[1] (5점) Set {(1, 1,0), (1, 0, 1), (2, 1, 1)} 가 span하는 vector space는 몇차원 공간인가?

답.

$$\begin{bmatrix}1 & 1 & 2\\ 1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix} \xrightarrow[e.r.o]{} \begin{bmatrix}1 & 1 & 2\\ 0 & 1 & 1\\ 0 & 1 & 1\end{bmatrix} \xrightarrow[e.r.o]{} \begin{bmatrix}1 & 1 & 2\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix} \xrightarrow[e.r.o]{} \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$$

따라서 {(1, 1,0), (1, 0, 1)} 만 linearly independent 하므로 이 집합으로 span 되는 공간은 2차원 공간이다.

[2] (5점) $X \coloneqq [x_1 \quad x_2 \quad x_3]$, x_i 는 2차원 column 벡터이다. X^TX과 XX^T 를 계산하고 각각의 차원(m × n형태)을 구하시요.

답.

$$X^TX = \begin{bmatrix}x_1^T\\ x_2^T\\ x_3^T\end{bmatrix}[x_1 \quad x_2 \quad x_3] = \begin{bmatrix}x_1^Tx_1 & x_1^Tx_2 & x_1^Tx_3\\ x_2^Tx_1 & x_2^Tx_2 & x_2^Tx_3\\ x_3^Tx_1 & x_3^Tx_2 & x_3^Tx_3\end{bmatrix} \rightarrow 3 \times 3 \quad \text{matrix}$$

$$XX^T = [x_1 \quad x_2 \quad x_3]\begin{bmatrix}x_1^T\\ x_2^T\\ x_3^T\end{bmatrix} = x_1x_1^T + x_2x_2^T + x_3x_3^T \rightarrow 2 \times 2 \quad \text{matrix}$$

[3] (5점) $A = \begin{bmatrix}1 & 2 & -1\\ -2 & -4 & 2\end{bmatrix}$ 의 $C(A)$, $N(A)$ 를 구하고 각 벡터공간의 차원과 basis를 제시하시오.

답.

$C(A) = \left\{c\begin{bmatrix}1\\ -2\end{bmatrix} | c \in R\right\}$, dimension = 1, basis = $\left\{\begin{bmatrix}1\\ -2\end{bmatrix}\right\}$

$N(A) = \left\{\begin{bmatrix}x\\ y\\ z\end{bmatrix} | x + 2y - z = 0\right\} = \left\{s\begin{bmatrix}-2\\ 1\\ 0\end{bmatrix} + t\begin{bmatrix}1\\ 0\\ 1\end{bmatrix} | s,t \in R\right\}$, dim. = 2, basis = $\left\{\begin{bmatrix}-2\\ 1\\ 0\end{bmatrix}, \begin{bmatrix}1\\ 0\\ 1\end{bmatrix}\right\}$

[4] (5 점) A가 m × n (m ≠ n) marix 일 때, Singular value decomposition 을 하면 $A = U\Sigma V^*$ 로 나타낼 수 있다. 여기서 Σ는 일부의 대각 요소에 singular value 가 위치하고 나머지 요소는 모두 0 인 m × n matrix 이다. Ax=b 이 해가 무수히 많을 때 $\|x\|$ 이 가장 적은 해를 구하시오.

답.

Psuedo inverse 를 구하면, $x = A^{\dagger}b = V\Sigma^{\dagger}U^*b$ 이 된다.

[5] (5점) Kullback – Leibler divergence $D_{p(X,Y)\|p(X)p(Y)} = 0$ 이면 Mutual information $I(X,Y)$의 값은 얼마인가? 왜 그렇게 되는지 이유를 쓰시오.
답.

$D_{p(X,Y)\|p(X)p(Y)} = 0$ 이면 $p(X,Y)$와 $p(X)p(Y)$의 분포가 같아서 X, Y 는 independent 하다. 따라서 $I(X,Y)$=0 이다.

[6] (5점) 다음의 확률 값이 주어져 있다.

$$P(X=0|Y=0) = \frac{2}{3},\ P(X=1|Y=0) = \frac{1}{3},$$
$$P(X=0|Y=1) = \frac{3}{8},\ P(X=1|Y=1) = \frac{5}{8},$$
$$P(Y=0) = 1/5,\ \ P(Y=1) = 4/5$$

$H(\mathrm{X},\mathrm{Y}) = -\sum_x \sum_y p(x,y) \log p(x,y)$를 구하시요. 계산은 하지 마세요..
답.

$$H(X,Y) = -\sum_x \sum_y p(x|y)p(y) \log p(x|y)p(y)$$
$$= -\frac{2}{3}\ \frac{1}{5} log \frac{2}{3}\ \frac{1}{5} - \frac{1}{3}\ \frac{1}{5} log \frac{1}{3}\ \frac{1}{5} - \frac{3}{8}\ \frac{4}{5} log \frac{3}{8}\ \frac{4}{5} - \frac{5}{8}\ \frac{4}{5} log \frac{5}{8}\ \frac{4}{5}$$

[7] (5점) 동전을 던졌을 때, 앞이 나오는 경우 확률변수 X는 0이되고, 뒤가 나오는 경우는 X가 1이된다. 앞이 나올 확률은 1/3, 뒤가 나올 확률은 2/3이다. 이 확률 변수 X가 가지는 entropy는 얼마인가?
답.

$$H(X) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$$

[8] (5점) $C_{p||q}(x;W) = -\sum_x \sum_k p_k(x)\ \log q_k(x;W)$에서 신경망의 입력 x 에 대해 신경망의 k 번째 노드의 출력이 $q_k(x;W)$ 이고 desired 출력이 $p_k(x)$ 이다. $C_{p||q}(x;W)$가 0이 되도록 W를 학습하게 되면 어떤 결과를 가져오는가?
답.

$q_k(x;W)$가 desired 출력 $p_k(x)$와 같아지게 된다.

[9] (각 5 점) 두 class 를 분류하는 classifier 는 $y = w^T x + b$로 주어진다. 즉, 입력 패턴 x 에 대해 y 가 양수이면 label 을 +1 로 부여하고, y 가 음수이면 label 을 -1 로 부여한다. 정확한 label 을 부여하도록 하기 위한 w 값은 n개의 훈련 데이터 $\{(x_i, y_i) | i = 1, \dots, n\}$ 를 이용하여 학습에 의해 구해진다. 이때, class 간의 margin(비무장지대)을 최대로 하는 Support Vector Machine(SVM)은 다음의 최적화문제를 풀어서 w를 구한다.

$$\underset{w}{\text{minimize}} \quad \frac{1}{2} w^T w$$
$$\text{subject to} \quad y_i(w^T x_i + b) \geq 1, \quad i = 1, \dots, n$$

A. Lagrangian $L(w, \alpha)$는 아래와 같이 주어지는데 괄호 안에 들어갈 수식은?
$L(w, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i ($ $)$

B. KKT 4 번째 조건인 $\nabla_w L(w, \alpha) = 0$ 을 만족하는 w를 구하시오.

C. KKT 3 번째 조건은 무엇인가? 이 조건은 support vector 가 아닌 패턴에 해당하는 α_i가 0 이 된다는 것을 의미한다. 왜 그런가?

D. B 에서 구한 w를 $L(w, \alpha)$에 대입하며 dual 함수 $L(\alpha) = -\frac{1}{2} \alpha^T Q \alpha + \alpha^T \mathbf{1}$를 구할 수 있고 이를 최대로 하는 α 를 구해서 B 에서 구한 식에 대입하여 최종 w를 구한다. 여기서 α 는 α_i를 요소로 하는 column vector 이고 $\mathbf{1}$은 모든 요소가 1 인 column vector 이다. $L(\alpha)$를 최대로 하는 α 를 구하시요.

답.

A. $y_i(w^T x_i + b) - 1$

B. $w = \sum_{i=1}^{n} \alpha_i y_i x_i$

C. $\alpha_i(y_i(w^T x_i + b) - 1) = 0$, 여기서 support vector 가 아닌 패턴에 대해서는 $y_i(w^T x_i + b) - 1 \neq 0$이다. 따라서 α_i 가 0 이 될 수밖에 없다.

D. $\nabla_\alpha L(\alpha) = 0 \rightarrow -\mathrm{Q}\alpha + \mathbf{1} = \mathbf{0} \rightarrow \alpha = Q^{-1}\mathbf{1}$

[10] (각 5 점) Softmax(σ)를 출력노드의 activation 함수로 쓰는 신경망의 loss 는 cross entropy 를 사용한다. 그리고 은닉층 노드의 activation 함수는 사용하지 않는다. 신경망의 출력 노드는 2 개이며, 은닉층은 하나만 있으며, 3 개의 노드를 가지고 있다. 입력층 노드는 2 개이다. 2 개의 입력값을 x_1, x_2, 출력층의 각 노드의 출력을 o_1, o_2, 은닉층 각 노드의 출력을 h_1, h_2, h_3 라고 표시한다. 그러면 이 신경망의 matrix-vector form 은 아래와 같다.

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} o_1 \\ o_2 \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix} \circ \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

A. Desired 출력을 t_1, t_2,라고 했을 때 error signal vector $\delta_o = \begin{bmatrix} \delta_{o1} \\ \delta_{o2} \end{bmatrix}$를 구하시요.

B. 은닉층 error signal vector $\delta_h = \begin{bmatrix} \delta_{h1} \\ \delta_{h2} \\ \delta_{h3} \end{bmatrix}$ 를 구하시요.

C, D. $W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}, V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{33} \end{bmatrix}$ 라고 했을 때, error backpropagation에 의한 W, V의 update는 아래 식으로 이루어진다.

$$W^{new} = W^{old} + \Delta W, \qquad V^{new} = V^{old} + \Delta \mathrm{V}.$$

$\Delta W, \Delta \mathrm{V}$를 learning rate η, $\begin{bmatrix} \delta_{o1} \\ \delta_{o2} \end{bmatrix}, \begin{bmatrix} \delta_{h1} \\ \delta_{h2} \\ \delta_{h3} \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 를 이용하여 구하시요. 단, 연산시 dimension이 맞도록 필요시column vector를 적절하게 row vector 로 바꾸어 사용하시요.

답.

A. $\delta_o = \begin{bmatrix} \delta_{o1} \\ \delta_{o2} \end{bmatrix} = \begin{bmatrix} t_1 - o_1 \\ t_2 - o_2 \end{bmatrix}$

B. $\delta_h = \begin{bmatrix} \delta_{h1} \\ \delta_{h2} \\ \delta_{h3} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \begin{bmatrix} \delta_{o1} \\ \delta_{o2} \end{bmatrix}$

C. $\Delta W = \eta \begin{bmatrix} \delta_{o1} \\ \delta_{o2} \end{bmatrix} [h_1 \quad h_2 \quad h_3]$,

D. $\Delta V = \eta \begin{bmatrix} \delta_{h1} \\ \delta_{h2} \\ \delta_{h3} \end{bmatrix} [x_1 \quad x_2]$,

[11] (5 점) n 개의 2-차원 벡터 샘플 집합 $\{\boldsymbol{x}_1, .., \boldsymbol{x}_n\}$ 의 평균 백터는 $m = \frac{1}{n}\sum_{k=1}^{n} x_k$ 로 주어진다. $\boldsymbol{x}_k - m$ 백터를 크기가 1 인 벡터 e가 span 하는 공간에 projection 하면 크기가 얼마가 되는가?

답: $|(\boldsymbol{x}_k - m)^T e|$, $(\boldsymbol{x}_k - m)^T e = e^T(\boldsymbol{x}_k - m)$ 도 정답처리.

[12] (5 점) 2 차원 공간의 데이터로부터 구한 평균 m 이 $\binom{1}{2}$라고 하자. 데이터의 Scatter matrix S 의 orthonormal eigenvector 가 $\binom{1/\sqrt{5}}{2/\sqrt{5}}$ $\binom{1/\sqrt{5}}{-2/\sqrt{5}}$ 이고 각각의 eigenvalue 가 2, 1 일 때, 벡터 $\binom{4}{3}$을 1 차원 값으로 projection 하였을 때 projection error 가 가장 적은 1-st principal component 를 구하시오.

답: $a_1 = \mathrm{e}_1^T(\boldsymbol{x}_k - m) = \begin{bmatrix}1/\sqrt{5} & 2/\sqrt{5}\end{bmatrix}\left(\begin{bmatrix}4\\3\end{bmatrix} - \begin{bmatrix}1\\2\end{bmatrix}\right) = \frac{5}{\sqrt{5}} = \sqrt{5}$

[13] (5 점) Simple linear regression model 이 아래와 같다.

$$Y = \theta_0 + \theta_1 X + \epsilon$$

Correlation coefficient 는 $\rho(Y, X) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{\sigma_y}\right)\left(\frac{x_i - \bar{x}}{\sigma_x}\right)$으로 주어진다. 이때 100 개의 pair data $\{((x_i, y_i) | i = 1, \dots, 100\}$ 로 regression 하였더니 $\theta_0 = 3$, $\theta_1 = -5$로 추정되었다. $\rho(Y, X)$이 부호는?

답: 음수, $\rho(Y, X)$는 θ_1과 부호가 같으므로 음수이다.

[14] (5 점) 일반적인 regression 문제는 아래와 같이 주어진다.

$$\boldsymbol{y} = \Phi\theta + \boldsymbol{\epsilon},$$

이를 Least Squares Estimation을 하면 아래 최적화 문제를 풀어야 한다.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \|\boldsymbol{\epsilon}\|^2 = \|\boldsymbol{y} - \Phi\theta\|^2 \cong S(\theta)$$

$\hat{\theta}$ 의 closed form solution을 유도하시오.

답:

$\nabla_\theta(\boldsymbol{y} - \Phi\theta)^T(\boldsymbol{y} - \Phi\theta) = 0$ at $\hat{\theta}$

$2\Phi^T(\boldsymbol{y} - \Phi\hat{\theta}) = 0$

$\Phi^T\boldsymbol{y} - \Phi^T\Phi\hat{\theta} = 0$

$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$

[15] (5 점) Regression model $y_i = \theta_0 + \theta_1 x_i + \epsilon_i,\ i = 1, \ldots, n$ 의 fitting 정확도를 측정하는 R^2을 구하는 식을 쓰시오.

답: $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)2}{\sum_{i=1}^{n}(y_i - \bar{y})2}$, 여기서 $\bar{y}$는 $\{y_i\}$의 평균이고, $\hat{y}_i$는 fitting 매개변수에 의해 추정된 y_i 값이다.

[16] (5 점) Elastic Regression 에서 목적함수에 추가되는 regularization 항은 무엇인가?

답: $\gamma_1 \|\theta\|_2^2 + \gamma_2 \|\theta\|_1$

[17] (각 5점) p 차원의 독립 랜덤 변수 벡터의 관측 값 n개를 column vector로 하는 행렬을 X라고 하면 $\mathrm{X} = [x_1 \quad x_2 \quad \cdots \quad x_n]$과 같다. 평균을 0이 되도록 조정된 값이다.

A. Data $\{x_1, \quad x_2, \quad \cdots, \quad x_n\}$ 의 Scatter matrix S를 X를 이용하여 표현하면?

답: $\mathrm{S} = \mathrm{X}X^T$

B. $p \times n$ 차원의 $\mathbf{X}$를 $q \times n$ 차원의 $\mathbf{Z}$벡터로 차원을 축소하고자 한다. 이를 위해 $\mathbf{Z} = \bar{\mathbf{U}}^T\mathbf{X}$를 이용하여 구한다. 다음 중 틀린 내용은?

(1) $\bar{\mathbf{U}}$ 의 column들은 S의 eigenvalue가 큰 순서대로 q개에 해당하는 orthonormal eigenvector 들이다.

(1) $\hat{\theta} = (\mathbf{Z}\boldsymbol{Z}^{\boldsymbol{T}})^{-1}\mathbf{Z}\boldsymbol{y}$ 이 된다.

(2) $\boldsymbol{y} = \mathbf{X}^T\theta + \boldsymbol{\epsilon}$ 대신에 $\boldsymbol{y} = \mathbf{Z}^{\boldsymbol{T}}\boldsymbol{\vartheta} + \boldsymbol{\epsilon}$ 를 이용하여 regression을 하면 underfitting 문제를 완화할 수 있다.

(3) $\boldsymbol{y} = \mathbf{X}^T\theta + \boldsymbol{\epsilon}$대신에 $\boldsymbol{y} = \mathbf{Z}^{\boldsymbol{T}}\boldsymbol{\vartheta} + \boldsymbol{\epsilon}$를 이용하여 regression을 하면 계산량이 줄어든다.

(4) 틀린 내용 없음

답: (3), underfitting이 아니고 overfitting을 완화하는 효과가 있음.

[18] (5점) Gaussian Process Regression에서는 $y = f(\mathbf{x}) + \epsilon$ 에서 $f(\mathbf{x})$ 의 분포를 $\boldsymbol{f}(\boldsymbol{x}) \sim \boldsymbol{N}\left(\boldsymbol{\mu_f}(\mathbf{x}),\, \boldsymbol{k}(\mathbf{x},\, \mathbf{x}')\right)$ 로 가정을 하고 $\boldsymbol{\mu_f}(\mathbf{x}) = \mathbb{E}[\boldsymbol{f}(\mathbf{x})]$, $\boldsymbol{k}(\mathbf{x},\, \mathbf{x}') = \mathbb{E}\left[\left(\boldsymbol{f}(\mathbf{x}) - \boldsymbol{\mu_f}(\mathbf{x})\right)\left(\boldsymbol{f}(\mathbf{x}') - \boldsymbol{\mu_f}(\mathbf{x}')\right)\right]$를 구하는 것이 목적이다. 다음 중 틀린 내용은?

(1) $\boldsymbol{f}(\mathbf{x})$를 모르므로 $\mathbf{x}$ 값 만을 이용하여 $\boldsymbol{k}(\mathbf{x},\, \mathbf{x}')$ 를 구하는 함수를 kernel 이라고 한다.

(2) Kernel 함수를 $k(\mathbf{x},\, \mathbf{x}') = \sigma_f^2 \mathbf{exp}\left(-\frac{1}{2\lambda}(\mathbf{x} - \mathbf{x}')^T\,(\mathbf{x} - \mathbf{x}')\right)$ 로 선택했을 때, $\boldsymbol{f}(\mathbf{x})$와 $\boldsymbol{f}(\mathbf{x}')$간의 연관성은 $\mathbf{x}$ 와 $\mathbf{x}'$의 거리가 클 수록 낮아지는데, 낮아지는 정도는 λ 값이 작을수록 더 천천히 낮아진다.

(3) σ_f^2 값이 크면 $\mathbf{x}$ 의 변동에 $\boldsymbol{f}(\mathbf{x})$가 크게 변동한다는 것을 의미한다.

(4) 최적 λ, σ_f^2 는 $p(\mathbf{y}|\mathbf{X})$를 최대로 하도록 구한다.

(5) 틀린 내용 없음

답: (2) 낮아지는 정도는 λ 값이 작을수록 더 급격히 낮아짐..

[19] (5점) Kalman filter에서 추정해야 할 상태 값은 $x_k^a = E[x_k|z_k, \dots, z_1]$이다. 여기서 $x_k^a = E\left[x_k\middle|z_{k-1}, \dots, z_1\right] + E[x_k|z_k] = x_k^f + K_k(z_k - H_k x_k^f)$ 로 구한다. K_k 를 Kalman Gain이라고 한다. 다음 중 틀린 내용은?

(1) x_k^f는 $x_k^f = A_{k-1}x_{k-1}^a + B_{k-1}u_{k-1}$식으로부터 구한다.

(2) K_k 는 estimation error covariance matrix $P_k = E[(x_k - x_k^a)(x_k - x_k^a)^T]$의 trace 를 최소로 하도록 구한다.

(3) $z_k - H_k x_k^f$ 값의 의미는 새로 관측된 출력 z_k 와 $k-1$ 까지의 관측데이터부터 prediction된 출력 $\hat{z}_k$간의 오차이다.

(4) Kalman filter는 visual tracking 문제에도 적용할 수 있다.

(5) 틀린 내용 없음

답: (5)

[20] (5점) 병원에서 암표지자 검사를 받아서 표지자 값이 $x = 3.2$ 가 나왔다. 이 값으로 posteriori 확률을 계산하였더니 암일 확률이 0.1이다. 정상인을 암환자로 판정할 시 risk를 1으로 하고 암환자를 정상인으로 판단시 risk 는 10로 준다. 정확히 판정하는 경우 risk는 0으로 한다. Expected risk 에 기반하여 암 환자 여부를 판단하시오.

답:

정상으로 판단시

risk $R(\alpha_1|x = 3.2) = \lambda_{12} p\left(\text{암}\middle|x = 3.2\right) = 10 \times 0.1$

암환자로 판단시

risk $R(\alpha_2|x = 3.2) = \lambda_{21} p\left(\text{정}\middle|x = 3.2\right) = 1 \times 0.9$

$R(\alpha_2|x = 3.2) < R(\alpha_1|x = 3.2)$ 이므로 암환자로 판정한다.

[21] (5 점) 동전을 100 번 던져서 head 가 40 번 나오고 tail 이 60 번 나왔다. Head 가 나올 확률 μ_1 과 tail 이 나올 확률 μ_2 를 Bayesian learning 으로 추정하려고 한다.

μ_i 는 Dirichlet 분포를 따른다. Prior 분포 $p(\mu_1, \mu_2|\alpha_1, \alpha_2)$ 에서 hyper-parameter 가 $\alpha_1 = 10, \alpha_2 = 20$로 주어 졌을 때, μ_1, μ_2 를 추정하시오.

답:

$\mu_1 = \frac{c_1+\alpha_1}{\sum_i(c_i+\alpha_i)} = \frac{40+10}{130} = \frac{5}{13}$

$\mu_2 = \frac{c_2+\alpha_2}{\sum_i(c_i+\alpha_i)} = \frac{60+20}{130} = \frac{8}{13}$

[22] (5 점) 관측된 10 개의 word w_i 를 MCMC 로 clustering 하려고 한다. Vocabulary 는 6 개이고, cluster 수는 2 개라 가정한다. Gibbs sampling 을 위해 $p(z_i|z_{-i})$를 산출하여 sampling 하여 아래 표와 같이 cluster label 을 assign 하였다. w_i와 z_i는 multinomial 분포를 갖는다. w_i와 z_i의 분포의 파라미터 φ, θ 를 inference 하기위해 아래 표 결과로부터

$h_\theta(k) = \sum_{i=1}^{10} \delta(z_i - k)$, $h_\phi(k, v) = \sum_{i=1}^{10} \delta(w_i - v)\delta(z_i - k)$ 를 산출해야 한다.

$h_\theta(2)$, $h_\phi(1,3)$, $h_\phi(2,3)$를 구하시오.

i	1	2	3	4	5	6	7	8	9
w_i	1	3	2	3	3	5	4	1	6
z_i	1	2	2	2	1	1	2	1	2

답: $h_\theta(2) = 5$, $h_\phi(1,3) = 1$, $h_\phi(2,3) = 2$