Chapter 3: Linear Regression

Linear regression in classical statistics

Linear regression in machine learning

- Given a dataset:

$\{(\mathbf{x}^{(i)}, y^{(i)}) \mid i_{=1}^{=n}\}$

Of **n** data points.

- Vector **x** represents the input features.

- Dimension of **x** denotes number of input features. We call this d.

- Scalar y represents the output.

- We propose a hypothesis h parameterized by θ in the form:

$$h_\theta(x) = \sum_{i=0}^{d} \theta_i x_i = \theta^T \mathbf{x}$$

- Where we define $x_0 = 1$ for all data points.

- Thus $\theta \in \mathbb{R}^{n+1}$

Linear regression in machine learning

- To assess how "good" the prediction is, we define the **cost function** as:

$$L(\theta) = (\hat{\mathbf{y}} - \mathbf{y})^2 = \frac{1}{2}\sum_{i=1}^{n}(h_\theta(x^{(i)}) - y^{(i)})^2$$

- We need to choose θ so as to minimize $L(\theta)$

- For this we use an optimization algorithm called **Gradient Descent**. Which is defined as:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} L(\theta)$$

- Where α is the **learning rate**. And θ gets updated in each step.

Linear regression in machine learning

$$\frac{\partial}{\partial \theta_j} L(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2$$

$$= \frac{1}{2} .2 (h_\theta(x) - y) \, . \frac{\partial}{\partial \theta_j} (h_\theta(x) - y)$$

$$= (h_\theta(x) - y) \, . \frac{\partial}{\partial \theta_j} (\sum_{j=1}^{n} \theta_j x_j - y)$$

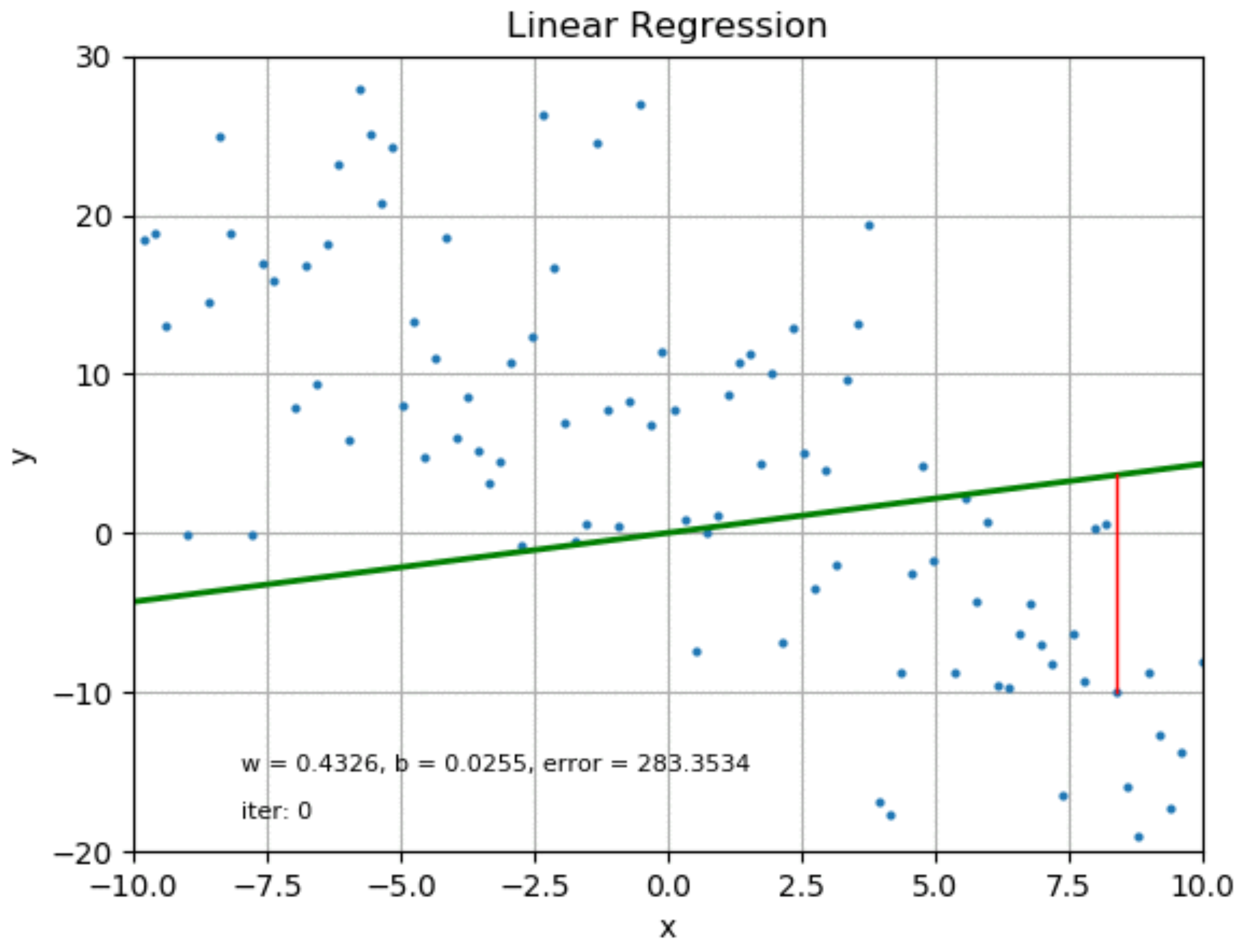
$$= (h_\theta(x) - y) x_j$$

- Thus:

$$\theta_j := \theta_j + \alpha \sum_{i=1}^{n} (y^{(i)} - h_\theta(x^{(i)})) x_j^{(i)}$$

Linear regression in machine learning

$$\theta_j := \theta_j + \alpha \sum_{i=1}^{n} (y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$$



هوش مصنوعی
رگرسیون خطی

learning supervised LR

x_0	x_1	x_2	y
1	75	2	4,5
1	123	3	11,5
1	40	1	2,5
1	230	4	17

n

feature

target

$\{ (x, y) , i \; (1-n) \}$

$[x_1 , x_2]$ $[y]$

1	415	8

learning

1) فرض خطی بودن

2) معیار برای خوب بودن عملکرد پیش‌بینی

3) عملکرد ⇑

Loss کمینه کردن

gradient descent

$$h_\theta : \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2$$
$$= [\theta_0 , \theta_1 , \theta_2] \cdot [x_0 , x_1 , x_2]$$
$$= \boxed{\theta^T x}$$

بردار θ رو رندوم بچین = مرحله اول

یا $\theta = [\theta_0 \quad \theta_1 \quad \theta_2]$
$= [7 \quad 1,5 \quad 3]$

$$x^{(1)} \quad \boxed{1 \quad 75 \quad 2 \quad 4,5}$$

$$x_0^{(1)} \quad x_1^{(1)} \quad x_2^{(1)} \quad y^{(1)}$$

$$\theta = [7 \quad 1,5 \quad 3]$$

پیش‌بینی

$$\theta^T x = \begin{bmatrix} 7 \\ 1,5 \\ 3 \end{bmatrix} [1 \quad 75 \quad 2] =$$

$$7 \times 1 + 1,5 \times 75 + 3 \times 2$$

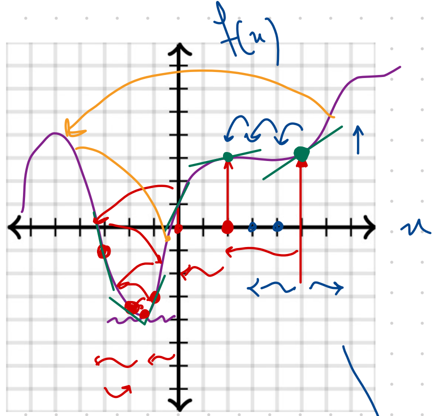
$$y^{(1)} = 4,5 \qquad \hat{y}^{(1)} = \boxed{1462,5}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \frac{1}{2} \left(\hat{y}^{(i)} - y^{(i)} \right)^2$$

loss function

Optimization

↳ gradient descent



$$h_\theta = \theta^T x = \hat{y}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \frac{1}{2}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

اول θ رو رندوم

$$\theta_{t+1} = \theta_t - \alpha \frac{\partial L}{\partial \theta}$$

learning rate

$$\theta_{t+1} = \theta_t - \alpha \nabla \mathcal{L}$$

$$\underbrace{(\theta^T x - y)\, x}$$

$$\begin{bmatrix} \theta_j \\ \vdots \\ \theta_j \end{bmatrix} = \begin{bmatrix} \theta_j \end{bmatrix} - \alpha \boxed{\frac{\partial \mathcal{L}}{\partial \theta_j}}$$

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} \frac{1}{2} \left(\hat{y}^{(i)} - y^{(i)} \right)^2$$

$$= \frac{\partial}{\partial \theta_j} \boxed{\sum_{i=1}^{n} \frac{1}{2} \left(\theta^T x^{(i)} - y^{(i)} \right)^2}$$

یکا خوبة . $\frac{\partial}{\partial \theta_j} \frac{1}{2} \left(\theta^T x - y \right)^2$

chain rul

$$= \cancel{\frac{1}{2}} \times 2 \left(\theta^T x - y \right) \cdot \frac{\partial}{\partial \theta} \left(\theta^T x - y \right)$$

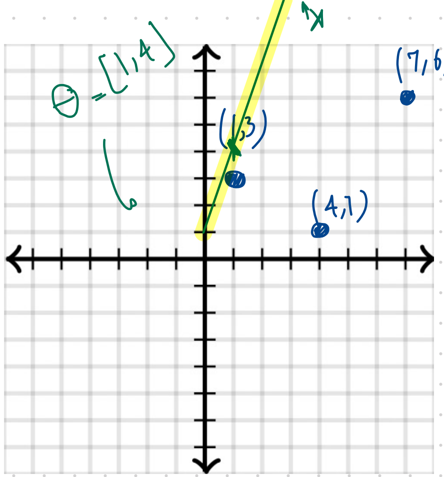
اختلاف $\leftarrow \left(\theta^T x - y \right) \cdot \frac{\partial}{\partial \theta} \underbrace{\left(\theta^T x - y \right)}$

$$\boxed{\left(\theta^T x - y \right) \cdot x}$$

$$\frac{\partial}{\partial \theta_0} \left(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 - y \right) = x_0$$

$$\frac{\partial}{\partial \theta_1} \left(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 - y \right) = x_1$$

$$\frac{\partial}{\partial \theta_2} \qquad = x_2$$



x_0	x_1	y	$\theta_0 x_0 + \theta_1 x_1$	$\hat{y}$
1	1	3	$1\times1 + 4\times1$	5
1	4	1	$1\times1 + 4\times4$	17
1	7	6	$1\times1 + 4\times7$	29

$$h_\theta = \hat{y} = \theta^T x$$

θ [illegible]

$$\theta = [\theta_0, \theta_1]$$

$$\theta = [1, 4]$$

$$\mathcal{L}_\theta = \sum \frac{1}{2}(\hat{y} - y)^2$$

$$\mathcal{L} \quad 394,5 = \frac{1}{2}\overset{4}{(5-3)^2} + \frac{1}{2}\overset{256}{(17-1)^2} + \frac{1}{2}\overset{529}{(29-6)^2}$$

x_0	x_1	y	$\hat{y} = \theta_0 x_0 + \theta_1 x_1$	
1	1	3	$1\times1 + 4\times1$	5
1	4	1	$1\times1 + 4\times4$	17
1	7	6	$1\times1 + 4\times7$	29

x_0	x_1	y	$\hat{y}$	$(\hat{y}-y)$	$(\hat{y}-y)x_0$	$(\hat{y}-y)x_1$
1	1	3	5	2	2	2
1	4	1	17	16	16	64
1	7	6	29	23	23	203
					41	269

$$\theta = [1 \quad 4]$$

$$\theta_0 = \theta_0 - \alpha \boxed{\frac{\partial J}{\partial \theta_0}} \qquad \sum (\hat{y}-y)x_0$$

$$\theta_1 = \theta_1 - \alpha \frac{\partial J}{\partial \theta_1} \qquad \sum (\hat{y}-y)x_1$$

$$\theta_0 = 1 - \alpha\, 41 = \qquad 1 - ,41 = ,59$$

$$\theta_1 = 4 - \alpha\, 269 \qquad 4 - 2,69 = 1,31$$

$$\alpha = 0,01$$

$$\theta = \begin{bmatrix} .59 & 1.31 \end{bmatrix}$$