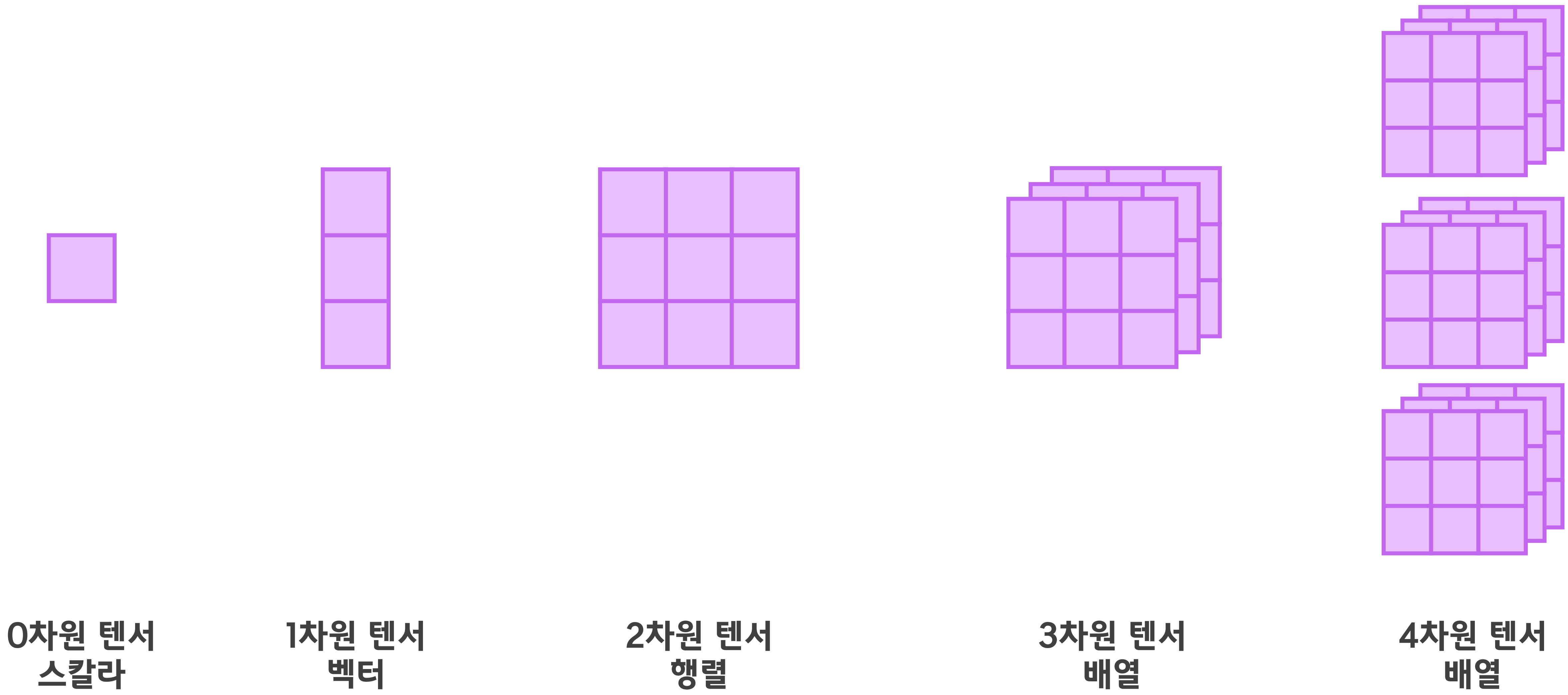
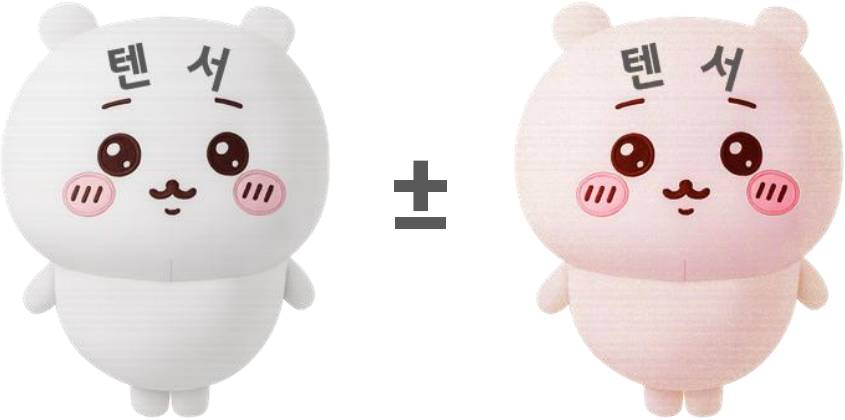

0차원 텐서
스칼라
1차원 텐서
벡터
2차원 텐서
행렬
3차원 텐서
배열
4차원 텐서
배열

3장

텐서



view

squeeze

인덱싱

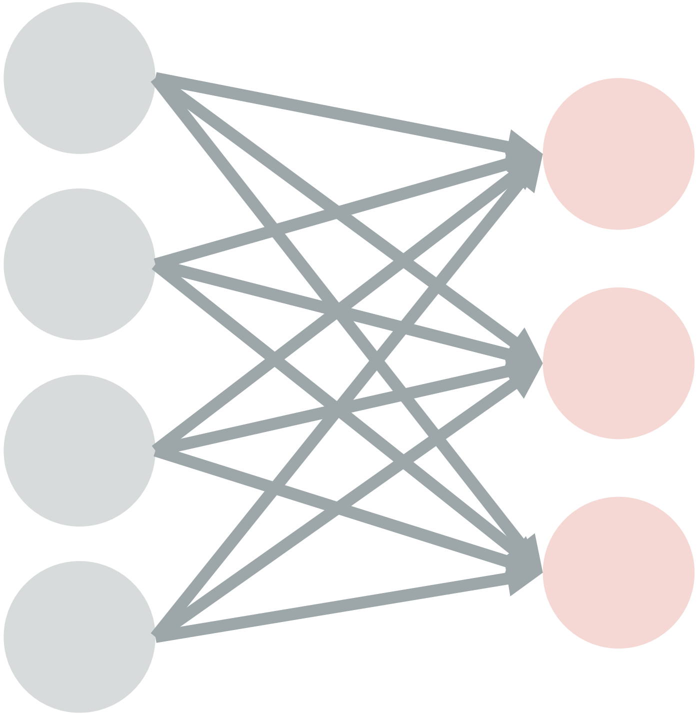
concatenate

stack

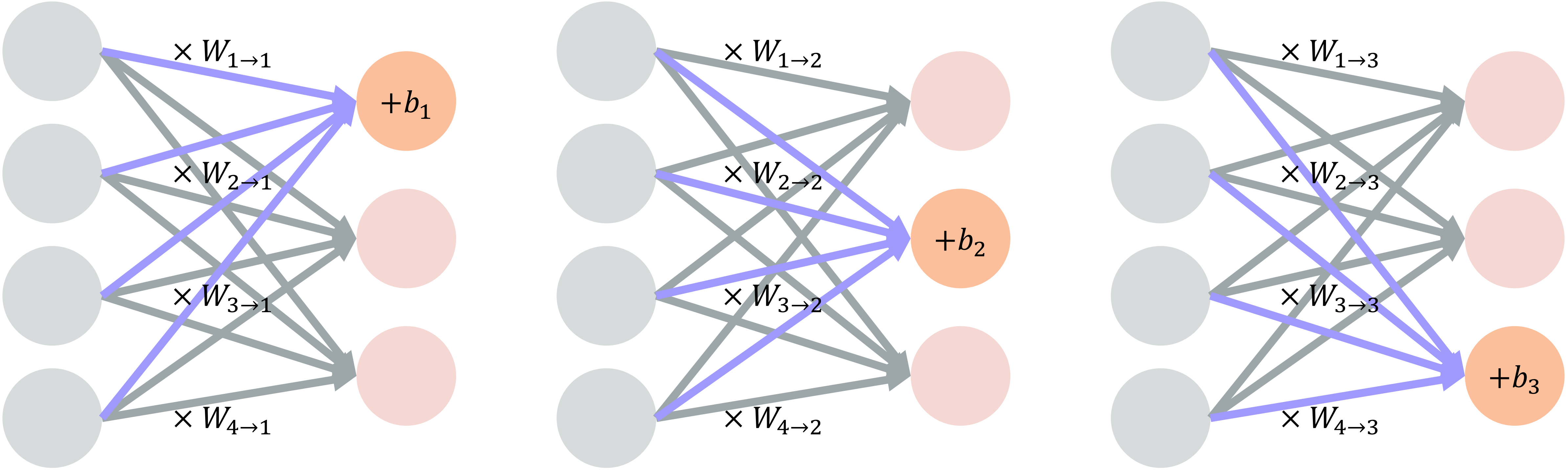
$$
\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix}\cdot\begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix}
$$
$$
=\begin{bmatrix}1\cdot1+2\cdot3+3\cdot5 & 1\cdot2+2\cdot4+3\cdot6\\4\cdot1+5\cdot3+6\cdot5 & 4\cdot2+5\cdot4+6\cdot6\end{bmatrix}
$$
$$
=\begin{bmatrix}1+6+15 & 2+8+18\\4+15+30 & 8+20+36\end{bmatrix}=\begin{bmatrix}22 & 28\\49 & 64\end{bmatrix}
$$

$$
v^\top M=\begin{bmatrix}1 & 2 & 3\end{bmatrix}\cdot\begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix}
$$
$$
=\begin{bmatrix}1\cdot1+2\cdot3+3\cdot5 & 1\cdot2+2\cdot4+3\cdot6\end{bmatrix}
$$
$$
=\begin{bmatrix}1+6+15 & 2+8+18\end{bmatrix}=\begin{bmatrix}22 & 28\end{bmatrix}
$$

선형 계층



선형 계층: 어떤 함수를 근사 계산하고 싶을 때 사용되는 가장 기본적인 모델, n개의 입력을 받아 m개의 출력을 반환하는 함수처럼 사용



출력 노드 = 입력 노드 × Weight + bias

손실 함수: 모델의 예측값과 실제값 간의 차이를 수치로 표현한 함수.
손실 함수의 값이 작아질수록 모델이 학습 데이터에 대해 더 잘 예측함

$$Loss = \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

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

L1

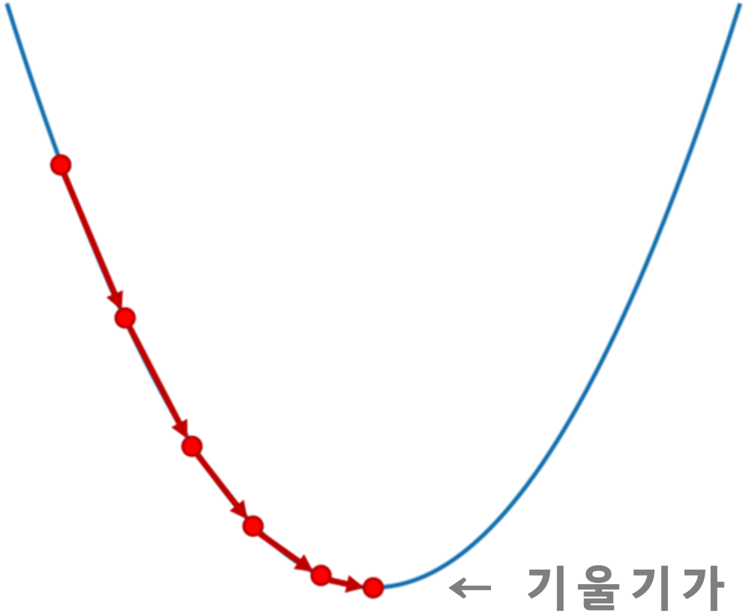
L2

RMSE

MSE

경사 하강법

$$Loss = \sum_{i=1}^{n} |y_i - \hat{y}_i|$$ → 이 값이 가장 작을수록 이득...

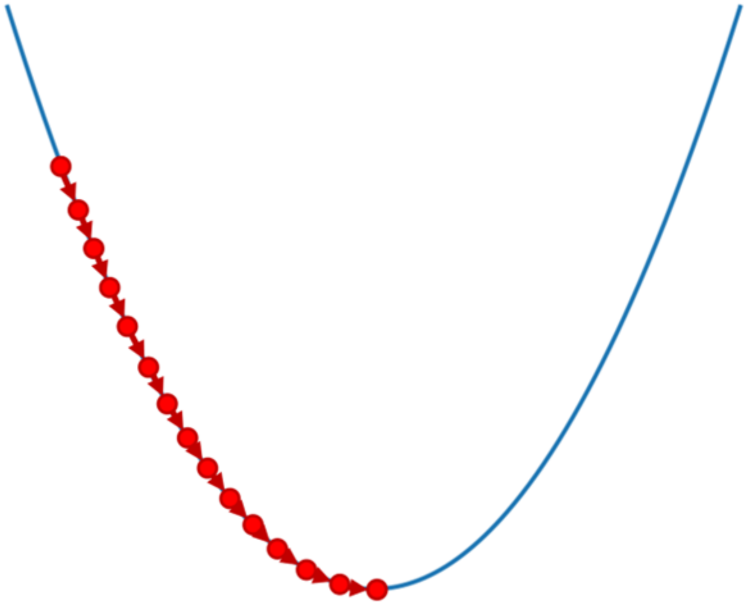


← 기울기가 0인 곳이 최솟값일 것!

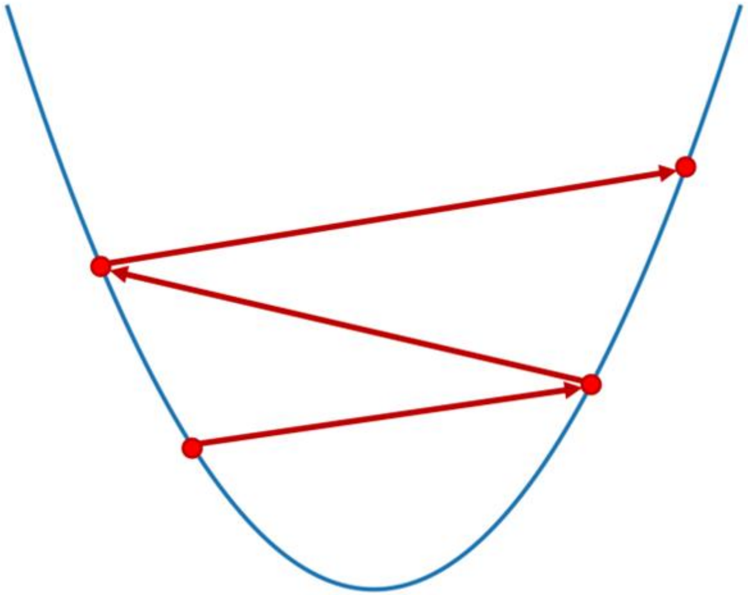
$$
\begin{aligned}
W_{i+1} &= W_i - \alpha \frac{\partial}{\partial W} MSE(W, b) \\
&= W_i - \alpha \frac{\partial}{\partial W} \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
&= W_i - \alpha \frac{\partial}{\partial W} \sum_{i=1}^{n} \left[\frac{1}{n} \{y_i - (W_i \times x + b_i)\}^2 \right] \\
&= W_i - \alpha \times \frac{2}{n} \sum_{i=1}^{n} [y_i - (W_i \times x + b_i) \times (-x)] \\
&= W_i - \alpha \times \frac{2}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) \times (-x) \\
&= W_i - \alpha \times \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) \times x \\
&= W_i - \alpha \times 2E[(\hat{y}_i - y_i) \times x] \\
\therefore W_{i+1} &= W_i - \alpha \times E[(\hat{y}_i - y_i \times x)]
\end{aligned}
$$

6장

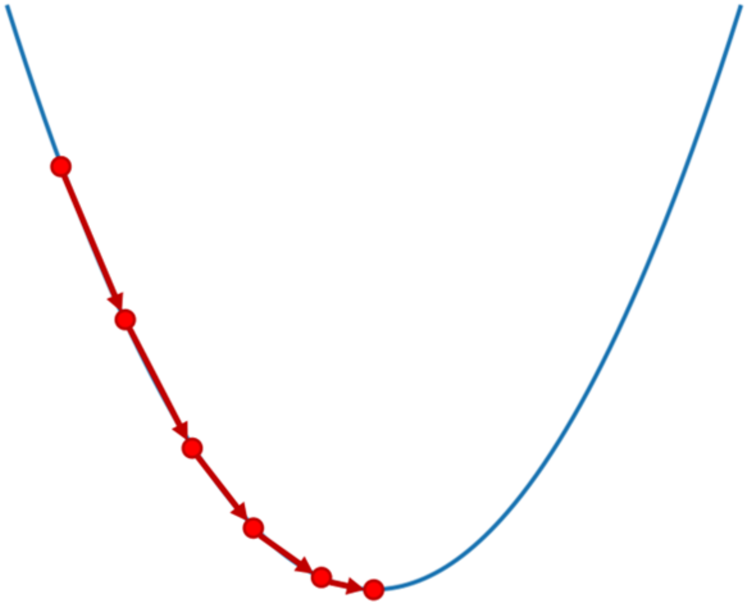
경사 하강법



낮은 학습률

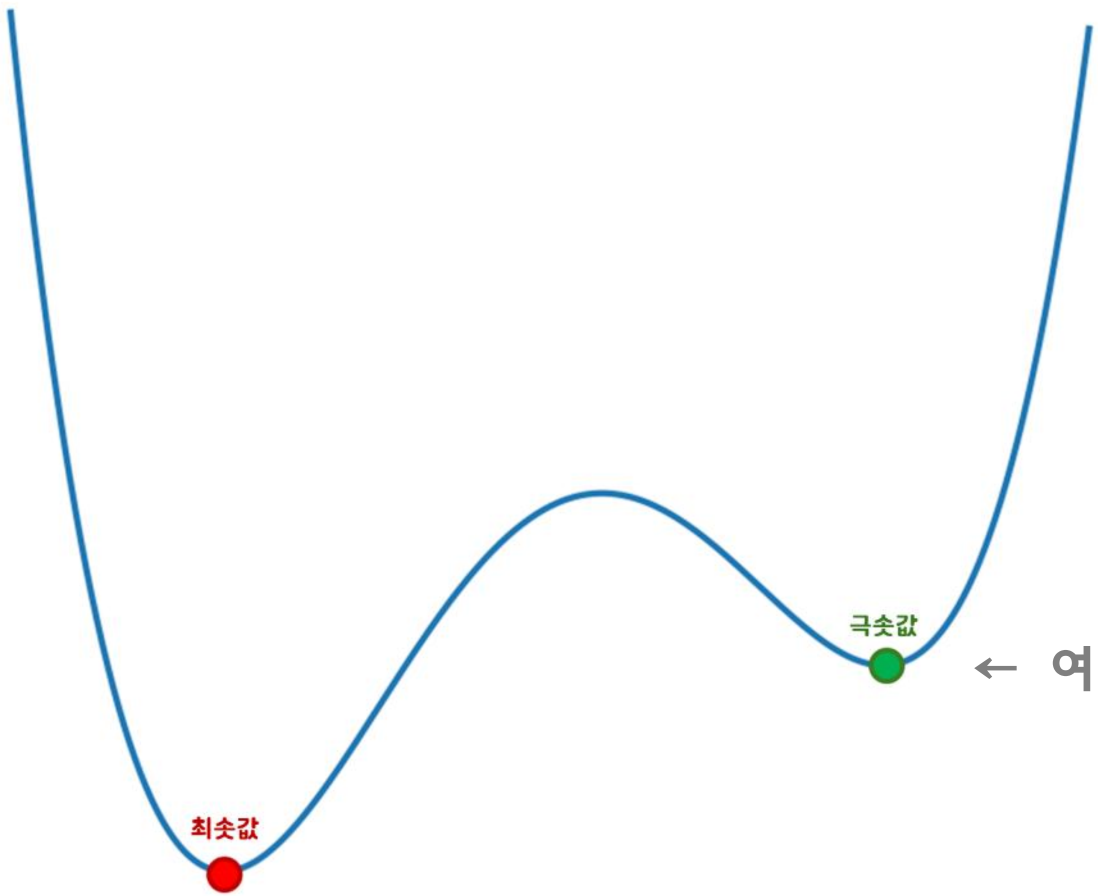


높은 학습률

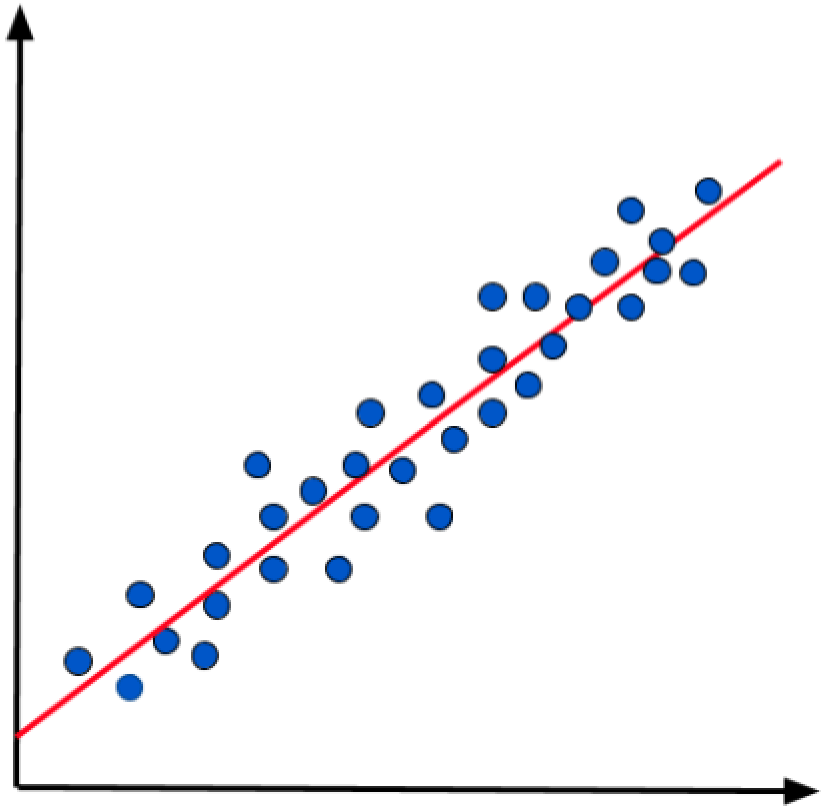


좋은 학습률

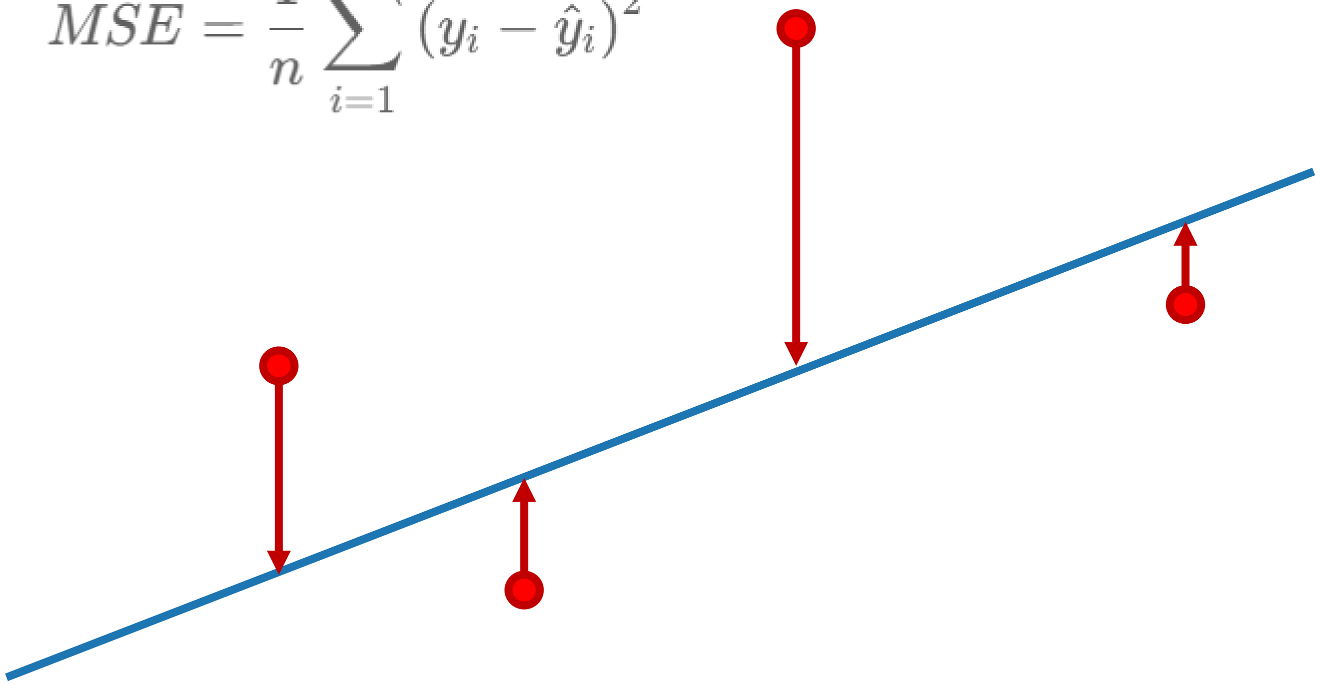
경사 하강법



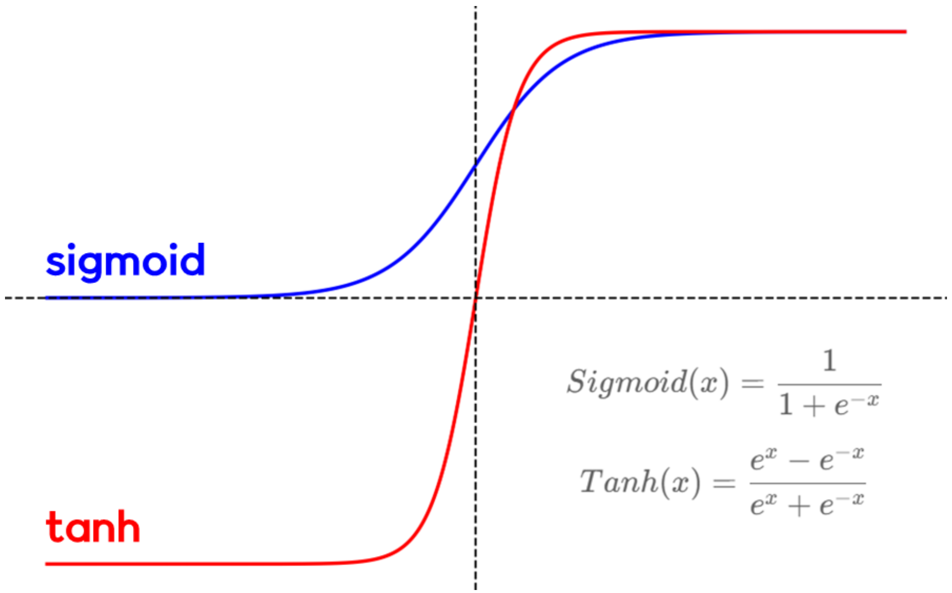
← 여기도 기울기 0인데?



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

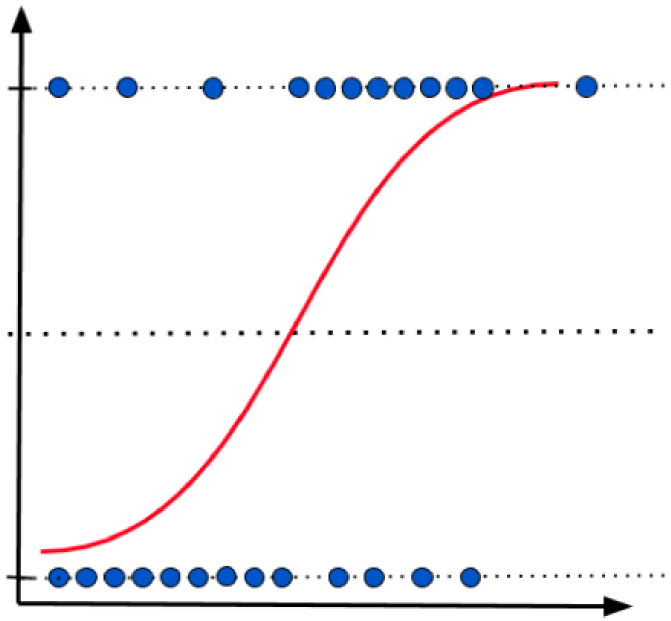


로지스틱 회귀



$$Sigmoid(x) = \frac{1}{1+e^{-x}}$$

$$Tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$



$$BCE = -\frac{1}{n}\left[y\log(\hat{y}) + (1-y)\log(1-\hat{y})\right]$$

$$BCE = -\frac{1}{n}\left[\qquad\qquad + (1-y)\log(1-\hat{y})\right]$$

$$BCE = -\frac{1}{n}\left[y\log(\hat{y}) + \qquad\qquad\qquad\qquad\right]$$