Linear Regression :

$X_{n \times d}$

$$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$g(x) = w^T x$

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

$$\boxed{\underset{n\times d}{X}\ \underset{d\times 1}{w} = \underset{n\times 1}{y}}$$

$$\boxed{\min_{w} \|Xw - y\|_2^2}$$

MSE

$Xw = y$

$n = d$

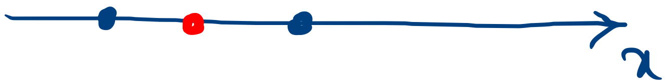
$$\Rightarrow \boxed{w = X^{-1} y}$$

$\boxed{n < d}$

$\boxed{n \geq d}$ ✓

$n = 2$

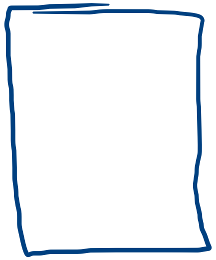
$d = 1$



$$w = \underbrace{(X^TX)^{-1}X^T}_{\text{pseudo-inverse}}\, y$$

$$\boxed{X_{n\times d}}$$

$n \geq d$



A $\quad$ B

$$AB = BA = I$$

$$\underline{(X^TX)^{-1}X^T}\, X = I$$

$$Xw = y$$

$$\boxed{w = (X^TX)^{-1}X^Ty}$$

Logistic Regression

- ⟷ Binary Classifier
- → Linear ✓
- → Bayes ✓
- → Discriminative → $p(y|x)$

$$\hat{y} = \arg\max_{i} \; p(y=i \mid x)$$

$$\Rightarrow \frac{p(y=1 \mid x)}{p(y=0 \mid x)} \underset{0}{\overset{1}{\gtrless}} 1$$

classifier

- Generative $P(x,y)$ x y
- Discriminative $\boxed{P(y|x)}$

$P(x,y)$

- $P(x)$
- $P(y)$
- $P(x|y) = \dfrac{P(x,y)}{P(y)}$
- $P(y|x) = \dfrac{P(x,y)}{P(x)}$

$$\boxed{P(y \mid x)} = \frac{P(x, y)}{P(x)} = \frac{P(y)\,P(x \mid y)}{P(x)}$$

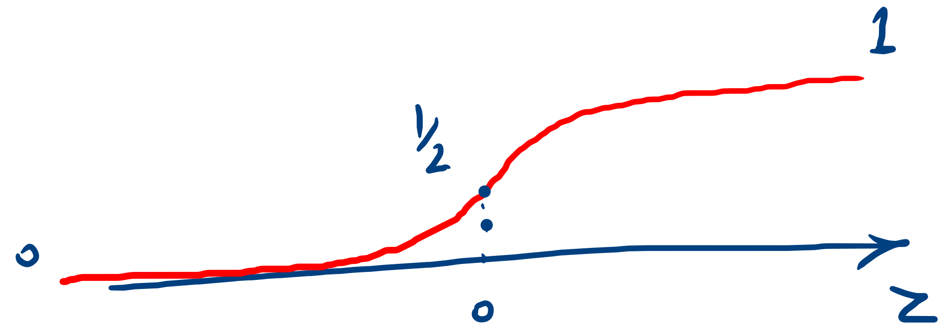
$$\frac{P(y=1|x)}{P(y=0|x)} \underset{0}{\overset{1}{\gtrless}} 1 \Longrightarrow \underbrace{\ln \frac{P(y=1|x)}{1-P(y=1|x)}}_{g(x) = w^T x} \underset{0}{\overset{1}{\gtrless}} 0$$

$$\underline{\ln \frac{P(y=1|x)}{1-P(y=1|x)} = w^T x} \Longrightarrow \boxed{P(y=1|x) = \frac{1}{1+e^{-w^T x}}}$$

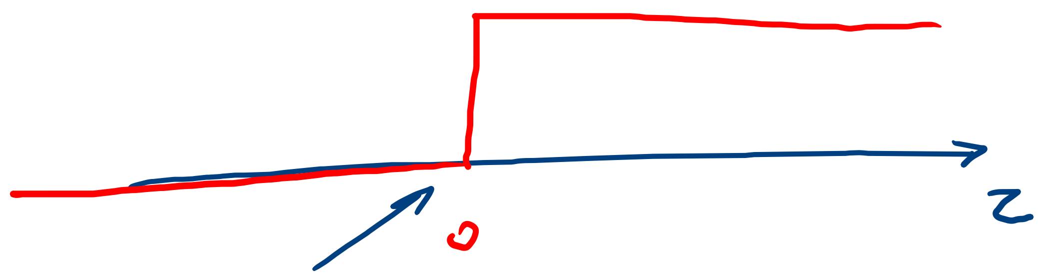
Logistic function

$$\boxed{\sigma(z) = \frac{1}{1+e^{-z}}}$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



$\text{sign}(z)$



$$\sigma'(z) = \sigma(z)\,(1-\sigma(z))$$

$\rightarrow D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ $\quad y \in \{0, 1\}$

$$P(y=1|x) = \frac{1}{1+e^{-w^T x}} = \boxed{\sigma(w^T x)}$$

$$\boxed{P(y|x) = \theta^{y} (1-\theta)^{1-y}} = \boxed{\sigma(w^T x)^{y} \left(1 - \sigma(w^T x)\right)^{1-y}}$$

$$\hat{w}_{ML} = \arg\max_{w} P(D|w) = \arg\max_{w} \ln P(D|w)$$

$$= \arg\max_{w} \ln P(y_1, \dots, y_n | x_1, \dots, x_n, w)$$

$$\overset{i.i.d.}{=} \arg\max_{w} \sum_{i=1}^{n} \ln P(y_i | x_i; w)$$

$$= \arg\max_{w} \underbrace{\sum_{i} \left(y_i \ln \sigma(w^T x_i) + (1-y_i) \ln \left(1 - \sigma(w^T x_i)\right) \right)}_{L(w)}$$

$$\nabla_w L = \boxed{\sum_{i=1}^{n} x_i \left(y_i - \sigma(w^T x_i) \right) = 0}$$

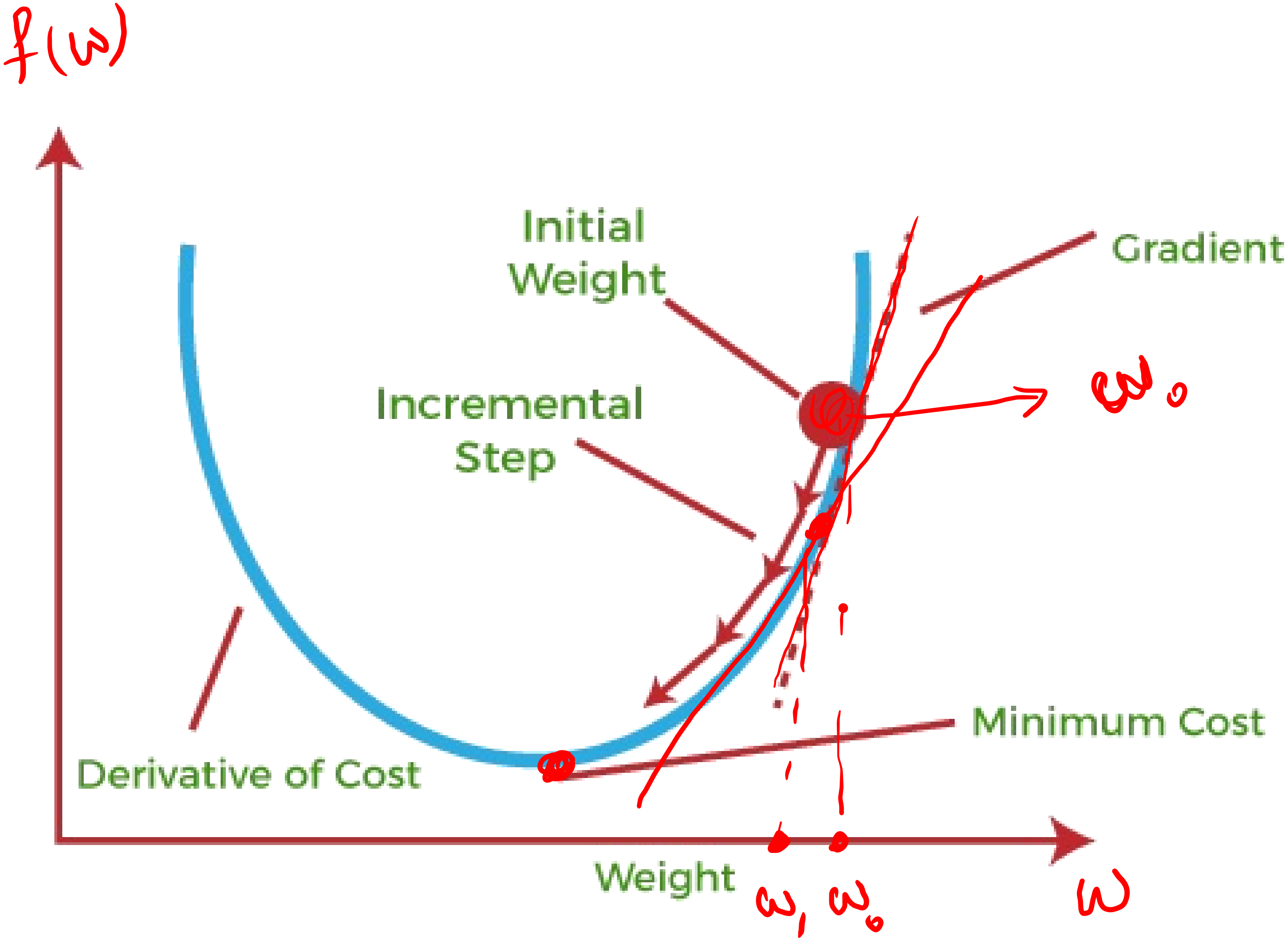
Gradient Descent

w_0

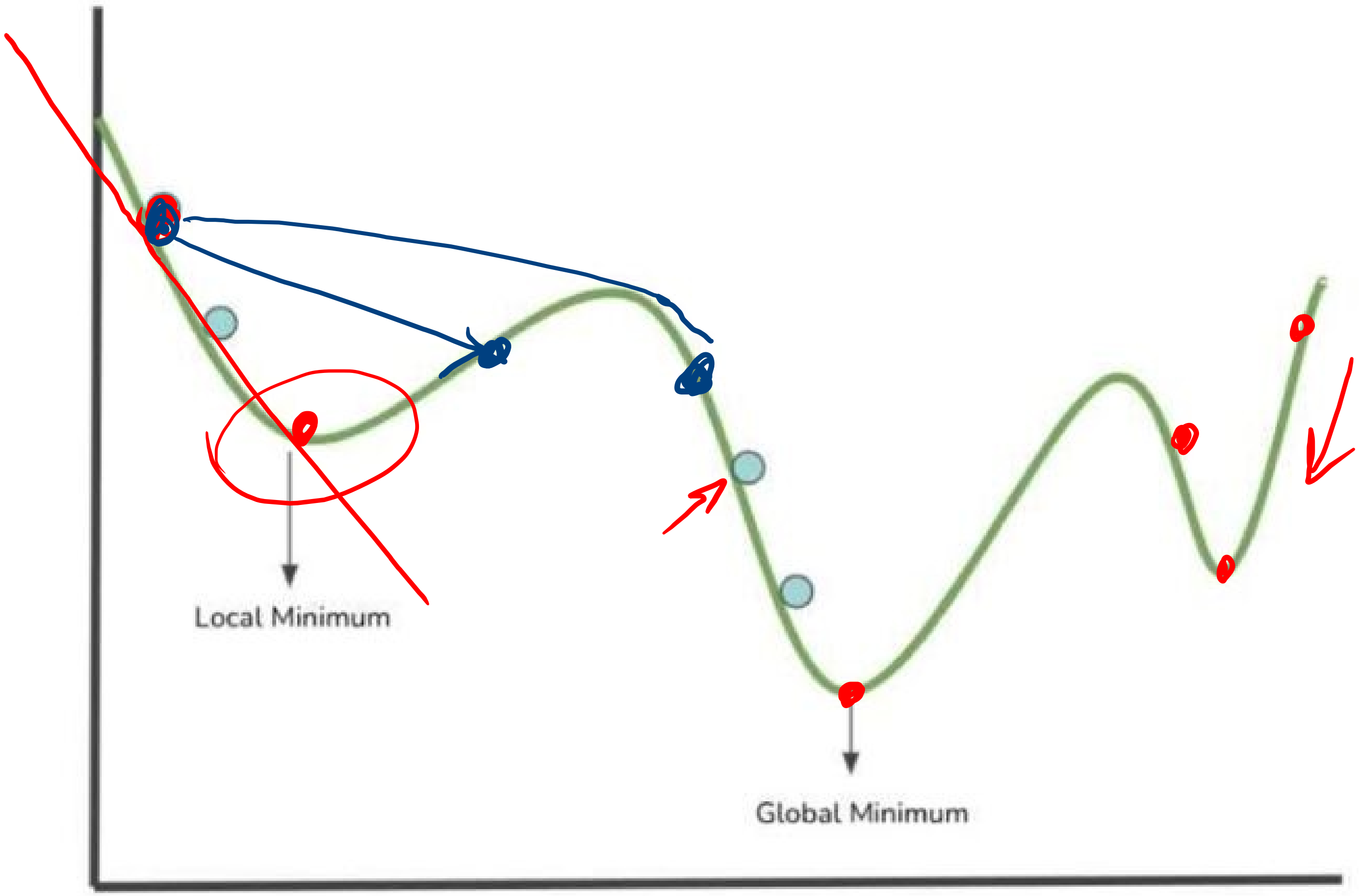
$$\boxed{w_{t+1} = w_t \overset{+}{-} \eta \, \underline{\nabla L(w_t)}}$$

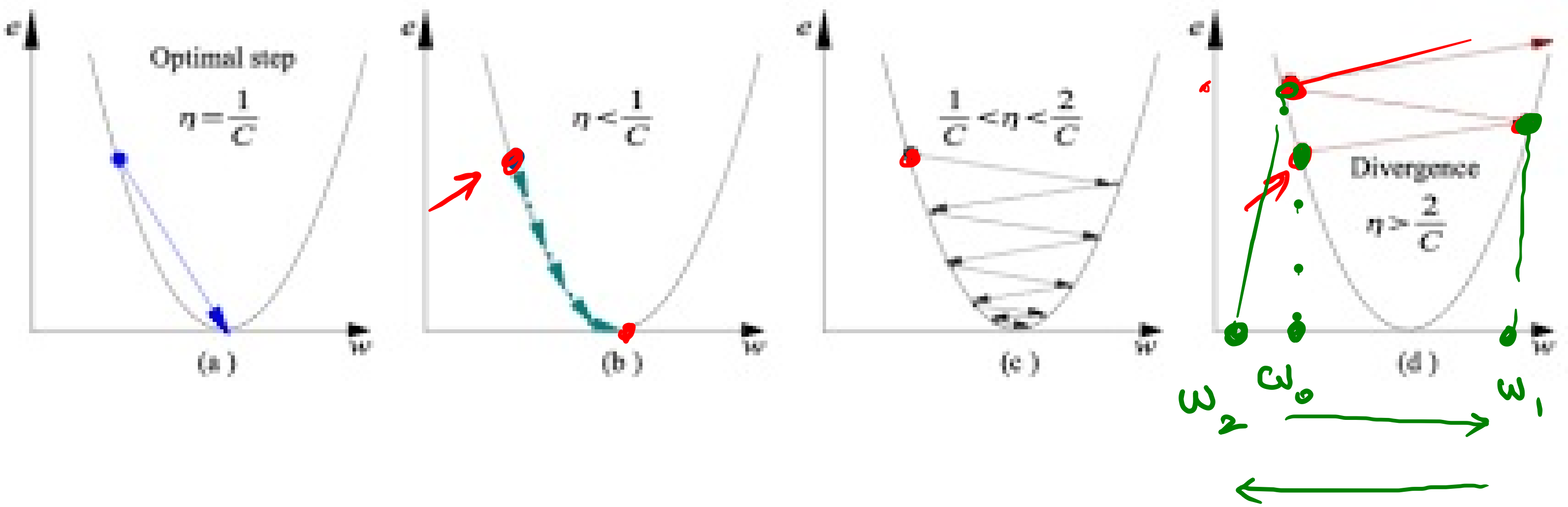
$w \in \mathbb{R}$

Learning Rate

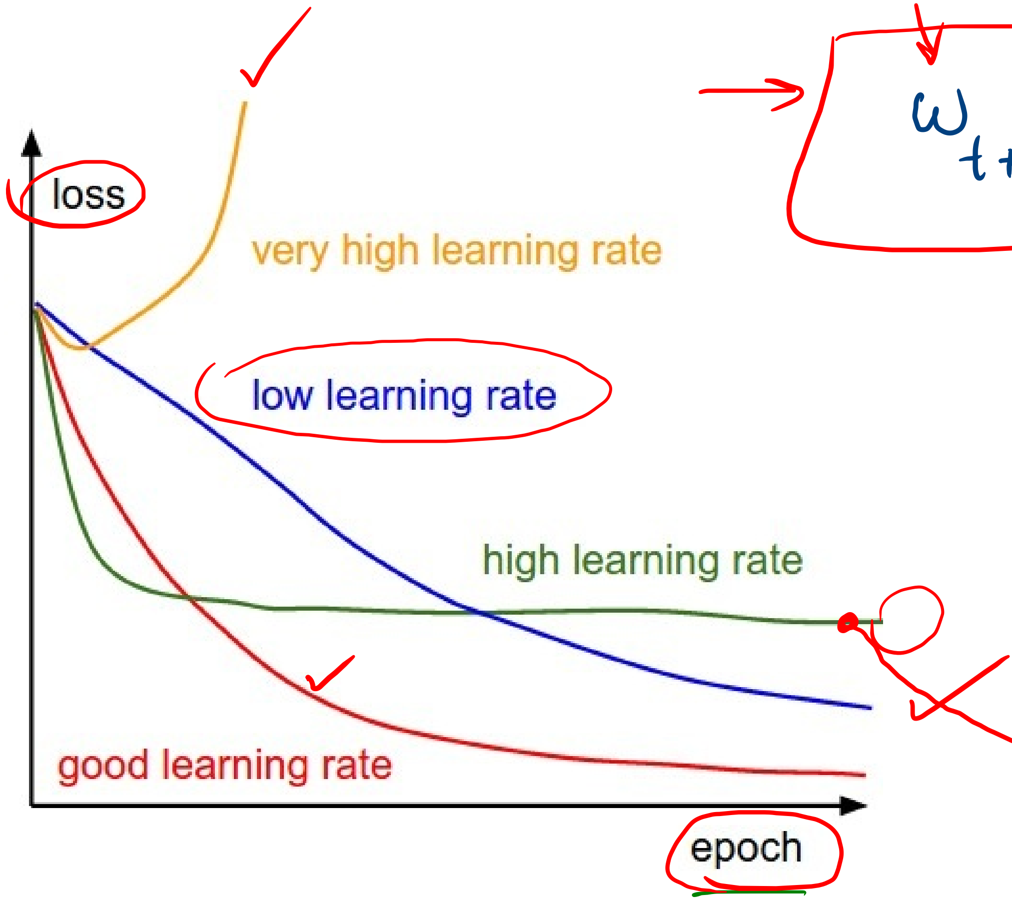


f(w)
Initial
Weight
Gradient
w0
Incremental
Step
Minimum Cost
Derivative of Cost
Weight
w1
w0
w


Local Minimum
Global Minimum



Optimal step
$\eta = \frac{1}{C}$
(a)
$\eta < \frac{1}{C}$
(b)
$\frac{1}{C} < \eta < \frac{2}{C}$
(c)
Divergence
$\eta > \frac{2}{C}$
(d)
e
w
w_2
w_0
w_1



$$\boxed{w_{t+1} = w_t - \eta \nabla_w L(w_t)}$$