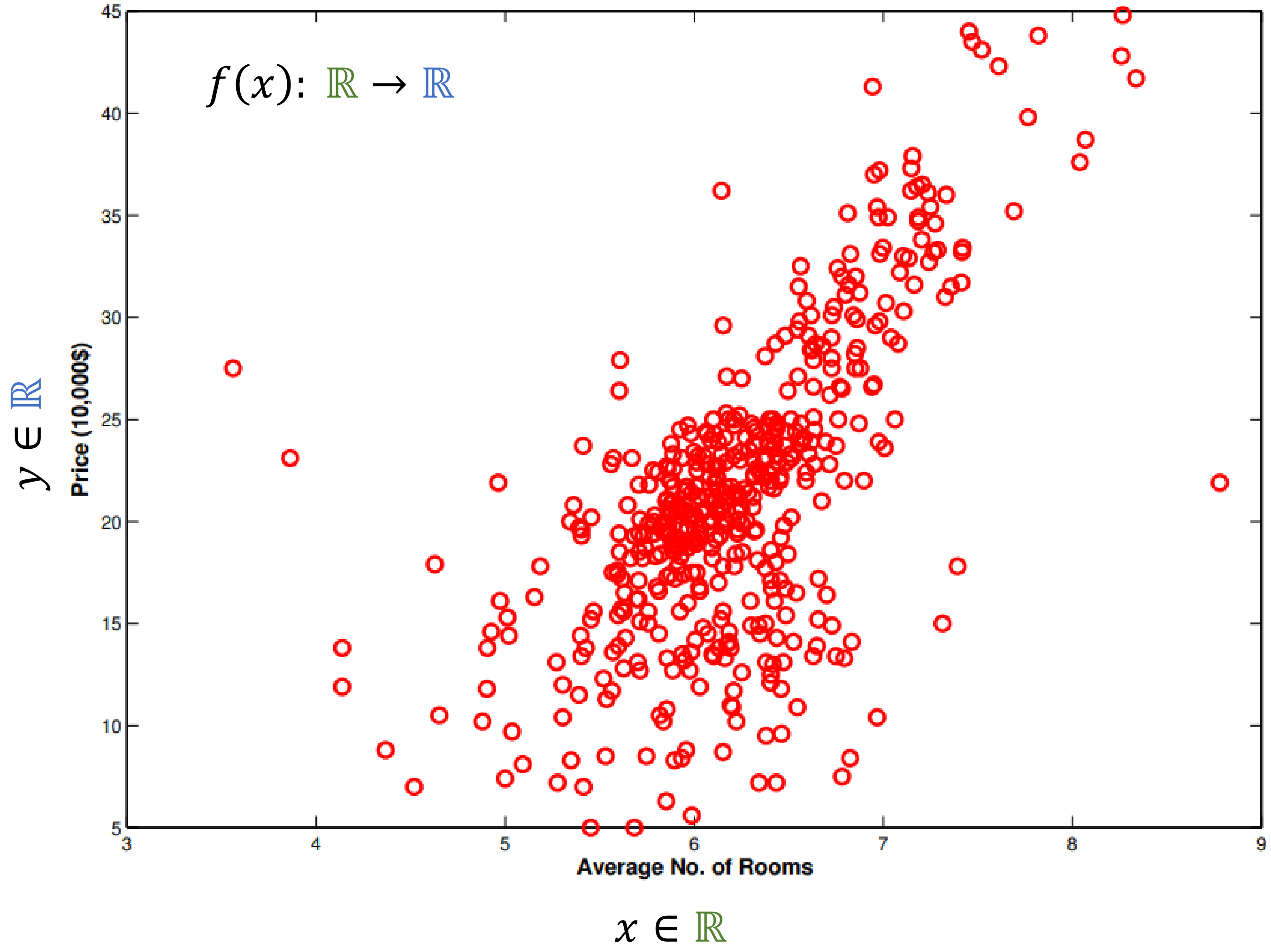
# Linear Regression

Dr. Uzair Ahmad

# Agenda

- Regression formulation
- Assumptions, Implications, Diagonostics
- Analytic solution
- Iterative solution
- Pitfalls
- Ridge Regression

# Example



$f(x)\colon \mathbb{R} \to \mathbb{R}$
$y \in \mathbb{R}$
Price (10,000$)
45
40
35
30
25
20
15
10
5
3
4
5
6
7
8
9
Average No. of Rooms
$x \in \mathbb{R}$

# A linear model

Parameters

$$y = f(x) = \underbrace{w_0 + w_1}_{} x$$

Dependent variable
(house price)

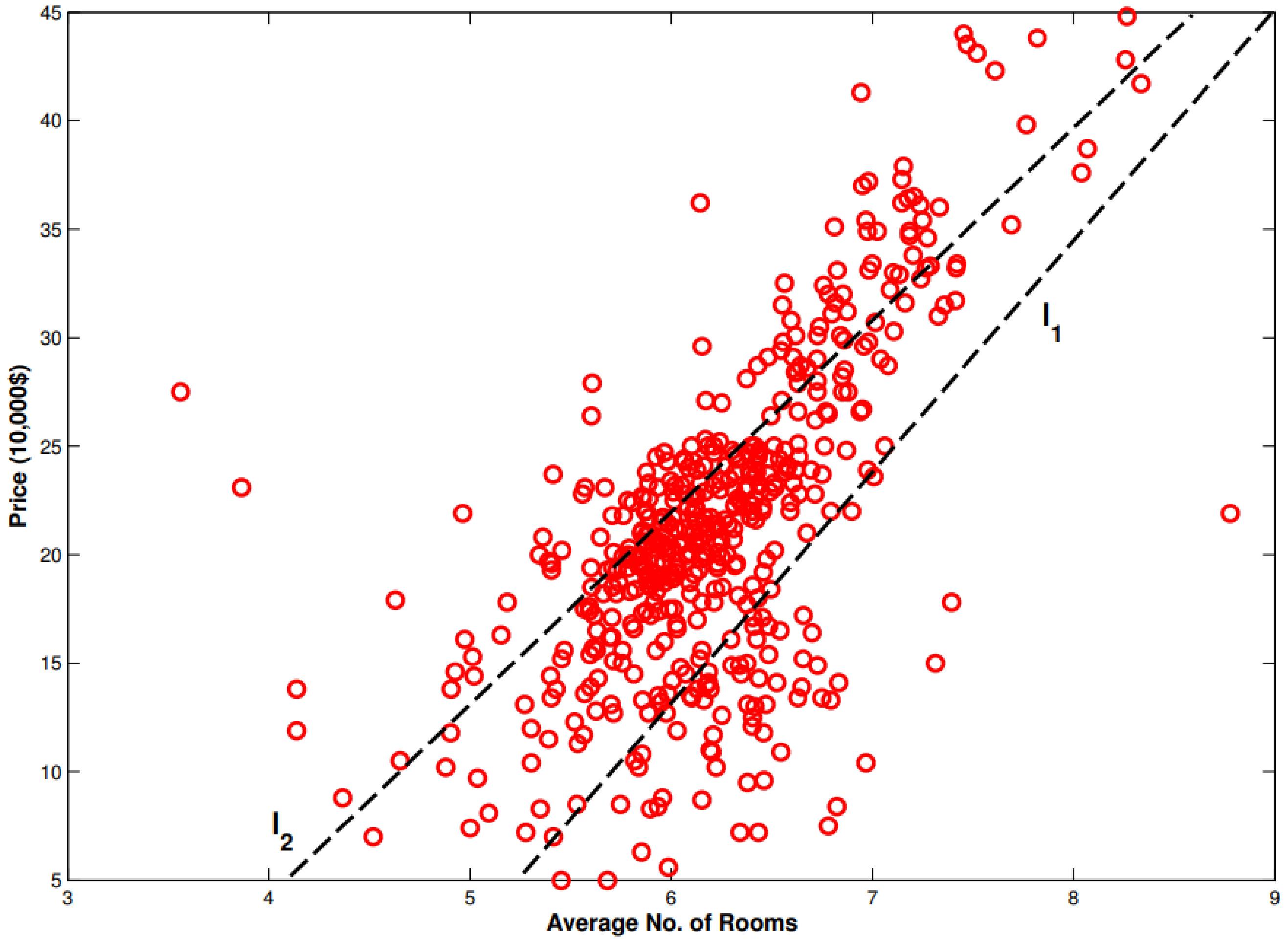
Independent variable
(# of rooms)

# A linear model

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

$$y_1 = w_0 + w_1 x_1$$
$$y_2 = w_0 + w_1 x_2$$
$$y_3 = w_0 + w_1 x_3$$
$$\dots$$
$$y_N = w_0 + w_1 x_N$$

# A linear model



# A linear model: multiple features

$$\boxed{x_i \in \mathbb{R}^m}$$

$$\boxed{y_i \in \mathbb{R}^N}$$

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$f(x): \mathbb{R}^m \to \mathbb{R}$$

$$y = f_w(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_m x_m$$

# A linear model: multiple features

$$f(x)\colon \mathbb{R}^{m+1} \rightarrow \mathbb{R}$$

$$y = f_w(x) = w_0\, x_0 + w_1 x_1 + w_2 x_2 + \ldots + w_m\, x_m$$

$$y = f_w(x) = \sum_{j=1}^{m} w_j x_j$$

# A linear model: multiple features

$$y = f_w(x) = w_0\, x_0 + w_1 x_1 + w_2 x_2 + \ldots + w_m\, x_m$$

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nm} \end{bmatrix}$$

# Assumptions

- Linearity
- IID
- Homoscedasticity
- Normality
- No Multicollinearity

# The learning task

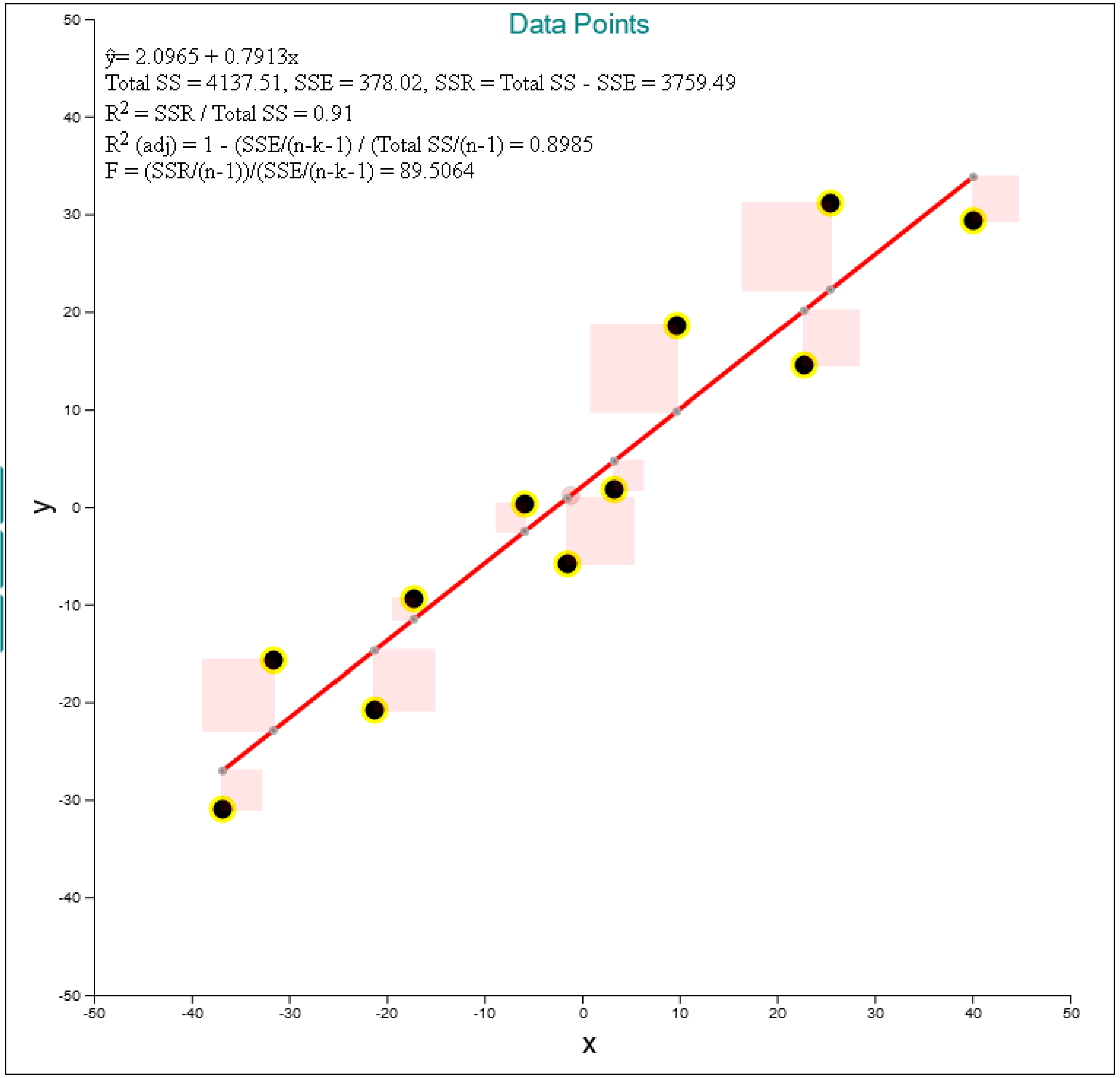
$f(x)\colon \mathbb{R}^{m+1} \rightarrow \mathbb{R}$

$$\hat{y} = f_w(x) = \sum_{j=1}^{m} w_j x_j$$

$$E(w_m) = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{N} \Big(\sum_{j=1}^{m} w_j x_j \; - y_i\Big)^2 = (Xw - y)^T (Xw - y)$$

$$\underset{w}{\arg\min} (Xw - y)^T (Xw - y)$$

# Least Squares



# Ordinary Least Squares

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$E = \begin{bmatrix} f_{\mathbf{w}}(\mathbf{x}_1) - y_1 \\ \cdots \\ f_{\mathbf{w}}(\mathbf{x}_N) - y_N \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \mathbf{w} \\ \cdots \\ \mathbf{x}_N^T \mathbf{w} \end{bmatrix} - \begin{bmatrix} y_1 \\ \cdots \\ y_N \end{bmatrix} = X\mathbf{w} - \mathbf{y}$$

# Ordinary Least Squares

$$J(w) = \frac{1}{2}\sum_{i=1}^{N}(f_w(\mathbf{x}_i) - y_i)^2 = \frac{1}{2}E^T E = \frac{1}{2}(X\mathbf{w} - \mathbf{y})^T(X\mathbf{w} - \mathbf{y})$$

$$\nabla_w J(\mathbf{w}) \quad = \quad \nabla_w \frac{1}{2}E^T E = \nabla_w \frac{1}{2}(X\mathbf{w} - \mathbf{y})^T(X\mathbf{w} - \mathbf{y})$$

$$= \quad X^T X\mathbf{w} - X^T\mathbf{y}$$

$$\boxed{\hat{\mathbf{w}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}}$$

# Ordinary Least Squares

$$J(w) = \frac{1}{2}\sum_{i=1}^{N}(f_w(\mathbf{x}_i) - y_i)^2 = \frac{1}{2}E^T E = \frac{1}{2}(X\mathbf{w} - \mathbf{y})^T(X\mathbf{w} - \mathbf{y})$$

$$\nabla_w J(\mathbf{w}) = \frac{1}{2}\left(X^T\mathbf{w}^T X\mathbf{w} - X^T\mathbf{w}^T y - \mathrm{y}^T X\mathbf{w} + y^T y\right)$$

$$\nabla_w(X^T\mathbf{w}^T X\mathbf{w}) = 2X^T X$$

$$\nabla_w(X^T\mathbf{w}^T y) = X^T y$$

$$\nabla_w(\mathrm{y}^T X\mathbf{w}) = \mathrm{y}^T X$$

$$\nabla_w(\mathrm{y}^T \mathrm{y}) = 0$$

$$\nabla_w J(\mathbf{w}) = \frac{1}{2}(2X^T X\mathbf{w} - X^T y - \mathrm{y}^T X + 0)$$

$$\nabla_w J(\mathbf{w}) = \frac{1}{2}(2X^T X\mathbf{w} - 2X^T y) = \frac{2}{2}(X^T X\mathbf{w} - X^T y) = X^T X\mathbf{w} - X^T y$$

$$\nabla_w J(\mathbf{w}) = X^T X\mathbf{w} - X^T y = 0$$

$$\boxed{\mathbf{w} = (X^T X)^{-1} X^T y}$$

# Iterative Least Squares

$$E(w_m) = \sum_{i=1}^{N}(\hat{y}_i - y_i)^2 = \sum_{i=1}^{N}(\sum_{j=1}^{m} w_j x_j \ - y_i)^2 = \sum_{i=1}^{N}(W^T x_i - y_i)^2$$

$$J = \frac{1}{2}\sum_{i=1}^{N}(W^T x_i - y_i)^2$$

$$\frac{\partial J}{\partial w_j} J = \frac{\partial J}{\partial w_j}\ \frac{1}{2}\sum_{i=1}^{N}(W^T x_i - y_i)^2$$

$$\frac{\partial J}{\partial w_j} J = \frac{1}{2}\frac{\partial J}{\partial w_j}\sum_{i=1}^{N}(W^T x_i - y_i)^2$$

$$\frac{\partial J}{\partial w_j} J = \frac{2}{2}\frac{\partial J}{\partial w_j}\sum_{i=1}^{N}(W^T x_i - y_i)^1 . \frac{\partial J}{\partial w_j}\,(W^T x_i - y_i)$$

# Iterative Least Squares

$$E(w_m) = \sum_{i=1}^{N} (\widehat{y_i} - y_i)^2 = \sum_{i=1}^{N} (\sum_{j=1}^{m} w_j x_j \;- y_i)^2$$

$$\frac{\partial J}{\partial w_j} J = \frac{2}{2} \sum_{i=1}^{N} (W^T x_i - y_i)^1 . \frac{\partial J}{\partial w_j} (W^T x_i - y_i)$$

$$\frac{\partial J}{\partial w_j} J = \sum_{i=1}^{N} (W^T x_i - y_i) . \frac{\partial J}{\partial w_j} (W^T x_i - y_i)$$

$$\frac{\partial J}{\partial w_j} J = \sum_{i=1}^{N} (W^T x_i - y_i) . (x_i)$$

Repeat until converged:

$$w_j^{(t+1)} = w_j^{(t)} \text{ - } \eta \frac{\partial J}{\partial w_j}$$

# Regression Challenges

- Function choice
  - Linear
  - Polynomial
- Generalizability
  - Overfitting
    - Cross validation
- Feature engineering
- Multicollinearity