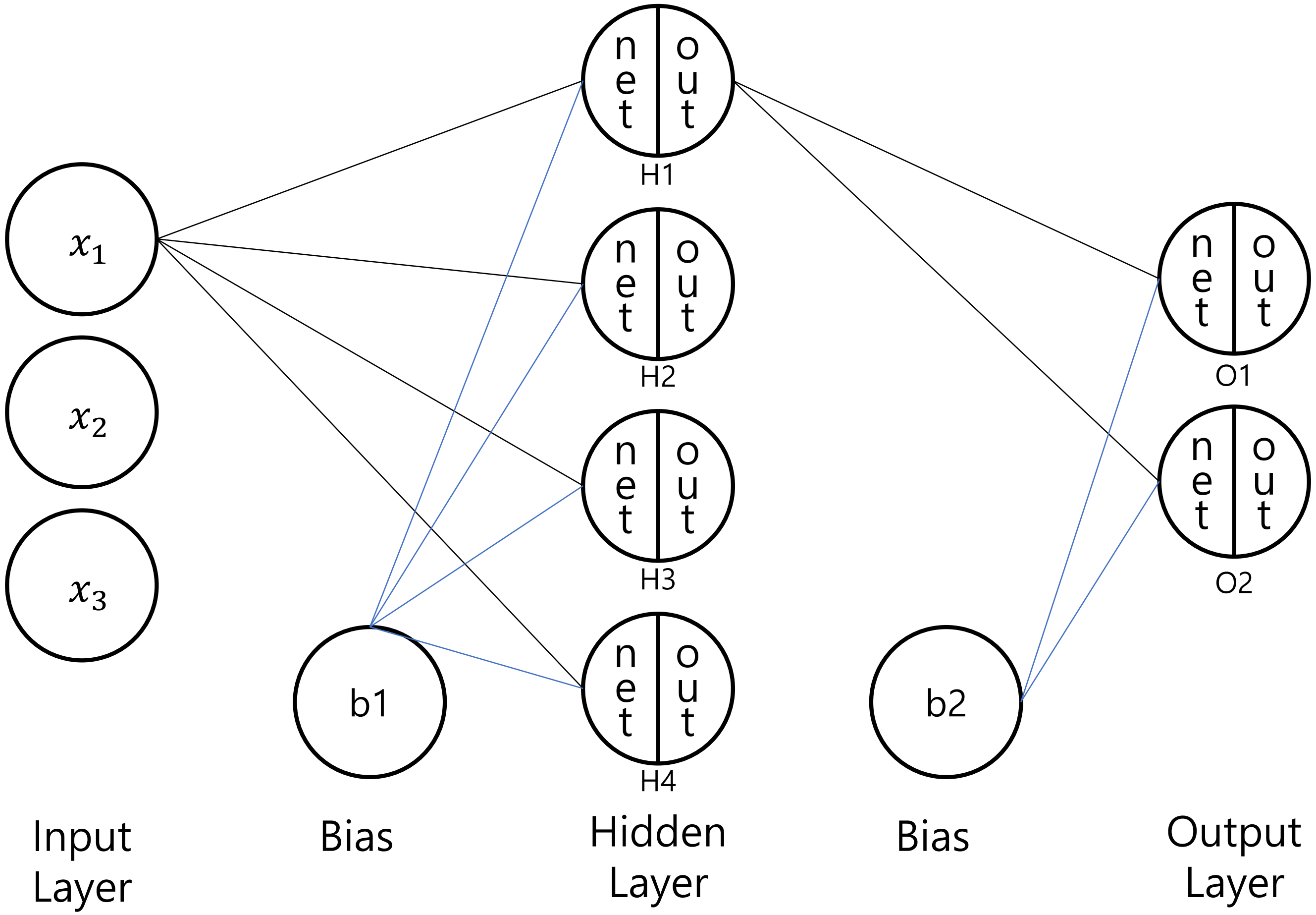
오차역전파 계산하기

By. JongSeobJ

0. 모델 모형



설명을 위해 3개의 인풋과 4개의 히든 레이어, 2개의 아웃풋으로 구성된 신경망입니다.

1. 변수 설명

$$W_{IH} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{pmatrix}, \ w_{ij} = x_i \text{ to } net_{Hj}$$

W_{IH}: 인풋 레이어에서 히든 레이어로의 가중치

w_{ij}: i번째 인풋에서 j번째 히든 레이어로의 가중치

$$net_{H1} = b_{11} + x_1 w_{11} + x_2 w_{21} + x_3 w_{31}$$

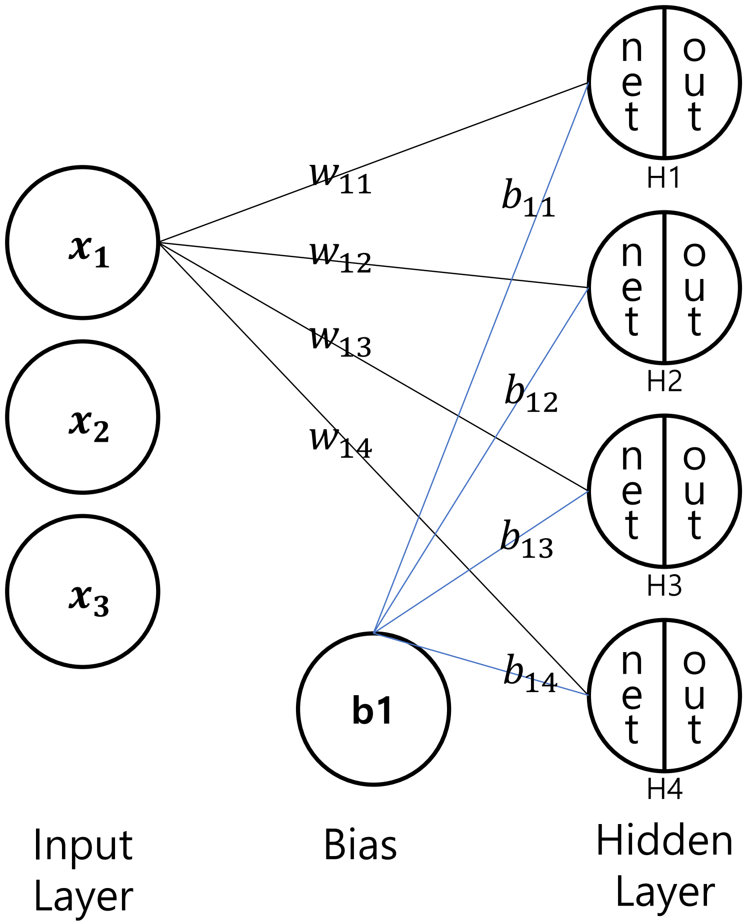
$$net_{H2} = b_{12} + x_1 w_{12} + x_2 w_{22} + x_3 w_{32}$$

$$net_{H3} = b_{13} + x_1 w_{13} + x_2 w_{23} + x_3 w_{33}$$

$$net_{H4} = b_{14} + x_1 w_{14} + x_2 w_{24} + x_3 w_{34}$$

$$out_{Hi} = \frac{1}{1 + e^{-net_{Hi}}}$$

이 예에서 활성화 함수는 시그모이드 함수를 사용



1. 변수 설명

$$W_{HO} = \begin{pmatrix} w'_{11} & w'_{12} \\ w'_{21} & w'_{22} \\ w'_{31} & w'_{32} \\ w'_{41} & w'_{42} \end{pmatrix}, \; w'_{ij} = out_{Hi} \text{ to } net_{Oj}$$

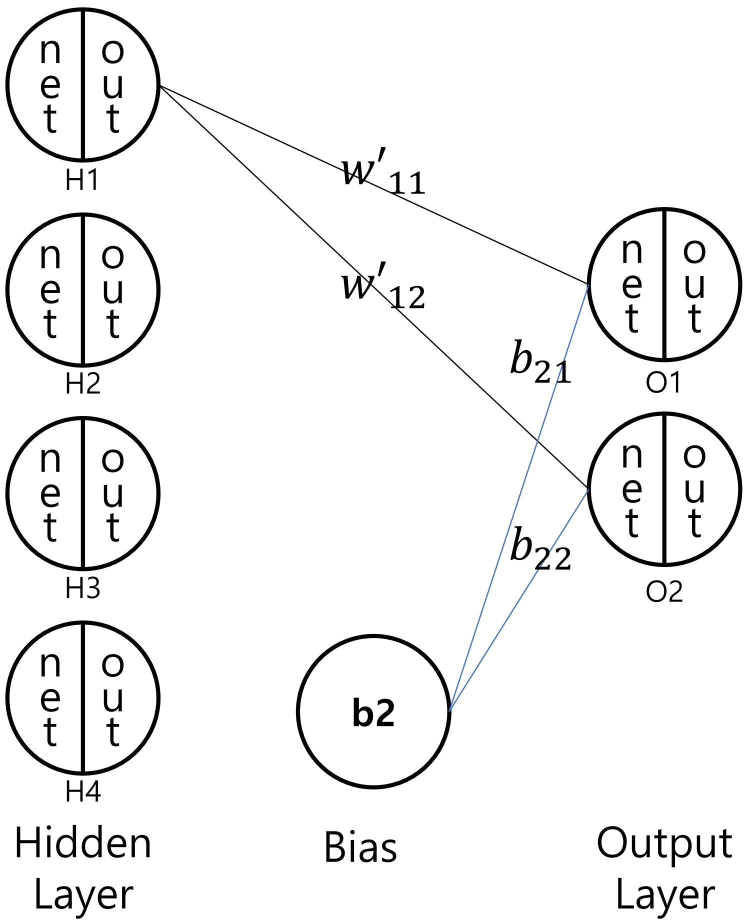
W_{HO}: 히든 레이어에서 아웃풋 레이어로의 가중치

w'_{ij}: i번째 히든레이어에서 j번째 아웃풋 레이어로의 가중치

$$net_{O1} = b_{21} + out_{H1}w'_{11} + out_{H2}w'_{21} + out_{H3}w'_{31} + out_{H4}w'_{41}$$
$$net_{O2} = b_{22} + out_{H1}w'_{12} + out_{H2}w'_{22} + out_{H3}w'_{32} + out_{H4}w'_{42}$$

$$out_{Oi} = \frac{1}{1 + e^{-net_{Oi}}}$$

이 예에서 활성화 함수는 시그모이드 함수를 사용



1. 변수 설명

에러함수

$$E_{Total} = \frac{1}{2}\sum(target - out)^2$$

$$= \frac{1}{2}\{(target_1 - out_{O1})^2 + (target_2 - out_{O2})^2\}$$

2. W_{HO}

우선 W_{HO}에 대해 계산하겠습니다.

1번째 히든레이어에서 1번째 아웃풋으로의 가중치를

체인룰을 이용하면 다음과 같이 계산할 수 있습니다

$$\frac{\partial E_{Total}}{\partial w'_{11}} = \frac{\partial E_{Total}}{\partial out_{O1}} \times \frac{\partial out_{O1}}{\partial net_{O1}} \times \frac{\partial net_{O1}}{\partial w'_{11}}$$

$$\frac{\partial E_{Total}}{\partial out_{O1}} = (out_{O1} - y_1) \quad \cdots\cdots\cdots (1)$$

$$\frac{\partial out_{O1}}{\partial net_{O1}} = out_{O1}(1 - out_{O1}) \quad \cdots (2)$$

$$\frac{\partial net_{O1}}{\partial w'_{11}} = out_{H1} \quad \cdots\cdots\cdots\cdots\cdots (3)$$

(1),(2),(3)을 곱하면

$$= (out_{O1} - y_1)out_{O1}(1 - out_{O1})out_{H1}$$

$$\delta_{O1} = (out_{O1} - y_1)out_{O1}(1 - out_{O1}) \Rightarrow \frac{\partial E_{Total}}{\partial w'_{11}} = \delta_{O1} out_{H1}$$

계속해서 첫번째 아웃풋으로의 값들을 계산하면

$$\frac{\partial E_{Total}}{\partial w'_{21}} = (out_{O1} - y_1)out_{O1}(1 - out_{O1})out_{H2} \quad , \frac{\partial E_{Total}}{\partial w'_{31}} = \delta_{O1} out_{H3} \quad , \frac{\partial E_{Total}}{\partial w'_{41}} = \delta_{O1} out_{H4}$$

$$= \delta_{O1} out_{H2}$$

2. W_{HO}

2번 아웃풋으로의 가중치들에 대해서 계산 과정입니다.

$$\frac{\partial E_{Total}}{\partial w'_{12}} = \frac{\partial E_{Total}}{\partial out_{O2}} \times \frac{\partial out_{O2}}{\partial net_{O2}} \times \frac{\partial net_{O2}}{\partial w'_{21}}$$

$$\frac{\partial E_{Total}}{\partial out_{O2}} = (out_{O2} - y_2) \quad \cdots\cdots\cdots (1)$$

$$\frac{\partial out_{O2}}{\partial net_{O2}} = out_{O2}(1 - out_{O2}) \quad \cdots (2)$$

$$\frac{\partial net_{O2}}{\partial w'_{21}} = out_{H1} \quad \cdots\cdots\cdots\cdots\cdots\cdots (3)$$

(1),(2),(3)을 곱하면

$$= (out_{O2} - y_2) out_{O2}(1 - out_{O2}) out_{H1}$$

$$\delta_{O2} = (out_{O2} - y_2) out_{O2}(1 - out_{O2}) \Rightarrow \frac{\partial E_{Total}}{\partial w'_{21}} = \delta_{O2} out_{H1}$$

위 과정을 계속 하면 W_{HO}를 갱신하기 위한 행렬은 다음과 같습니다.

$$\begin{pmatrix} \delta_{O1} out_{H1} & \delta_{O2} out_{H1} \\ \delta_{O1} out_{H2} & \delta_{O2} out_{H2} \\ \delta_{O1} out_{H3} & \delta_{O2} out_{H3} \\ \delta_{O1} out_{H4} & \delta_{O2} out_{H4} \end{pmatrix}$$

2. W_{HO}

W_{HO}를 갱신하기 위해서 위 값에 *learning rate*를 곱한 후 빼면

$$W_{HO}{}^{(new)} = \begin{pmatrix} w'_{11} & w'_{12} \\ w'_{21} & w'_{22} \\ w'_{31} & w'_{32} \\ w'_{41} & w'_{42} \end{pmatrix} - lr * \begin{pmatrix} \delta_{O1} out_{H1} & \delta_{O2} out_{H1} \\ \delta_{O1} out_{H2} & \delta_{O2} out_{H2} \\ \delta_{O1} out_{H3} & \delta_{O2} out_{H3} \\ \delta_{O1} out_{H4} & \delta_{O2} out_{H4} \end{pmatrix}$$

그리고 *b2*는 다음과 같습니다.

$$\frac{\partial E_{Total}}{\partial b_2} = \frac{\partial E_{Total}}{\partial out_{O2}} \times \frac{\partial out_{O2}}{\partial net_{O2}} \times \frac{\partial net_{O2}}{\partial b_{21}} = \delta_{O1} * 1$$

$$b2^{(new)} = (b_{21} \quad b_{22}) - l\mathrm{r} * (\delta_{O1} \quad \delta_{O2})$$

3. W_{IH}

이번에는 W_{IH}에 대해 계산해보겠습니다.

1번째 인풋에서 1번째 히든레이어로의 가중치를 계산해보겠습니다.

$$\frac{\partial E_{Total}}{\partial w_{11}} = \frac{\partial E_{Total}}{\partial out_{H1}} \times \frac{\partial out_{H1}}{\partial net_{H1}} \times \frac{\partial net_{H1}}{\partial w_{11}}$$

여기서 1번째 히든레이어는 1번째 아웃풋과 2번째 아웃풋 둘 다에게 영향을 미침으로

$$\frac{\partial E_{Total}}{\partial out_{H1}} = \frac{\partial E_1}{\partial out_{H1}} + \frac{\partial E_2}{\partial out_{H1}}$$

$$= \frac{\partial E_1}{\partial net_{O1}} \times \frac{\partial E_{Total}}{\partial out_{H1}} + \frac{\partial E_{Total}}{\partial out_{H1}} \times \frac{\partial E_{Total}}{\partial out_{H1}}$$

$$= \delta_{O1} w'_{11} + \delta_{O2} w'_{12} \quad \cdots\cdots\cdots (1)$$

$$\frac{\partial out_{H1}}{\partial net_{H1}} = out_{H1}(1 - out_{H1}) \quad \cdots\cdots\cdots (2)$$

$$\frac{\partial net_{H1}}{\partial w_{11}} = x_1 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (3)$$

(1),(2),(3)을 곱하면

$$= \underline{(\delta_{O1} w'_{11} + \delta_{O2} w'_{12}) out_{H1}(1 - out_{H1})} \, x_1$$

$$\delta_{H1} = (\delta_{O1} w'_{11} + \delta_{O2} w'_{12}) out_{H1}(1 - out_{H1}) \Rightarrow \frac{\partial E_{Total}}{\partial w_{11}} = \delta_{H1} x_1$$

3. W_{IH}

계속해서 다른 값들에 대해 계산하면

$$\frac{\partial E_{Total}}{\partial w_{12}} = \underbrace{(\delta_{O1}w'_{21} + \delta_{O2}w'_{22})out_{H2}(1 - out_{H2})}\, x_1$$

$$\delta_{H2} = (\delta_{O1}w'_{21} + \delta_{O2}w'_{22})out_{H2}(1 - out_{H2}) \quad \Rightarrow \quad \frac{\partial E_{Total}}{\partial w_{12}} = \delta_{H2}x_1$$

위 과정을 다른 가중치들에 대해서 계산을 하면 W_{IH}을 갱신하기 위한 행렬은 다음과 같습니다.

$$\begin{pmatrix} \delta_{H1}x_1 & \delta_{H2}x_1 & \delta_{H3}x_1 & \delta_{H4}x_1 \\ \delta_{H1}x_2 & \delta_{H2}x_2 & \delta_{H3}x_2 & \delta_{H4}x_2 \\ \delta_{H1}x_3 & \delta_{H2}x_3 & \delta_{H3}x_3 & \delta_{H4}x_3 \end{pmatrix}$$

3. W_{IH}

W_{IH}를 갱신하기 위해서 위 값에 *learning rate*를 곱한 후 빼면

$$W_{IH}^{(new)} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{pmatrix} - lr * \begin{pmatrix} \delta_{H1}x_1 & \delta_{H2}x_1 & \delta_{H3}x_1 & \delta_{H4}x_1 \\ \delta_{H1}x_2 & \delta_{H2}x_2 & \delta_{H3}x_2 & \delta_{H4}x_2 \\ \delta_{H1}x_3 & \delta_{H2}x_3 & \delta_{H3}x_3 & \delta_{H4}x_3 \end{pmatrix}$$

그리고 *b1*는 다음과 같습니다.

$$\frac{\partial E_{Total}}{\partial b_{11}} = \frac{\partial E_1}{\partial out_{H1}} * \frac{\partial out_{H1}}{\partial net_{H1}} * \frac{\partial net_{H1}}{\partial b_1} = \delta_{H1}$$

$$b1^{(new)} = (b_{11} \ \ b_{12} \ \ b_{13} \ \ b_{14}) - l\mathrm{r} * (\delta_{H1} \ \ \delta_{H2} \ \ \delta_{H3} \ \ \delta_{H4})$$

4. 구현

$$W_{HO}{}^{(new)} = \begin{pmatrix} w'_{11} & w'_{12} \\ w'_{21} & w'_{22} \\ w'_{31} & w'_{32} \\ w'_{41} & w'_{42} \end{pmatrix} - lr * \begin{pmatrix} \delta_{O1} out_{H1} & \delta_{O2} out_{H1} \\ \delta_{O1} out_{H2} & \delta_{O2} out_{H2} \\ \delta_{O1} out_{H3} & \delta_{O2} out_{H3} \\ \delta_{O1} out_{H4} & \delta_{O2} out_{H4} \end{pmatrix}$$

$$\delta_{O1} = (out_{O1} - y_1) out_{O1} (1 - out_{O1})$$

$$(out_{O1} \quad out_{O2}) - (y_1 \quad y_2) = (out_{O1} - y_1 \quad out_{O2} - y_2)$$

$$(out_{O1} - y_1 \quad out_{O2} - y_2) \odot (out_{O1}(1 - out_{O1}) \quad out_{O2}(1 - out_{O2})) = (\delta_{O1} \quad \delta_{O2})$$

$\odot$: 행렬의 가로 곱 eg) $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \odot \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 \times 4 & 2 \times 3 \\ 3 \times 2 & 4 \times 1 \end{pmatrix}$

$$\begin{pmatrix} out_{H1} \\ out_{H2} \\ out_{H3} \\ out_{H4} \end{pmatrix} \circ (\delta_{O1} \quad \delta_{O2}) = \begin{pmatrix} \delta_{O1} out_{H1} & \delta_{O2} out_{H1} \\ \delta_{O1} out_{H2} & \delta_{O2} out_{H2} \\ \delta_{O1} out_{H3} & \delta_{O2} out_{H3} \\ \delta_{O1} out_{H4} & \delta_{O2} out_{H4} \end{pmatrix}$$

4. 구현

$$W_{IH}{}^{(new)} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{pmatrix} - lr * \begin{pmatrix} \delta_{H1}x_1 & \delta_{H2}x_1 & \delta_{H3}x_1 & \delta_{H4}x_1 \\ \delta_{H1}x_2 & \delta_{H2}x_2 & \delta_{H3}x_2 & \delta_{H4}x_2 \\ \delta_{H1}x_3 & \delta_{H2}x_3 & \delta_{H3}x_3 & \delta_{H4}x_3 \end{pmatrix}$$

$$\delta_{H1} = (\delta_{O1}w'_{11} + \delta_{O2}w'_{12})out_{H1}(1 - out_{H1})$$

$$\Rightarrow (\delta_{O1} \quad \delta_{O2}) \circ \begin{pmatrix} w'_{11} & w'_{12} \\ w'_{21} & w'_{22} \\ w'_{31} & w'_{32} \\ w'_{41} & w'_{42} \end{pmatrix}^T = (\delta_{O1} \quad \delta_{O2}) \circ \begin{pmatrix} w'_{11} & w'_{21} & w'_{31} & w'_{41} \\ w'_{12} & w'_{22} & w'_{32} & w'_{42} \end{pmatrix}$$

$$= (\delta_{O1}w'_{11} + \delta_{O2}w'_{12} \quad \delta_{O1}w'_{21} + \delta_{O2}w'_{22} \quad \delta_{O1}w'_{31} + \delta_{O2}w'_{32} \quad \delta_{O1}w'_{41} + \delta_{O2}w'_{42})$$

$$(\delta_{O1}w'_{11} + \delta_{O2}w'_{12} \quad \delta_{O1}w'_{21} + \delta_{O2}w'_{22} \quad \delta_{O1}w'_{31} + \delta_{O2}w'_{32} \quad \delta_{O1}w'_{41} + \delta_{O2}w'_{42})$$
$$\odot(out_{H1}(1 - out_{H1}) \quad out_{H2}(1 - out_{H2}) \quad out_{H3}(1 - out_{H3}) \quad out_{H4}(1 - out_{H4}))$$

$$= (\delta_{H1} \quad \delta_{H2} \quad \delta_{H3} \quad \delta_{H4})$$

4. 구현

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \circ \begin{pmatrix} \delta_{H1} & \delta_{H2} & \delta_{H3} & \delta_{H4} \end{pmatrix} = \begin{pmatrix} \delta_{H1}x_1 & \delta_{H2}x_1 & \delta_{H3}x_1 & \delta_{H4}x_1 \\ \delta_{H1}x_2 & \delta_{H2}x_2 & \delta_{H3}x_2 & \delta_{H4}x_2 \\ \delta_{H1}x_3 & \delta_{H2}x_3 & \delta_{H3}x_3 & \delta_{H4}x_3 \end{pmatrix}$$