Regression

MACHINE LEARNING

Pakarat Musikawan

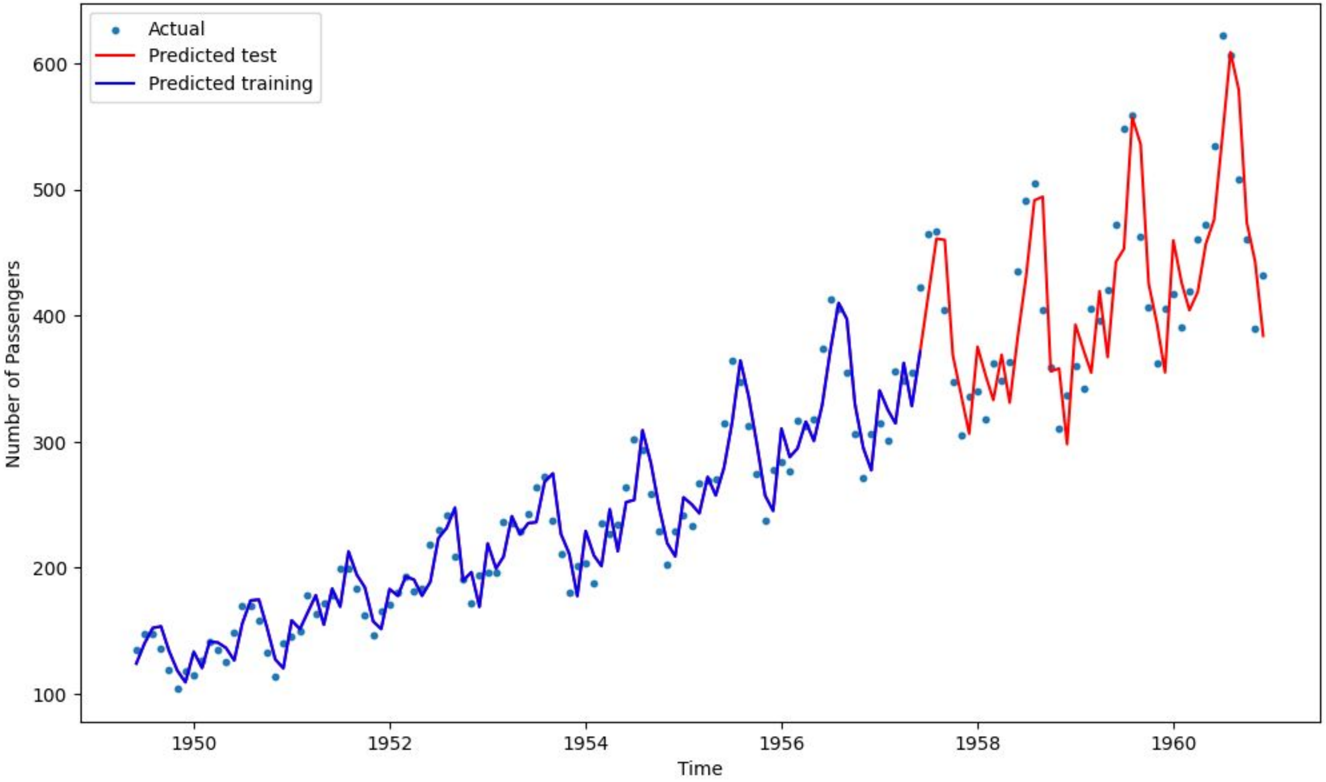
Regression

- Linear Regression
- Non-Linear Regression
- Ordinary Least Squares (OLS) Regression
- L1-Regularization (Lasso Regression)
- L2-Regularization (Ridge Regression)
- L1/L2-Regularization (Elastic Net)

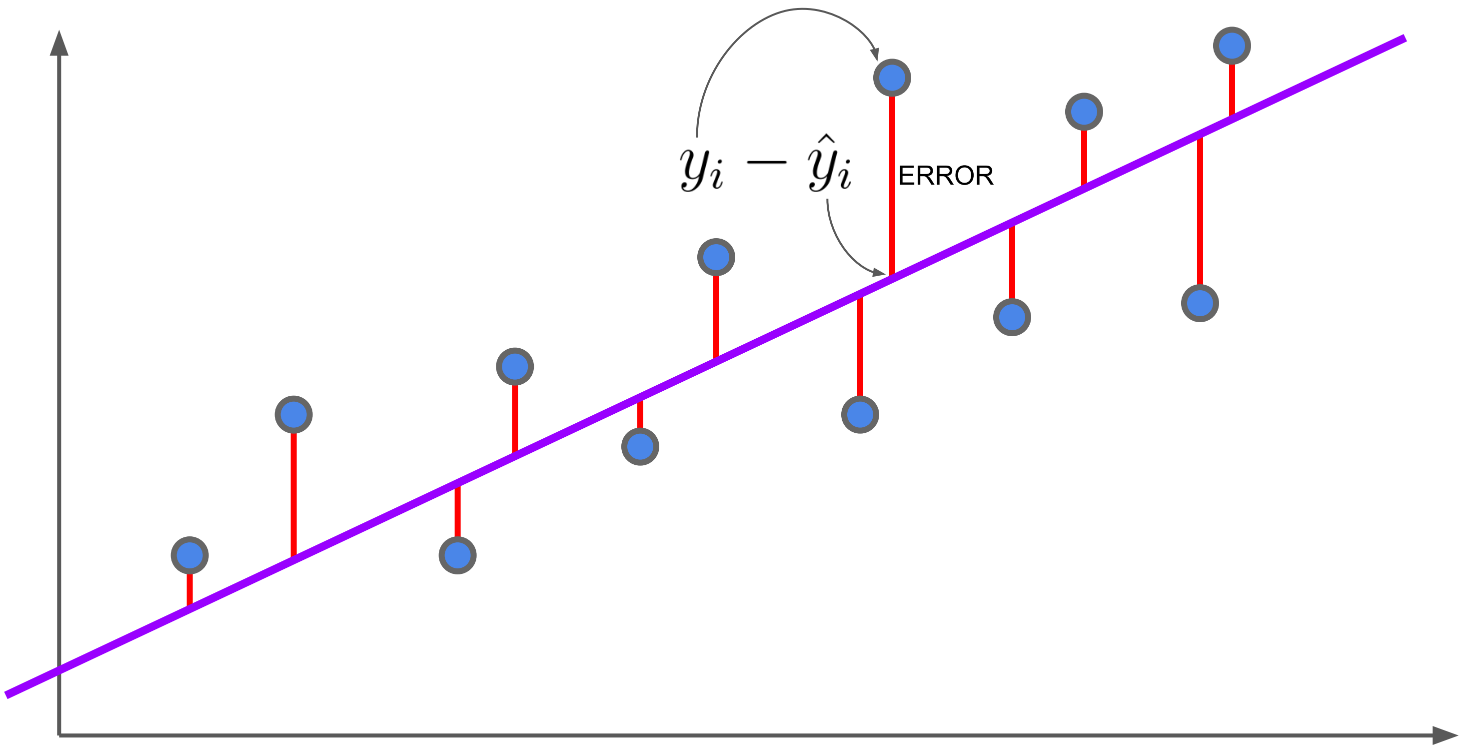
Regression Problem

longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
-114.31	34.19	15.0	5612.0	1283.0	1015.0	472.0	1.4936	66900.0
-114.47	34.40	19.0	7650.0	1901.0	1129.0	463.0	1.8200	80100.0
-114.56	33.69	17.0	720.0	174.0	333.0	117.0	1.6509	85700.0
-114.57	33.64	14.0	1501.0	337.0	515.0	226.0	3.1917	73400.0
-114.57	33.57	20.0	1454.0	326.0	624.0	262.0	1.9250	65500.0
...	...	...	...	...	...	...	...	...
-124.26	40.58	52.0	2217.0	394.0	907.0	369.0	2.3571	111400.0
-124.27	40.69	36.0	2349.0	528.0	1194.0	465.0	2.5179	79000.0
-124.30	41.84	17.0	2677.0	531.0	1244.0	456.0	3.0313	103600.0
-124.30	41.80	19.0	2672.0	552.0	1298.0	478.0	1.9797	85800.0
-124.35	40.54	52.0	1820.0	300.0	806.0	270.0	3.0147	94600.0

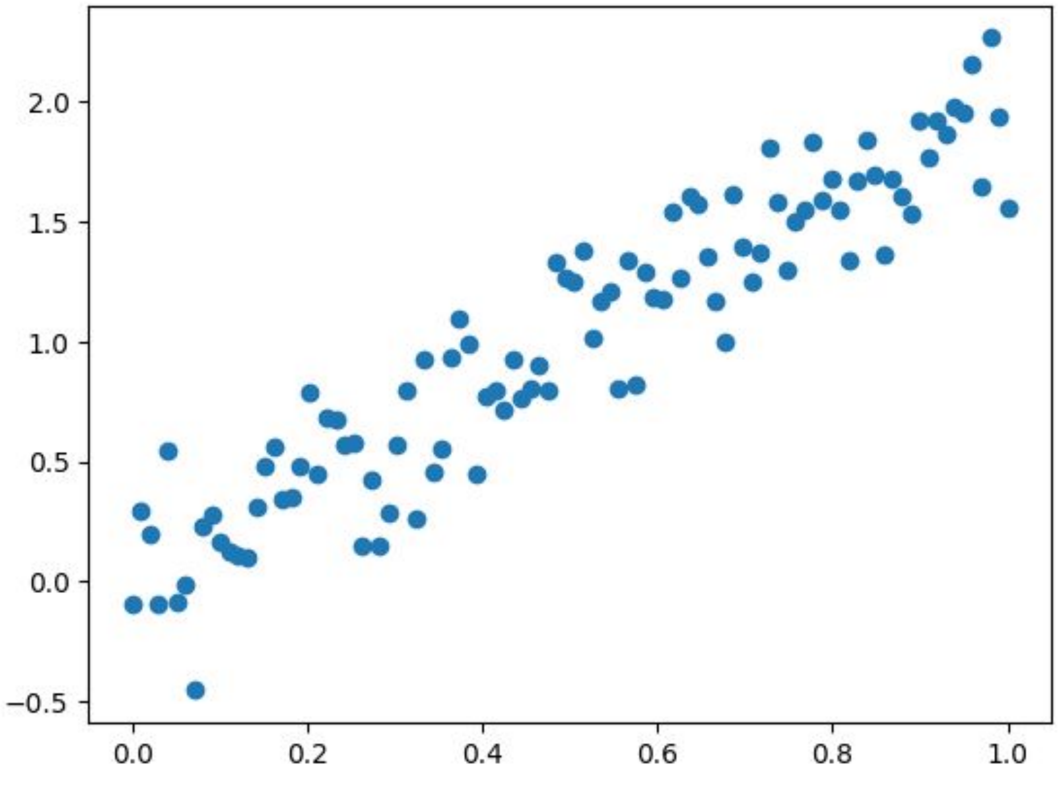
Regression Problem



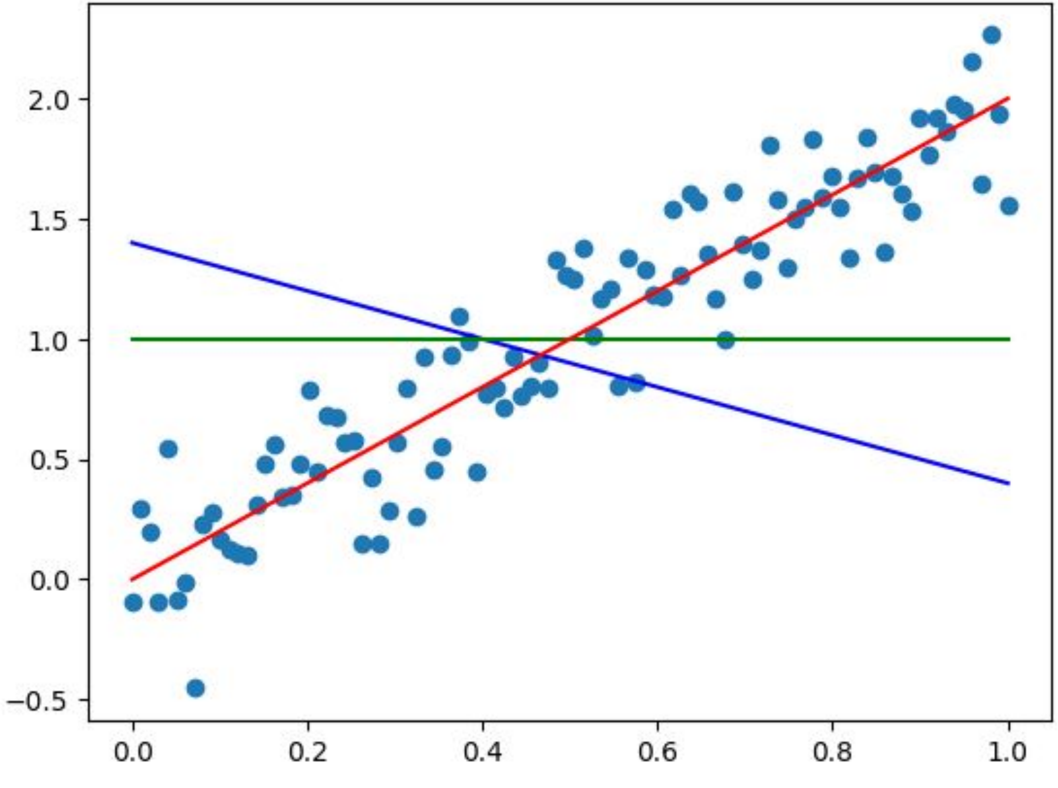
Regression Problem



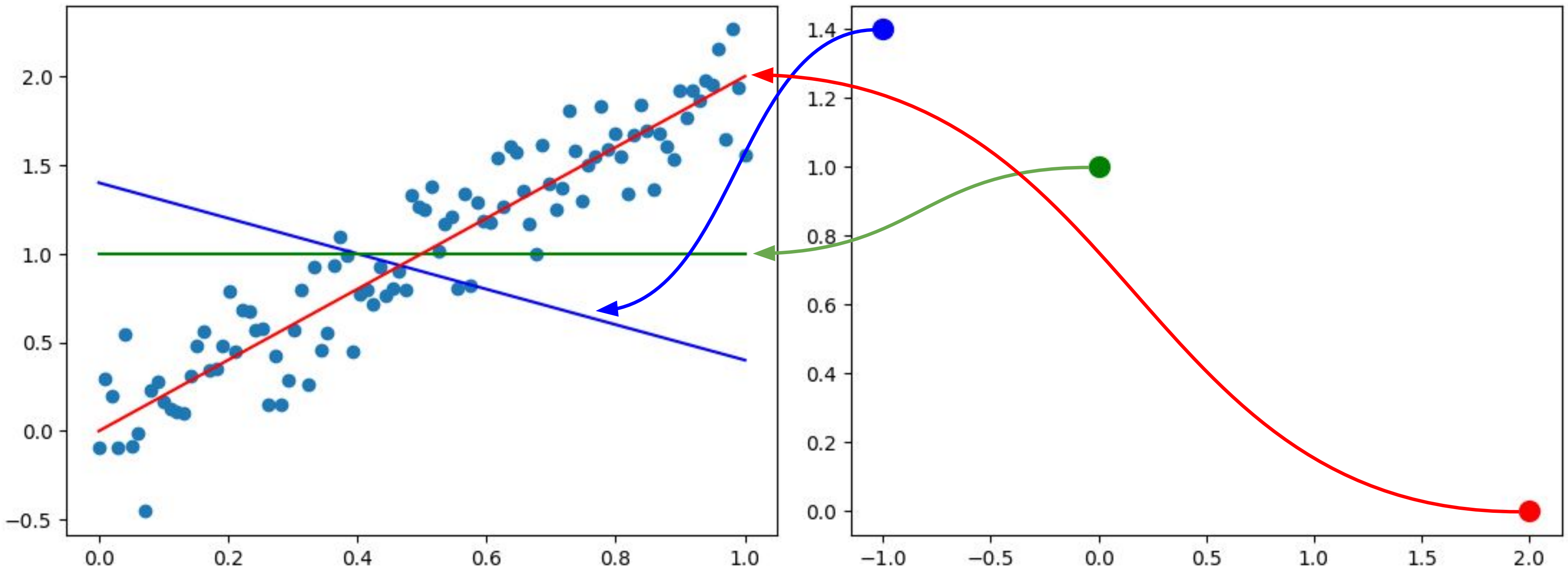
Linear Regression



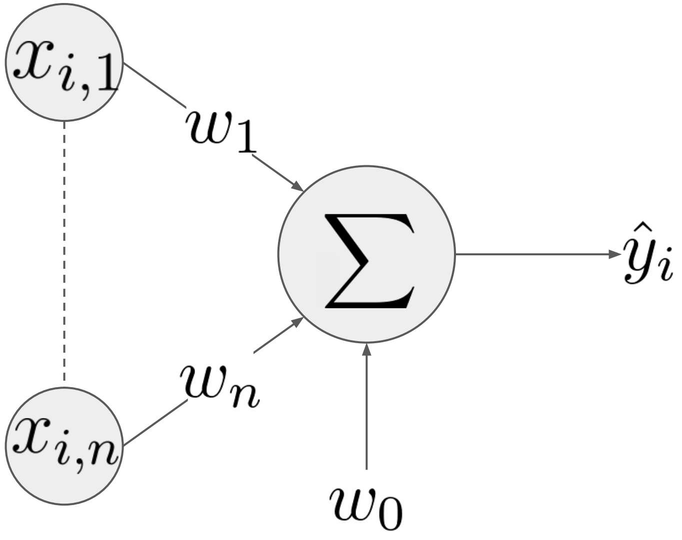
Linear Regression



Linear Regression



Linear Regression

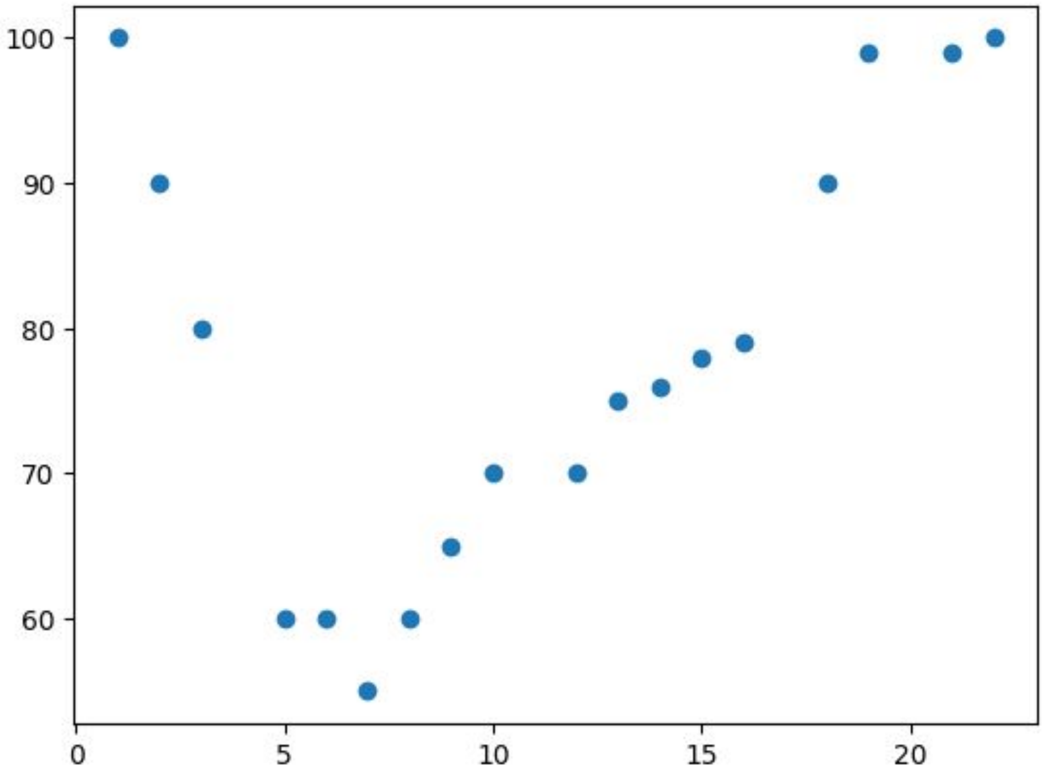


$$\hat{y}_i = w_0 + \sum_{j=1}^{n} w_j x_{i,j}$$

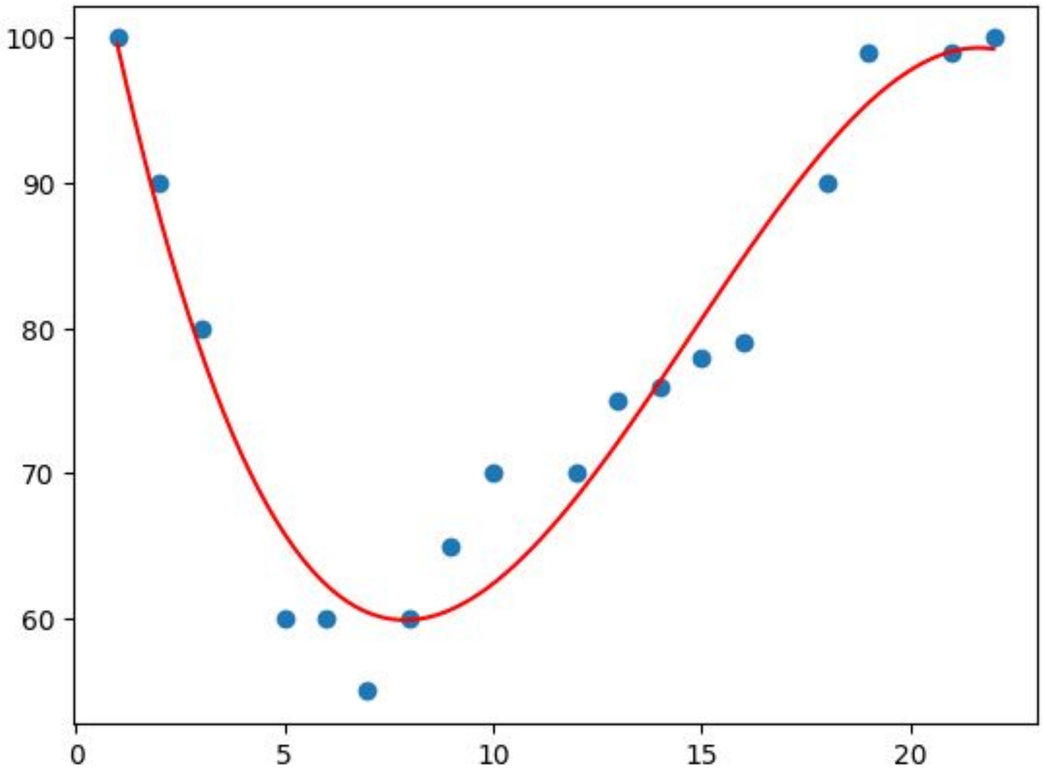
$$= w_0 + w_1 x_{i,1} + \cdots + w_n x_{i,n}$$

$$\boxed{\begin{aligned} \mathbf{X} &= \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \dots & x_{N,n} \end{bmatrix} \\ \boldsymbol{y} &= [y_1, \dots, y_N]^\top \\ \boldsymbol{w} &= [w_0, w_1, \dots, w_n] \end{aligned}}$$

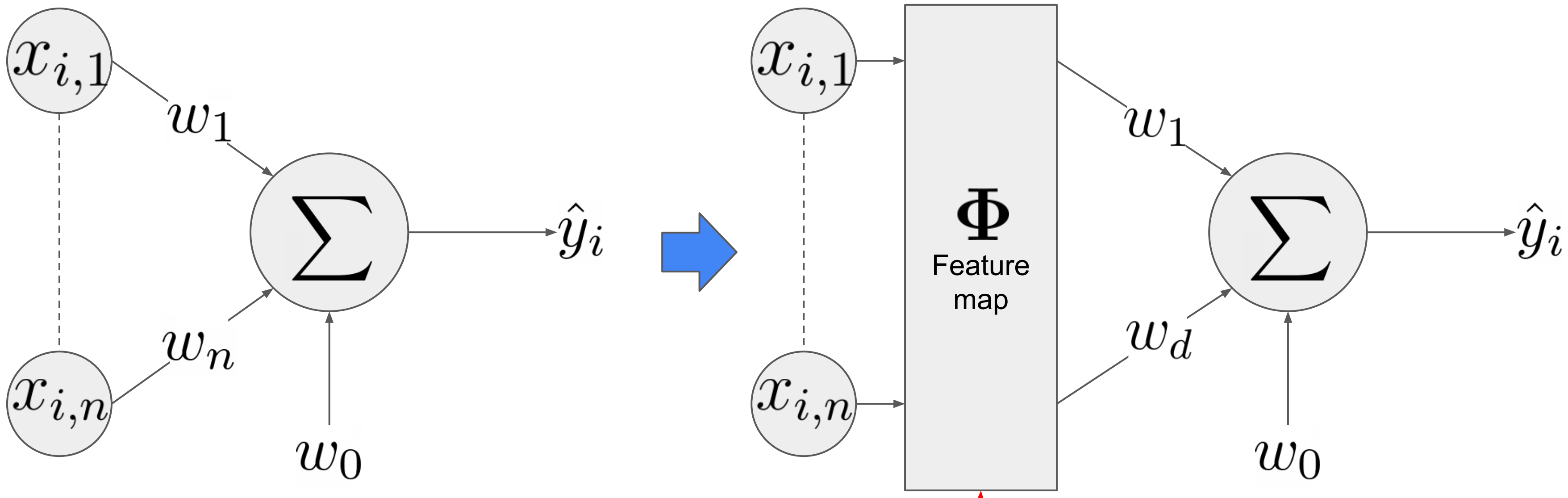
Non-Linear Regression



Non-Linear Regression

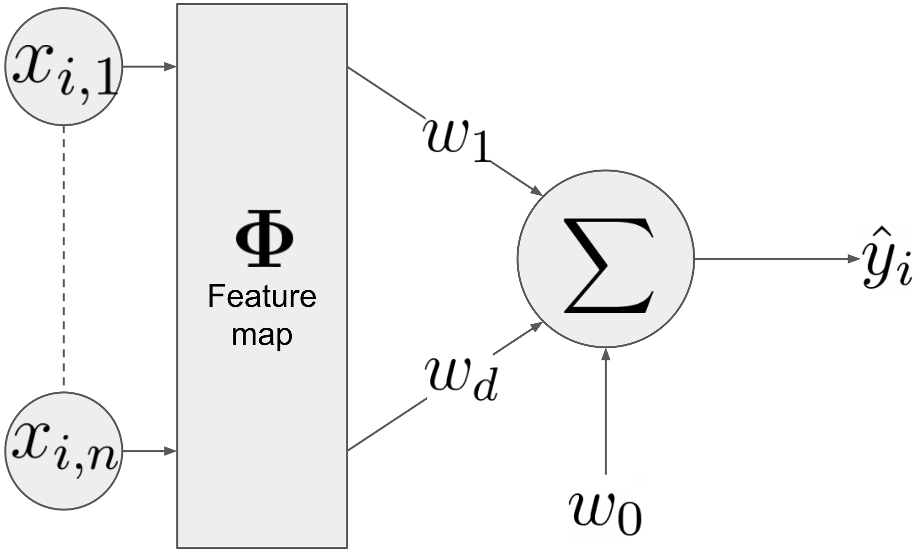


Non-Linear Regression



- Polynomial function
- Kernel function
- Neural network
- etc.

Non-Linear Regression



$$\hat{y}_i = w_0 + \sum_{j=1}^{d} w_j \phi_{i,j}$$

$$= w_0 + w_1 \phi_{i,1} + \cdots + w_n \phi_{i,d}$$

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \phi_{1,1} & \dots & \phi_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi_{N,1} & \dots & \phi_{N,d} \end{bmatrix}$$

$$\boldsymbol{y} = [y_1, \dots, y_N]^\top$$

$$\boldsymbol{w} = [w_0, w_1, \dots, w_d]$$

Non-Linear Regression - Polynomial

$$\boldsymbol{x}_i = [x_{i,1}, x_{i,2}]$$

$$\boldsymbol{\phi}_i = [1, \phi_{i,1}, \phi_{i,2}, \phi_{i,3}, \phi_{i,4}, \phi_{i,5}]$$

Dimension size of $\boldsymbol{\phi}_i$ is $\binom{n+o}{o} = \frac{(n+o)!}{o! \cdot n!}$

$$\binom{2+2}{2} = \frac{4!}{2! \cdot 2!} = \frac{24}{4} = 6$$

$$= [1, x_{i,1}, x_{i,2}, x_{i,1}^2, x_{i,1}x_{i,2}, x_{i,2}^2]$$

$$\boldsymbol{w} = [w_0, w_1, w_2, w_3, w_4, w_5]^\top$$

$$\hat{y}_i = \boldsymbol{\phi}_i \boldsymbol{w}$$

$$= w_0 + w_1\phi_{i,1} + w_2\phi_{i,2} + w_3\phi_{i,3} + w_4\phi_{i,4} + w_5\phi_{i,5}$$

$$= w_0 + w_1 x_{i,1} + w_2 x_{i,2} + w_3 x_{i,1}^2 + w_4 x_{i,1}x_{i,2} + w_5 x_{i,2}^2$$

Non-Linear Regression - Linear Kernel

$$
\begin{aligned}
\boldsymbol{\phi}_i &= [1, \phi_{i,1}, \phi_{i,2}, \ldots, \phi_{i,N}] \\
&= \left[1, \boldsymbol{x}_i^\top \boldsymbol{x}_1, \boldsymbol{x}_i^\top \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i^\top \boldsymbol{x}_N\right] \\
\boldsymbol{w} &= [w_0, w_1, w_2, \ldots, w_N]^\top \\
\hat{y}_i &= \boldsymbol{\phi}_i \boldsymbol{w} \\
&= w_0 + w_1 \phi_{i,1} + w_2 \phi_{i,2} + \ldots + w_N \phi_{i,N} \\
&= w_0 + w_1 \boldsymbol{x}_i^\top \boldsymbol{x}_1 + w_2 \boldsymbol{x}_i^\top \boldsymbol{x}_2 + \ldots + w_N \boldsymbol{x}_i^\top \boldsymbol{x}_N
\end{aligned}
$$

Regression Estimation & Regularization Technique

- Ordinary Least Squares (OLS) Regression
  - Closed-Form Solution
  - Iterative Method
- L1-Regularization (Lasso Regression)
- L2-Regularization (Ridge Regression)
- L1/L2-Regularization (Elastic Net)

Ordinary Least Squares (OLS) Regression

$$\boxed{\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \nabla_{\boldsymbol{w}} \mathcal{L}$$

$$w_j \leftarrow w_j - \eta \frac{\partial \mathcal{L}}{\partial w_j}$$

$$\leftarrow w_j + \eta \sum_{i=1}^{N} (y_i - \hat{y}_i) x_{i,j}$$

Closed-Form Solution

Ordinary Least Squares (OLS) Regression

$$\mathbf{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}, \boldsymbol{x}_i \in \mathbb{R}^n$$

$$\boldsymbol{y} = \{y_i\}_{i=1}^{N}, y_i \in \mathbb{R}$$

$$\boldsymbol{x}_i = [x_1, x_2, \ldots, x_n]$$

$$\boxed{\mathbf{I} = \Omega^{-1}\Omega}$$

$$\boxed{\mathbf{X}\mathbf{I} = \mathbf{X}}$$

$$\mathbf{X}\boldsymbol{w} = \boldsymbol{y}$$

$$\mathbf{X}^\top\mathbf{X}\boldsymbol{w} = \mathbf{X}^\top\boldsymbol{y}$$

$$(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{w} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}$$

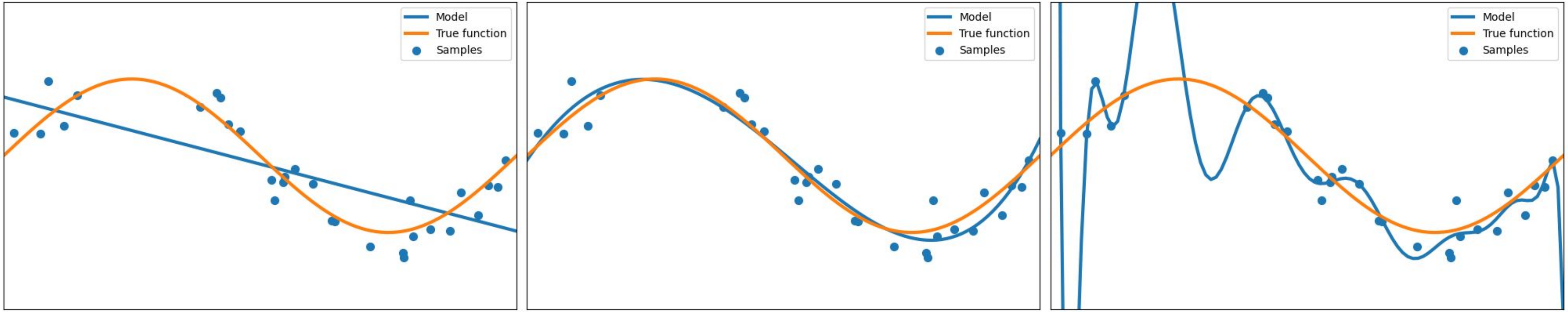
$$\mathbf{I}\boldsymbol{w} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}$$

$$\boldsymbol{w} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}$$

Ordinary Least Squares (OLS) Regression

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \\
&= \frac{1}{2} \left[\left(\boldsymbol{y} - \mathbf{X}\boldsymbol{w}\right)^\top \left(\boldsymbol{y} - \mathbf{X}\boldsymbol{w}\right) \right] \\
&= \frac{1}{2} \left[\boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \mathbf{X}\boldsymbol{w} - \boldsymbol{w}^\top \mathbf{X}^\top \boldsymbol{y} + \boldsymbol{w}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{w} \right] \\
\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} &\Rightarrow -\mathbf{X}^\top \boldsymbol{y} + \mathbf{X}^\top \mathbf{X}\boldsymbol{w} = 0 \\
&\Rightarrow \mathbf{X}^\top \mathbf{X}\boldsymbol{w} = \mathbf{X}^\top \boldsymbol{y} \\
&\Rightarrow \boldsymbol{w} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{y}
\end{aligned}
$$

Underfitting vs Overfitting



Underfitting

Overfitting

Iterative Method

L2-Regularization (Ridge Regression)

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \frac{\lambda}{2}\sum_{j=1}^{n} w_j^2$$

$$A = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$\frac{\partial A}{\partial w_j} = -\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i,j}$$

$$B = \frac{\lambda}{2}\sum_{j=1}^{n} w_j^2$$

$$\frac{\partial B}{\partial w_j} = \lambda w_j$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \nabla_{\boldsymbol{w}} \mathcal{L}$$

$$w_j \leftarrow w_j - \eta \frac{\partial \mathcal{L}}{\partial w_j}$$

$$\leftarrow w_j + \eta \left(\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i,j} - \lambda w_j \right)$$

Closed-Form Solution

L2-Regularization (Ridge Regression)

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \, \frac{\lambda}{2}\sum_{j=1}^{n} w_j^2$$

$$= \frac{1}{2}\left[(\boldsymbol{y} - \mathbf{X}\boldsymbol{w})^\top (\boldsymbol{y} - \mathbf{X}\boldsymbol{w})\right] + \frac{\lambda}{2}\|\boldsymbol{w}\|^2$$

$$= \frac{1}{2}\left[\boldsymbol{y}^\top\boldsymbol{y} - \boldsymbol{y}^\top\mathbf{X}\boldsymbol{w} - \boldsymbol{w}^\top\mathbf{X}^\top\boldsymbol{y} + \boldsymbol{w}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{w}\right] + \frac{\lambda}{2}\boldsymbol{w}^\top\boldsymbol{w}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{w}} \Rightarrow -\mathbf{X}^\top\boldsymbol{y} + \mathbf{X}^\top\mathbf{X}\boldsymbol{w} + \lambda\boldsymbol{w} = 0$$

$$\Rightarrow \mathbf{X}^\top\mathbf{X}\boldsymbol{w} + \lambda\boldsymbol{w} = \mathbf{X}^\top\boldsymbol{y}$$

$$\Rightarrow \left(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}\right)\boldsymbol{w} = \mathbf{X}^\top\boldsymbol{y}$$

$$\Rightarrow \boldsymbol{w} = \left(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^\top\boldsymbol{y}$$

Iterative Method

L1-Regularization (Lasso Regression)

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n}|w_j|$$

$$A = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$\frac{\partial A}{\partial w_j} = -\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i,j}$$

$$B = \lambda \sum_{j=1}^{n}|w_j|$$

$$\frac{\partial B}{\partial w_j} = \lambda \operatorname{sgn}(w_j)$$

sgn: Sign function

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \nabla_{\boldsymbol{w}}\mathcal{L}$$

$$w_j \leftarrow w_j - \eta \frac{\partial \mathcal{L}}{\partial w_j}$$

$$\leftarrow w_j + \eta \left(\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i,j} - \lambda \operatorname{sgn}(w_j) \right)$$

L1/L2-Regularization (Elastic Net)

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{n} w_j^2 + \alpha\sum_{j=1}^{n}|w_j|\right)$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta\nabla_{\boldsymbol{w}}\mathcal{L}$$

$$w_j \leftarrow w_j - \eta\frac{\partial\mathcal{L}}{\partial w_j}$$

$$\leftarrow w_j + \eta\left(\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i,j} - \lambda\left((1-\alpha)w_j + \alpha\,\mathrm{sgn}(w_j)\right)\right)$$

Workshop

ให้ทำการคำนวณเพื่อหาผลลัพธ์ของ Linear Regression โดยกำหนดให้

$$
\begin{aligned}
\boldsymbol{x}_i &= [x_{i,1}, x_{i,2}] \\
&= [0.25, 0.75] \\
\boldsymbol{w} &= [w_0, w_1, w_2]^\top \\
&= [0.25, 0.5, 0.75]^\top
\end{aligned}
$$

Workshop

ให้ทำการคำนวณเพื่อหาผลลัพธ์ของ Polynomial Regression โดยกำหนดให้

$$
\begin{aligned}
\boldsymbol{x}_i &= [x_{i,1}, x_{i,2}] \\
&= [0.25, 0.75] \\
\boldsymbol{\phi}_i &= [1, \phi_{i,1}, \phi_{i,2}, \phi_{i,3}, \phi_{i,4}, \phi_{i,5}] \\
&= \left[1, x_{i,1}, x_{i,2}, x_{i,1}^2, x_{i,1}x_{i,2}, x_{i,2}^2\right] \\
\boldsymbol{w} &= [w_0, w_1, w_2, w_3, w_4, w_5]^\top \\
&= [0.25, 0.35, 0.45, 0.55, 0.65, 0.75]^\top
\end{aligned}
$$