

# 인공지능개론

## 기계학습

# *Review: regression*

---

## Linear regression

$$h_{linear}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = \theta^T x$$

$$\theta^T x = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}^T \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

## Polynomial regression

$$h_{poly}(x) = \theta_0 x^0 + \theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n = \theta^T x$$

consider the feature vecgtor  $z_n = x^n$ , then

$$h_{poly}(x) = h_{linear}(z)$$

## Logistic regression

$$h_{log}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

# Review: regression

Consider

$$value = [10 \quad 8 \quad 9 \quad 13 \quad 15 \quad 18 \quad 14 \quad 20 \quad 24]^T$$

$$time = [1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9]^T$$

## Prediction (linear)

$$h_{linear}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = \theta^T x$$

$$h(x) = \theta_0 + \theta_1 t$$

$$\theta_0 + \theta_1 = 10$$

$$\theta_0 + 2\theta_1 = 8$$

:

$$\theta_0 + 9\theta_1 = 24$$



$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 9 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \\ \vdots \\ 24 \end{bmatrix}$$

$$\theta = (X^T X)^{-1} X^T y$$

$$h(x) = 7.8611 + 1.1167t$$

# Review: regression

Consider

$$value = [10 \ 8 \ 9 \ 13 \ 15 \ 18 \ 14 \ 20 \ 24]^T$$

$$time = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9]^T$$

## Prediction (polynomial)

$$h_{poly}(x) = \theta_0 x^0 + \theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n = \theta^T x$$

$$h(x) = \theta_0 + \theta_1 t + \theta_2 t^2$$

$$\theta_0 + \theta_1 + \theta_2 = 10$$

$$\theta_0 + 2\theta_1 + 4\theta_2 = 8$$

:

$$\theta_0 + 9\theta_1 + 81\theta_2 = 24$$

$$\begin{array}{c} \theta_0 + \theta_1 + \theta_2 = 10 \\ \theta_0 + 2\theta_1 + 4\theta_2 = 8 \\ \vdots \\ \theta_0 + 9\theta_1 + 81\theta_2 = 24 \end{array} \quad \rightarrow \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \vdots & \vdots & \vdots \\ 1 & 9 & 81 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \\ \vdots \\ 24 \end{bmatrix} \quad \theta = (X^T X)^{-1} X^T y$$

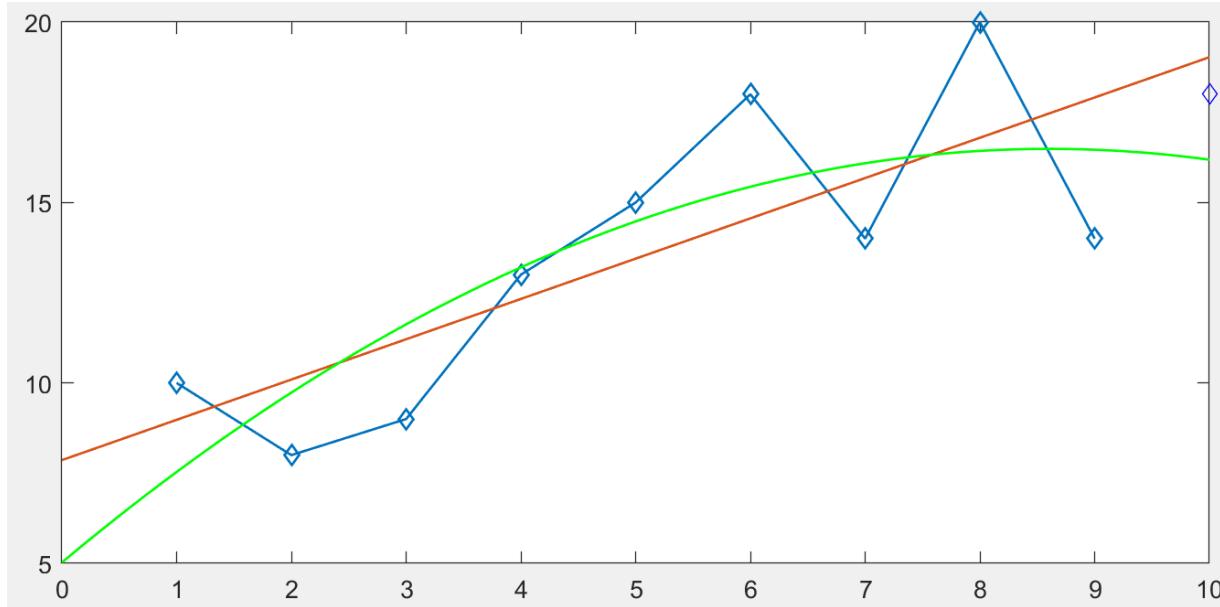
$$h(x) = 5.0238 + 2.6643t - 0.1548t^2$$

# *Review: regression*

Past

$$value = [10 \ 8 \ 9 \ 13 \ 15 \ 18 \ 14 \ 20 \ 24]^T$$
$$time = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9]^T$$

Today

$$value = [18]^T$$
$$time = [10]^T$$


How about tomorrow?

# Review: regression

Consider additional feature

$$value = [10 \ 8 \ 9 \ 13 \ 15 \ 18 \ 14 \ 20 \ 24]^T$$

$$time = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9]^T$$

$$index = [27 \ 29 \ 30 \ 30 \ 31 \ 28 \ 26 \ 22 \ 24]^T$$

## Prediction (linear)

$$h_{\text{linear}}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = \theta^T x$$

$$h(x) = \theta_0 + \theta_1 t + \theta_2 i$$

$$\theta_0 + \theta_1 + 27\theta_2 = 10$$

$$\theta_0 + 2\theta_1 + 29\theta_2 = 8$$

:

$$\theta_0 + 9\theta_1 + 24\theta_2 = 24$$

$$\begin{matrix} & \xrightarrow{\hspace{1cm}} & \begin{bmatrix} 1 & 1 & 27 \\ 1 & 2 & 29 \\ \vdots & \vdots & \vdots \\ 1 & 9 & 24 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \\ \vdots \\ 24 \end{bmatrix} & \theta = (X^T X)^{-1} X^T y \end{matrix}$$

$$h(x) = 9.1808 + 1.0862t - 0.0425i$$

## Prediction (polynomial)

$$h(x) = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 i + \theta_4 i^2$$

# Review: regression

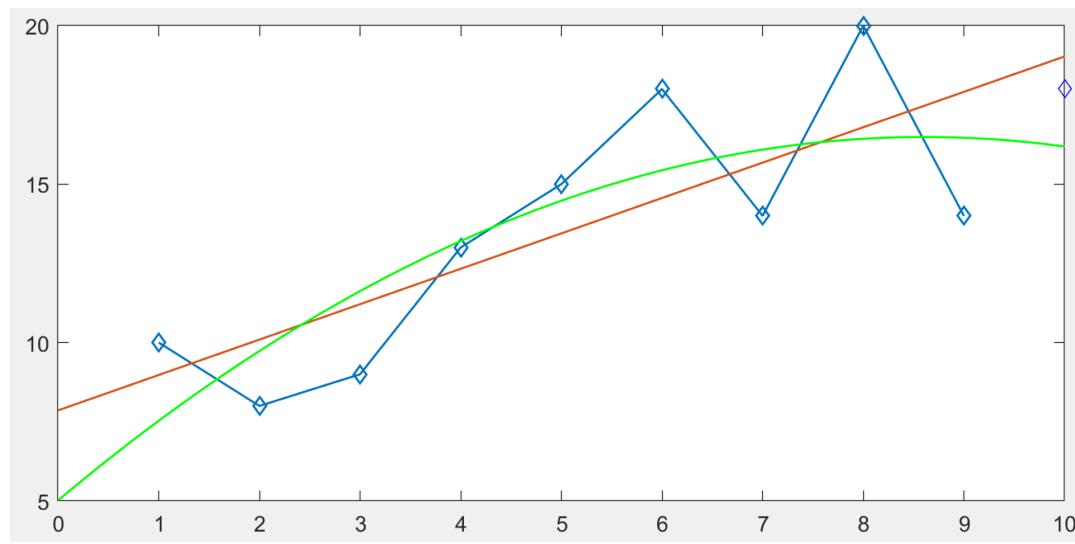
$$value = [10 \ 8 \ 9 \ 13 \ 15 \ 18 \ 14 \ 20 \ 24]^T$$

$$time = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9]^T$$

$$decision = [0 \ 1 \ 1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1]^T$$

0 = sell  
1 = buy

## Classification (linear)



$$h_{log}(x) = \frac{1}{1 + e^{-g(\theta, x)}}$$

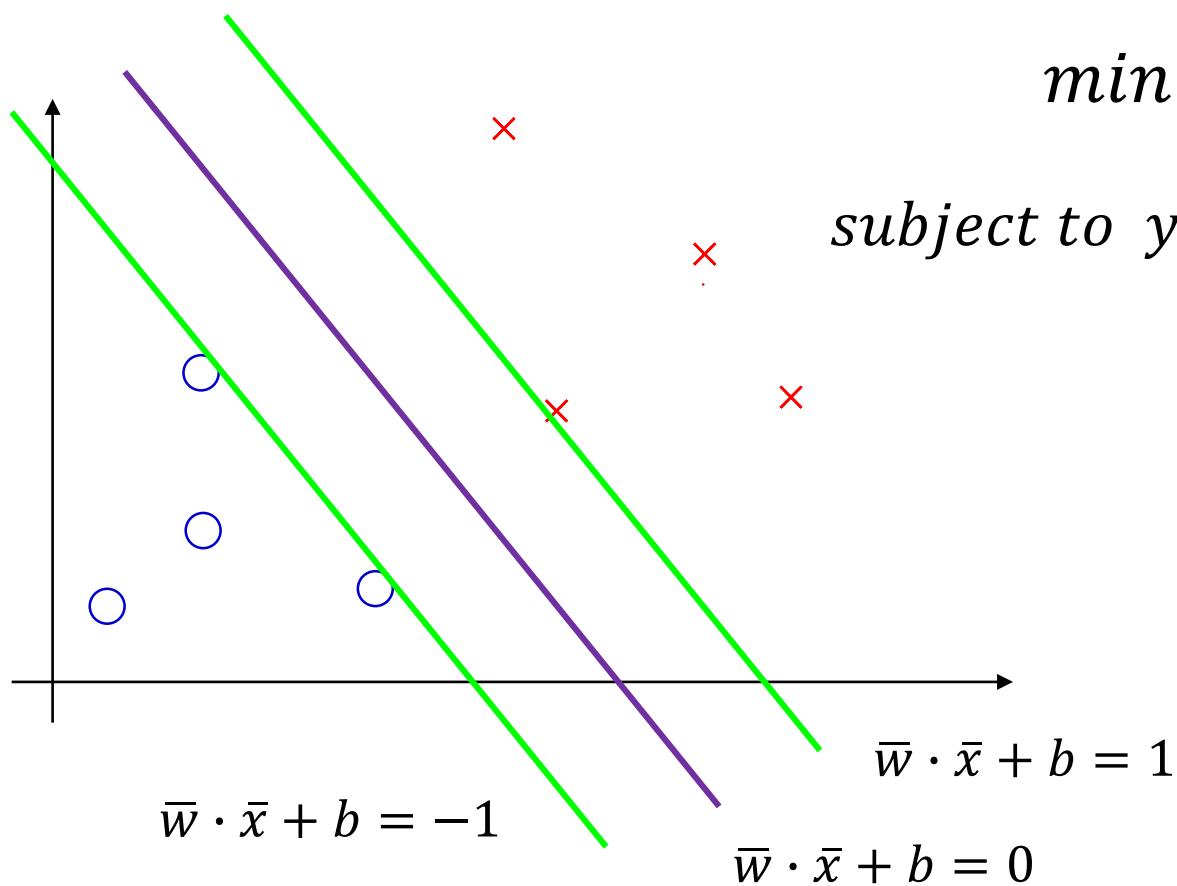
## Decision boundary

$$g(\theta, x) = \theta_0 + \theta_1 t + \theta_2 v$$

or

$$g(\theta, x) = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 v + \theta_4 v^2$$

# Review: SVM



$$\min \frac{1}{2} \|w\|^2$$

subject to  $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \geq 0$

# Review: SVM

$$\min \frac{1}{2} \|w\|^2$$

subject to  $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \geq 0$



$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum \alpha_i [y_i(\bar{w} \cdot \bar{x}_i + b) - 1]$$

subject to  $\alpha_i \geq 0$

$$\sum \alpha_i y_i = 0 \quad w = \sum \alpha_i y_i x_i$$



QP problem

$$\max_{\alpha} L = \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j x_i x_j \quad \sum \alpha_i y_i = 0 \quad \alpha_i \geq 0$$

# Review: SVM

Conditions for optimal solution of Lagrangian dual problem ( $w, b, \alpha$ )

## KKT (Karush-Kuhn-Tucker) conditions

### 1. Stationarity

$$\frac{\partial L}{\partial \bar{w}} = 0 \quad \Rightarrow \quad w = \sum \alpha_i y_i x_i \quad \frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum \alpha_i y_i = 0$$

### 2. Primal feasibility

$$y_i(\bar{w} \cdot \bar{x}_i + b) \geq 1$$

### 3. Dual feasibility

$$\alpha_i \geq 0$$

### 4. Complementary slackness

$$\alpha_i(y_i(\bar{w} \cdot \bar{x}_i + b) - 1) = 0$$

# Review: SVM

---

$$\alpha_i(y_i(\bar{w} \cdot \bar{x}_i + b) - 1) = 0$$

1.  $\alpha_i > 0$  and  $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0$

$\bar{x}_i$  are on the boundary => **Support vector**

2.  $\alpha_i = 0$  and  $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \neq 0$

$\bar{x}_i$  are not on the boundary

Only the support vectors construct Hyperplane

$$w = \sum \alpha_i y_i x_i \quad \rightarrow \quad w^* = \sum_{i \in S} \alpha_i^* y_i x_i$$

$$\begin{aligned} w^* \cdot x_S + b^* &= y_S \\ &= \sum_{i \in S} \alpha_i^* y_i x_i x_S + b^* \end{aligned}$$

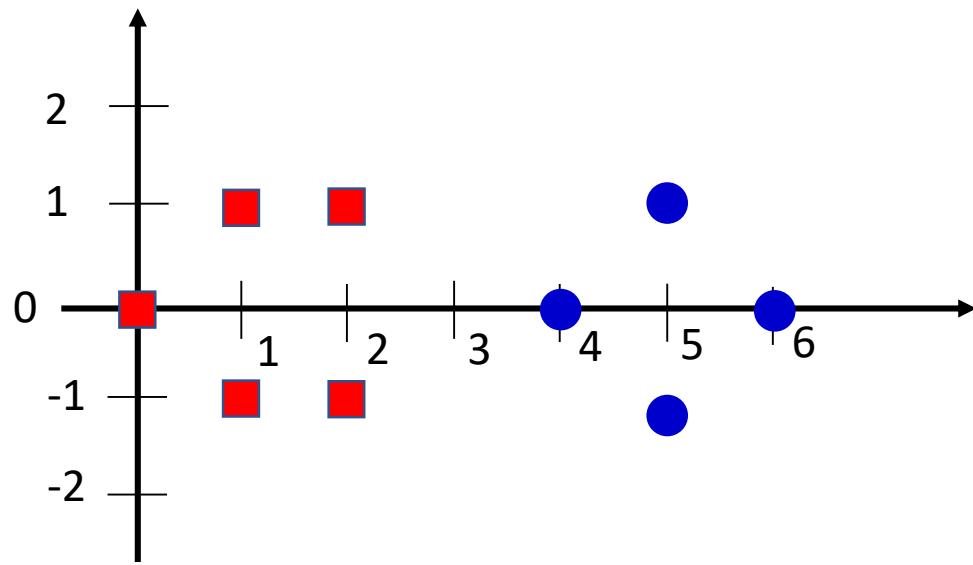
# Review: SVM

$$value = [4 \ 6 \ 1 \ 1 \ 2 \ 5 \ 2 \ 5 \ 0]^T$$

$$time = [0 \ 0 \ 1 \ -1 \ 1 \ 1 \ -1 \ -1 \ 0]^T$$

$$class = [1 \ 1 \ -1 \ -1 \ -1 \ 1 \ -1 \ 1 \ -1]^T$$

1 (Positive class)  
-1 (Negative class)



$$v_1 = [2 \ 1]^T$$

$$v_2 = [2 \ -1]^T$$

$$v_3 = [4 \ 0]^T$$

# Review: SVM

---

On the decision boundary

$$\bar{w} \cdot \bar{x}_i + b = -1$$

$$\bar{w} \cdot \bar{x}_i + b = 1$$

$$2w_1 + w_2 + b = -1$$

$$4w_1 + 0w_2 + b = 1$$

$$2w_1 - w_2 + b = -1$$

However, we need to calculate

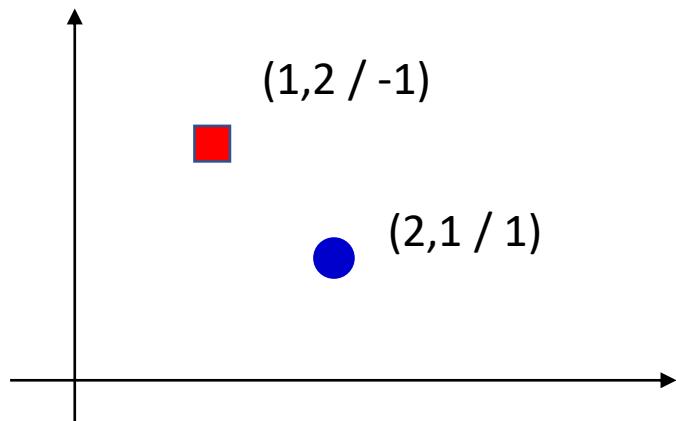
$$\max_{\alpha} L = \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j x_i x_j$$

$$\sum \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

# Review: SVM

However, we need to calculate



$$\begin{aligned} & \alpha_1 + \alpha_2 - \frac{1}{2} (\alpha_1 \alpha_1 \cdot 1 \cdot 1 \cdot \left\langle \binom{2}{1}, \binom{2}{1} \right\rangle \\ & + 2 \cdot \alpha_1 \alpha_2 \cdot 1 \cdot (-1) \cdot \left\langle \binom{1}{2}, \binom{2}{1} \right\rangle + \\ & + \alpha_2 \alpha_2 \cdot (-1) \cdot (-1) \cdot \left\langle \binom{1}{2}, \binom{1}{2} \right\rangle) \\ & = \alpha_1 + \alpha_2 - \frac{1}{2} (5\alpha_1^2 - 8\alpha_1\alpha_2 + 5\alpha_2^2) \end{aligned}$$

$$\alpha_1 y_1 + \alpha_2 y_2 = 0$$

$$\alpha_1 \geq 0$$

$$\alpha_2 \geq 0$$

$$\alpha_1 = \alpha_2 = 1$$

$$\max_{\alpha} L = \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j x_i x_j$$

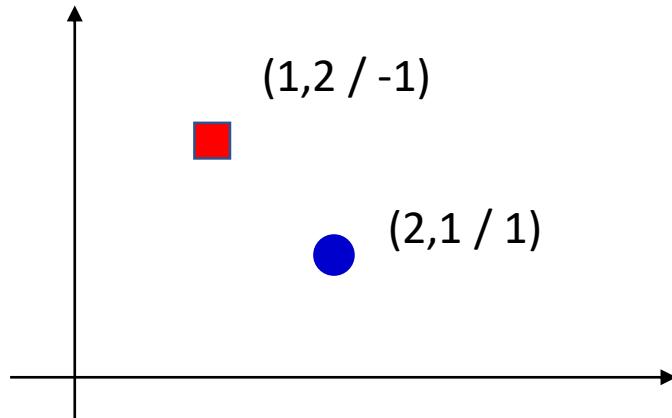
$$\sum \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

# Review: SVM

---

However, we need to calculate



$$\begin{aligned} w^* &= \sum \alpha_i^* y_i x_i = 1 \cdot 1 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 1 \cdot (-1) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{aligned}$$

$$b^* = y_S - \sum \alpha_i^* y_i x_i x_S = 1 - \left( 1 \cdot 1 \cdot \langle \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \rangle + 1 \cdot (-1) \cdot \langle \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \rangle \right) = 0$$

---

# Review: SVM

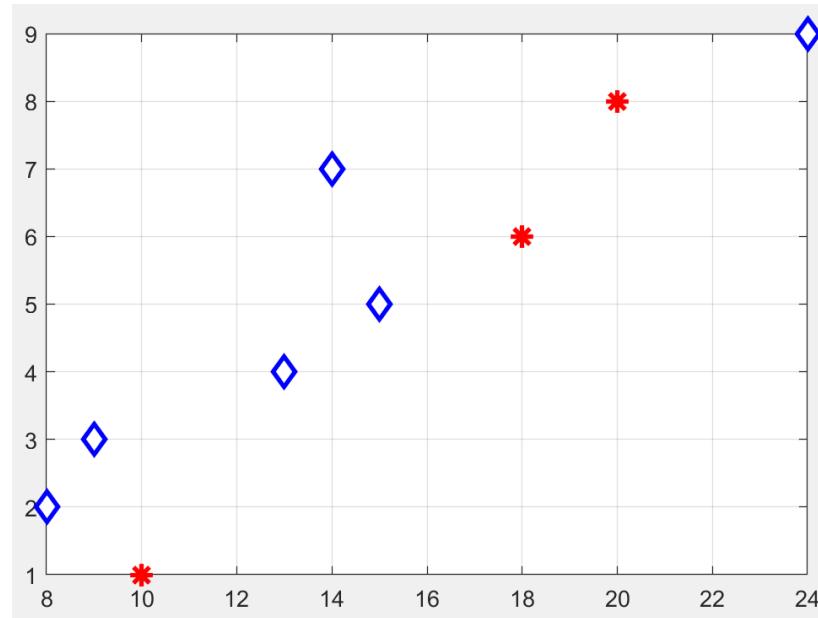
$$value = [10 \ 8 \ 9 \ 13 \ 15 \ 18 \ 14 \ 20 \ 24]^T$$

$$time = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9]^T$$

$$decision = [0 \ 1 \ 1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1]^T$$

0 = sell (Negative class)

1 = buy (Positive class)



# Review: SVM

## ❖ 커널 트릭을 사용할 때의 최적화 문제

$$\tilde{L}(\alpha) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j \Phi(x_i) \cdot \Phi(x_j) + \sum_{i=1}^N \alpha_i$$

$$h(x) = \sum_{i=1}^N a_i t_i \Phi(x_i) \cdot \Phi(x) + b$$

### ▪ 커널 함수 적용

$$\tilde{L}(\alpha) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j K(x_i, x_j) + \sum_{i=1}^N \alpha_i$$

$$h(x) = \sum_{i=1}^N a_i t_i K(x_i, x) + b$$