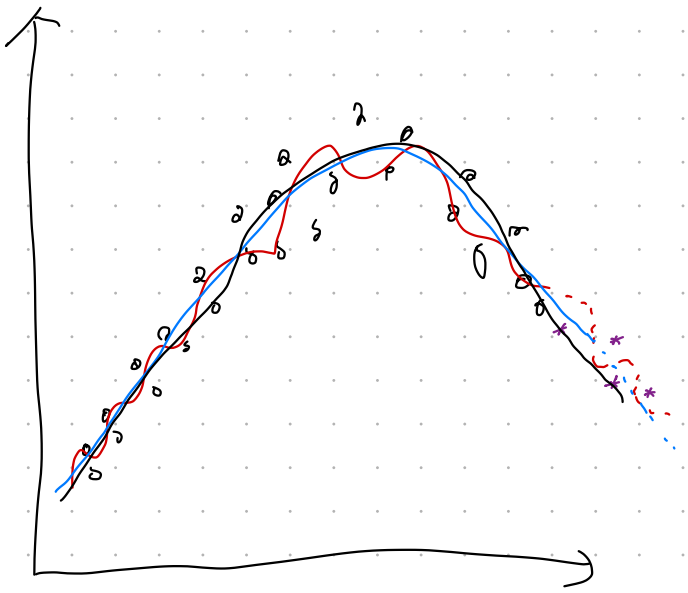
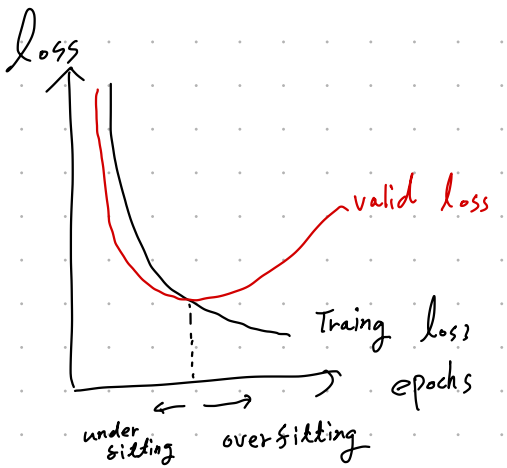
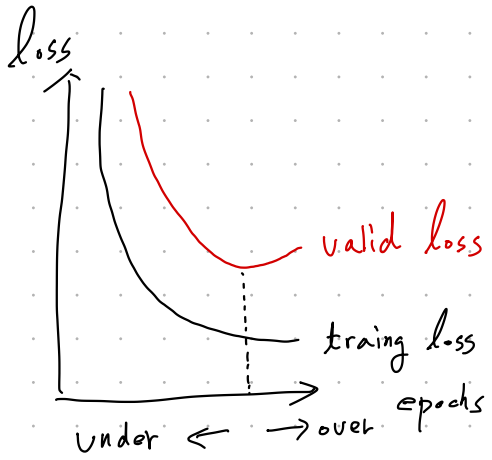
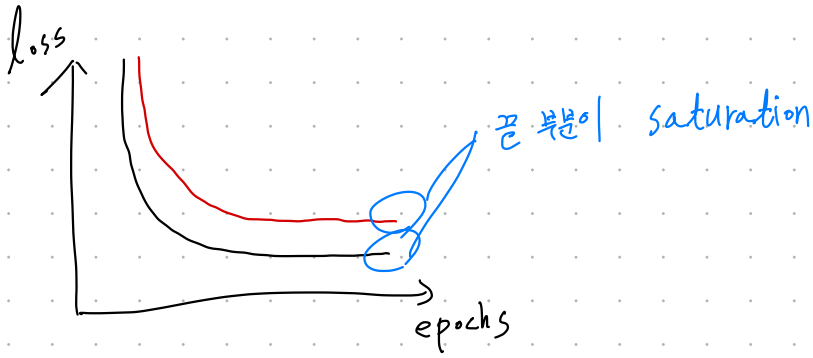
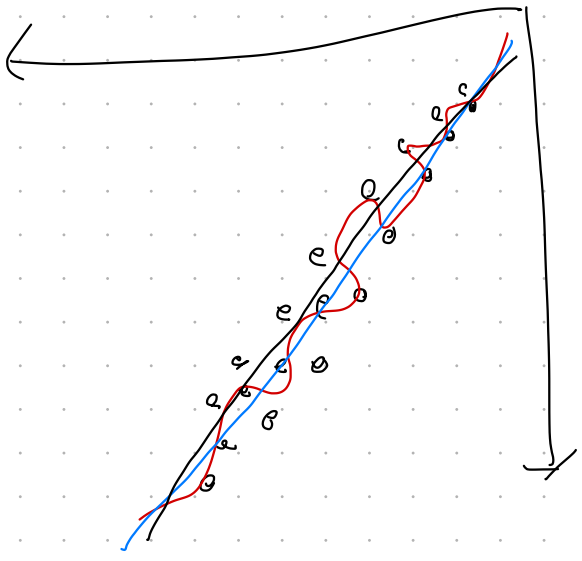
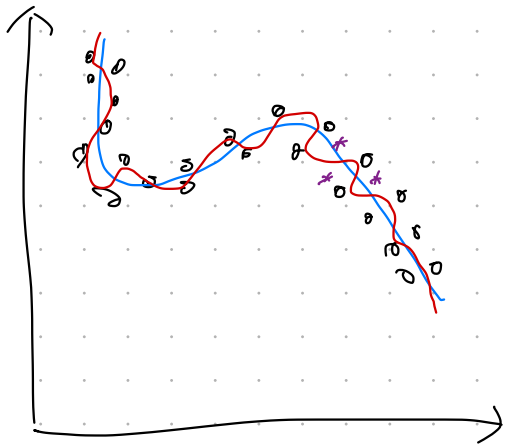


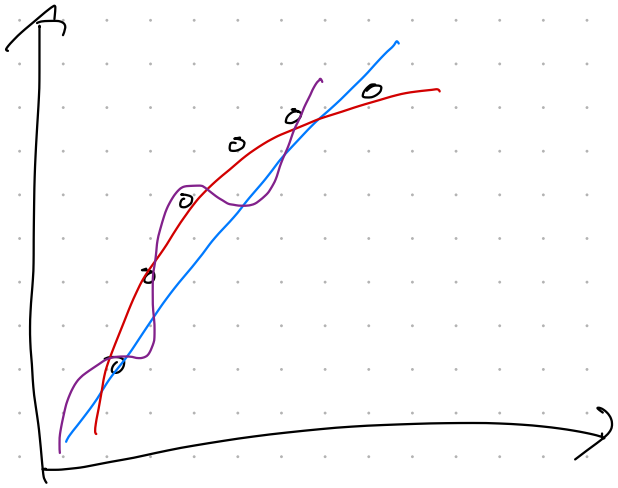

loss
valid loss
Traing loss
epochs
under fitting
overfitting

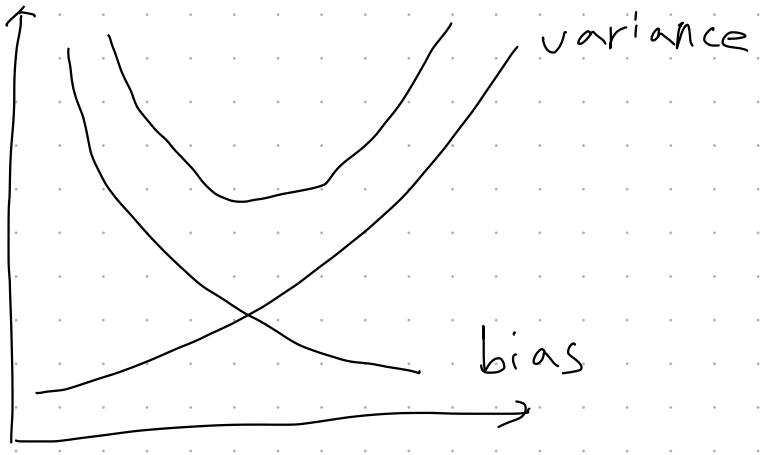

loss
valid loss
traing loss
epochs
under
over


loss
끝 부분이 saturation
epochs









under-fitting = high bias

$w_0 + w_1 x$

$w_0 + w_1 x + w_2 x^2$

$w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

over-fitting = high variance

모델의 복잡도

모델의 복잡도

$|y' - y|^2$ = 편차를 줄이는 방향 = loss → minma(loss)

훈련

$w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

↳ $w_0 + w_1 x + w_2 x^2 + \boxed{w_3 x^3 + w_4 x^4} \Rightarrow 0$으로 가까이

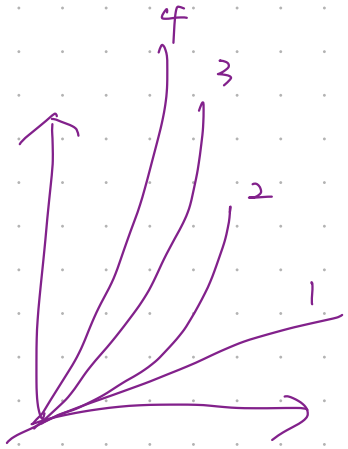
$\begin{pmatrix} loss + L_1 \\ loss + L_2 \end{pmatrix}$ 형태소 변환

minma(new loss)

훈련

$$\text{minma}\left(w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + l_1\right)$$

$$\text{minma}\left(\Sigma(wx) + \lambda \Sigma w\right) \qquad \lambda \text{ 값을 조작}$$



α = learning rate

$$\left(w_{\text{새로}} = w - \alpha\left|\hat{y} - y\right|^2\right) = \text{loss function}$$

$$\text{minma}\left(w - \alpha\left(w_0 + w_1 x_i + w_2 x_i^2 + w_3 x_k^3 + w_4 x_L^4\right) + \lambda w\right)$$

$\lambda \Rightarrow 100, 1000$

$x^3 \Rightarrow 0 \quad x^4 \Rightarrow 0$

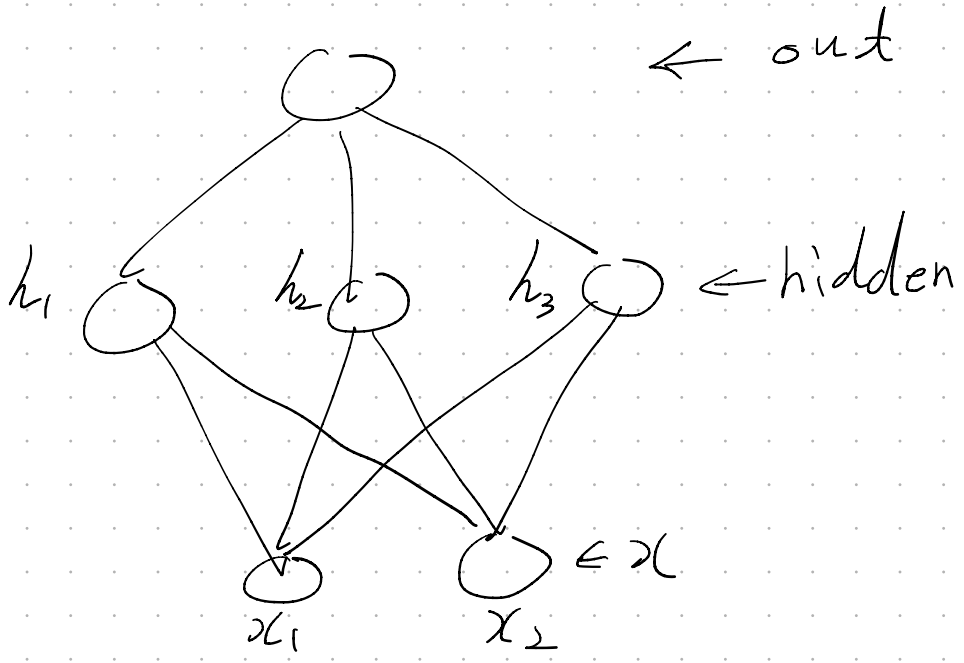
↗ 커지면

적절하게 ↗ 값을 조절하면?

$w_0 + w_1 x + w_2 x^2 + (w_3 \times 0.0000.1) + (w_4 \times 0.000.1)$

$$(\text{原loss} + \lambda w) \Rightarrow l_1$$

$$(\text{原loss} + \lambda w^2) \Rightarrow l_2$$



$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad h = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \qquad x^T W = h$$

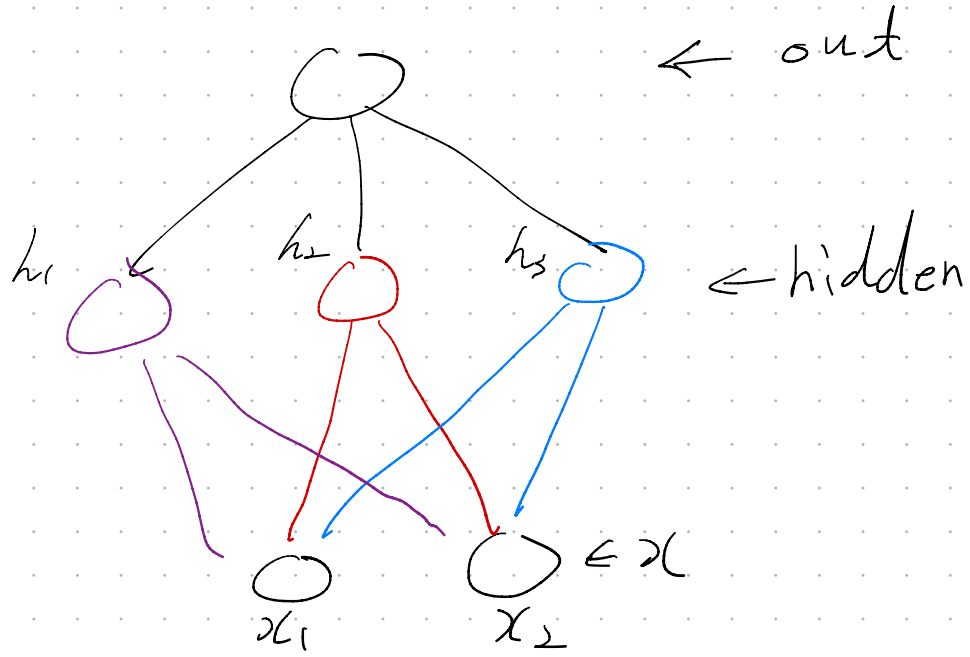
$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$$

$$[x_1\ x_2] \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

i = x의 index

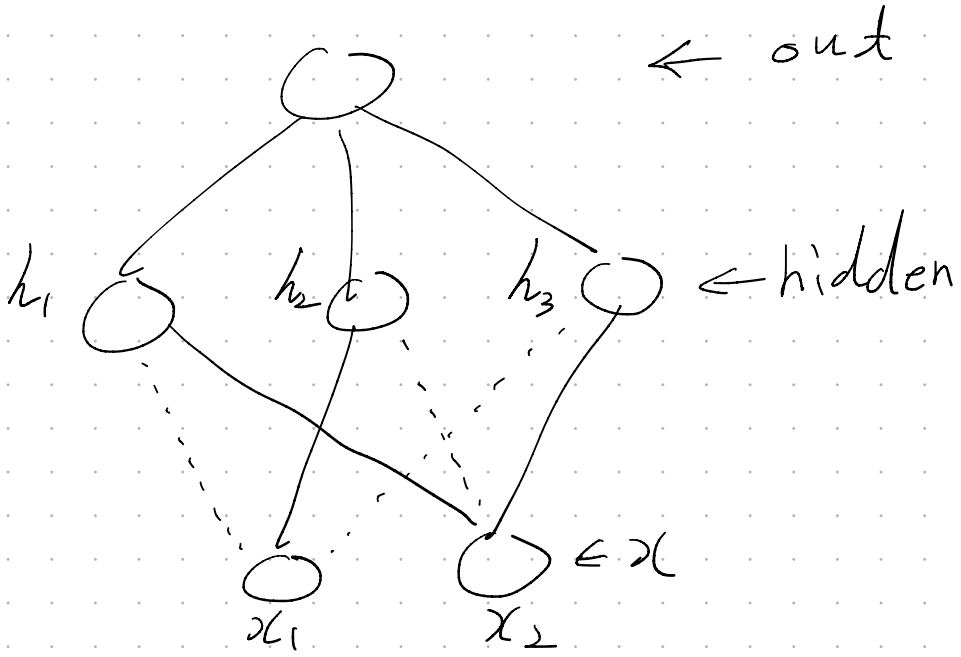
j = h의 index

drop



$$[x_1\ x_2]\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

$W[:-1, :]$ → h_3 제외



$$[x_1\ x_2]\begin{bmatrix} 0 & w_{12} & 0 \\ w_{21} & 0 & w_{23} \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

↓ 가중치

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$$

$\|w\|_2 \leq r$