Makine Öğrenmesi

2020

Perceptron

Gradyenli Alçalma

Ş. Sefa İşci

kave.bilgi.org.tr/

GRADYENLİ ALÇALMA

- Gradyenli (Dereceli) Alçalma
- Çok Etiketli Sınıflandırma / Regresyon
- Lojistik vs Perceptron – Neden Perceptron ?

Makine Öğrenmesi

Perceptron

Gradyenli (Dereceli) Alçalma

Genelleştirilmiş Lineer Modeller (GLM)

$$w := w - \alpha \frac{\partial J(w,b)}{\partial w} = w + \alpha \frac{1}{m} \sum_{i=1}^{m} x_{(i)} (y^{(i)} - a^{(i)}), \quad \frac{\partial J(w,b)}{\partial w} = dw = -\frac{1}{m} \sum_{i=1}^{m} x_{(i)} (y^{(i)} - a^{(i)}) = \frac{1}{m} \sum_{i=1}^{m} -dw_{(i)}$$

$$b := b - \alpha \frac{\partial J(w,b)}{\partial b} = b + \frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - a^{(i)}), \quad \frac{\partial J(w,b)}{\partial b} = db = -\frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - a^{(i)}) = \frac{1}{m} \sum_{i=1}^{m} -db_{(i)}$$

$dw_{(i)} = -x_{(i)} (y^{(i)} - a^{(i)}) \longrightarrow [n \times 1]$ kolon vektörü

$db_{(i)} = -(y^{(i)} - a^{(i)}) \longrightarrow [1 \times 1]$ skaler

$dw_{(i)} = x_{(i)} (a^{(i)} - y^{(i)}) = x_{(i)} (h_w(x^{(i)}) - y^{(i)})$

$db_{(i)} = a^{(i)} - y^{(i)} = h_w(x^{(i)}) - y^{(i)}$

$dw = -\frac{1}{m} x(y-a) = -\frac{1}{m} X^T (y-a)$

$db = -\frac{1}{m}$ np.sum$(y-a)$

$w := w - \alpha dw = w + \alpha \frac{1}{m} x(y-a) = w + \alpha \frac{1}{m} X^T (y-a)$

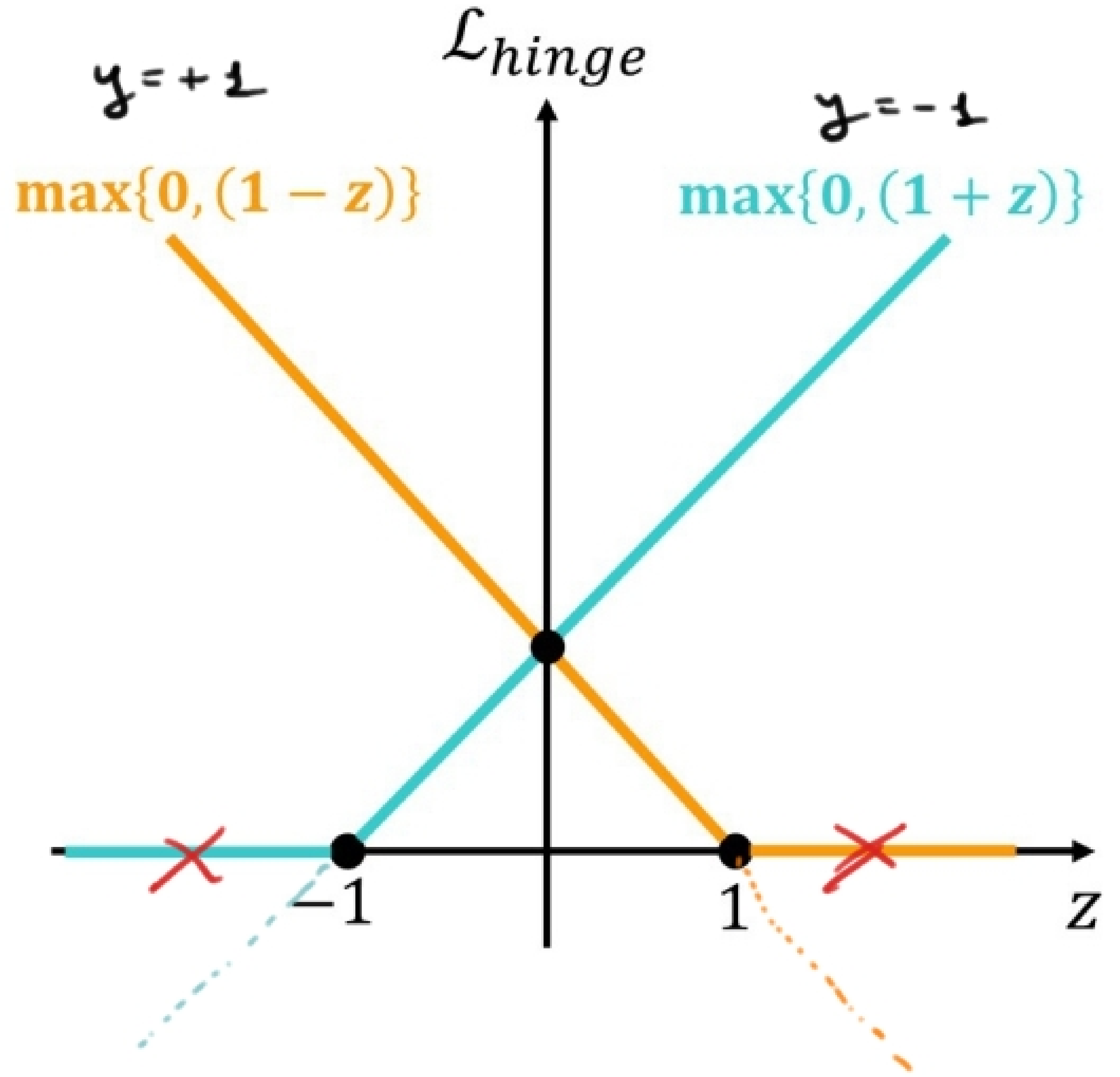
$b := b - \alpha db = b + \alpha \frac{1}{m}$ np.sum$(y-a)$

$a = \hat{y} = g(z) = h_w(x) = h(x)$

$z = Xw + b \rightarrow [m \times 1], \; w^T x + b \rightarrow [1 \times m]$

Gradyenli (Dereceli) Alçalma

-1 ve 1 durumu için : Tek örnekli $\rightarrow X \rightarrow [1 \times n] \rightarrow x \rightarrow [n \times 1]$



$$L(w,b) = \left(\frac{1+y}{2}\right)(1-z) + \left(\frac{1-y}{2}\right)(1+z)$$

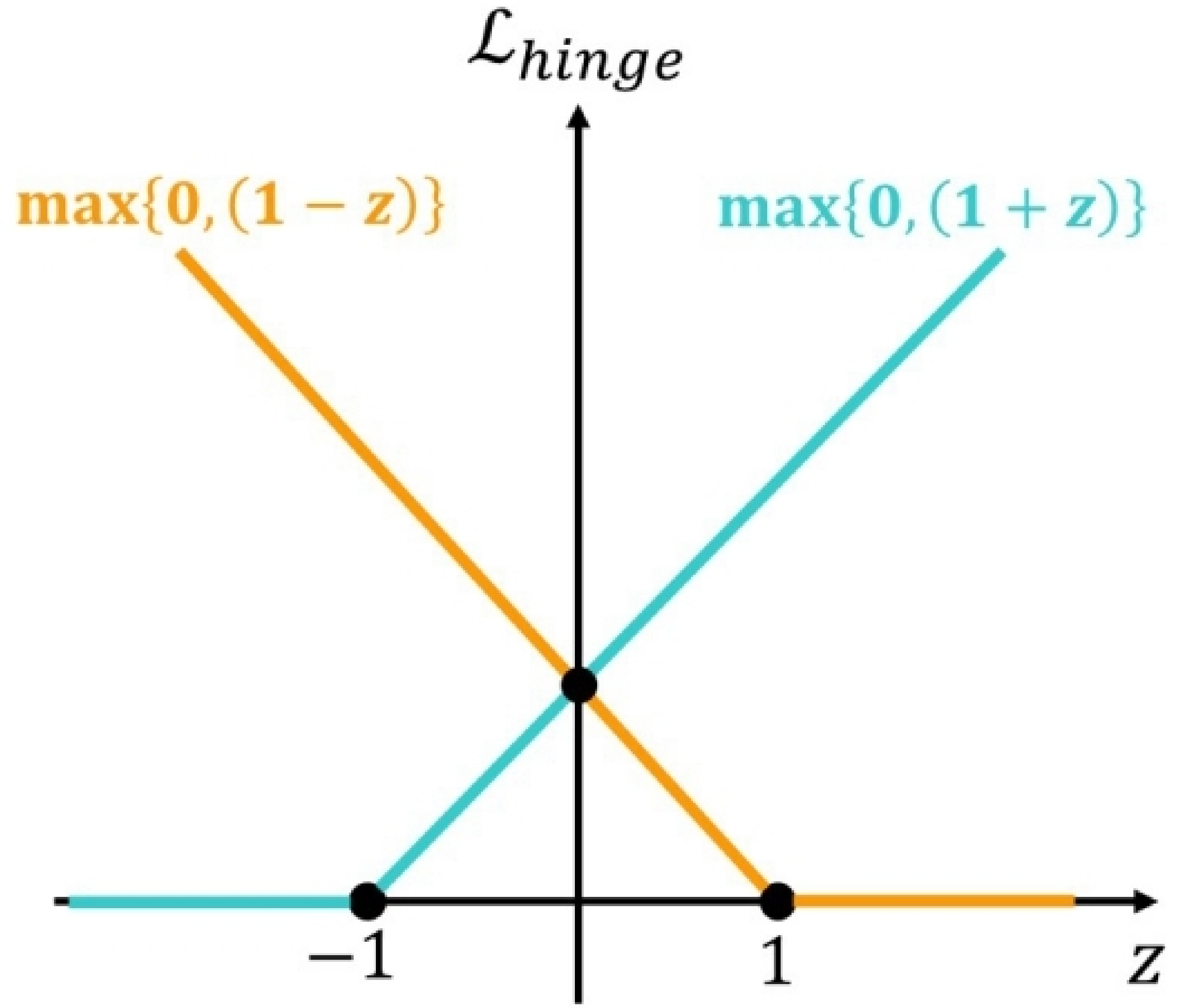
$$= \frac{1+y}{2} - z\frac{1+y}{2} + \frac{1-y}{2} + z\frac{1-y}{2}$$

$$= 1 + z(-y) = 1 - yz = 1 - y(w^Tx + b)$$

$$\frac{\partial L(w,b)}{\partial w} = -yx \qquad \frac{\partial L(w,b)}{\partial b} = -y$$

Gradyenli (Dereceli) Alçalma

-1 ve 1 durumu için : Tek örnekli $\rightarrow X \rightarrow [1 \times n] \rightarrow x \rightarrow [n \times 1]$



$\hat{y} = a$

$$\frac{\partial L(w,b)}{\partial w} = -yx \,, \quad \frac{\partial L(w,b)}{\partial b} = -y$$

Eğer $y \neq a$ ise;

$$w := w - \alpha \frac{\partial L(w,b)}{\partial w} = w + \alpha y x$$

$y = -1 \longrightarrow w := w - \alpha x$

$y = +1 \longrightarrow w := w + \alpha x$

$$b := b - \alpha \frac{\partial L(w,b)}{\partial b} = b + \alpha y$$

$y = -1 \longrightarrow b := b - \alpha$

$y = +1 \longrightarrow b := b + \alpha$

Gradyenli (Dereceli) Alçalma

-1 ve 1 durumu için : Çok örnekli $\rightarrow X \rightarrow [m \times n] \rightarrow x \rightarrow [n \times m]$

$$w \coloneqq w - \alpha \frac{\partial J(w,b)}{\partial w} \qquad \frac{\partial J(w,b)}{\partial w} = dw = -\frac{1}{m}\sum_{i=1}^{m} x_{(i)}\left(y^{(i)} - a^{(i)}\right) \qquad w \coloneqq w + \alpha \frac{1}{m}\sum_{i=1}^{m} x_{(i)}\left(y^{(i)} - a^{(i)}\right)$$

$$b \coloneqq b - \alpha \frac{\partial J(w,b)}{\partial b} \qquad \frac{\partial J(w,b)}{\partial b} = db = -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - a^{(i)}\right) \qquad b \coloneqq b + \alpha \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - a^{(i)}\right)$$

$dw_{(i)} = -x_{(i)}\left(y^{(i)} - a^{(i)}\right) \rightarrow [n \times 1]$ kolon vektörü

$db_{(i)} = -\left(y^{(i)} - a^{(i)}\right) \rightarrow$ skaler

$w := w - \alpha\, dw_{(i)} = w + \alpha x_{(i)}\left(y^{(i)} - a^{(i)}\right)$

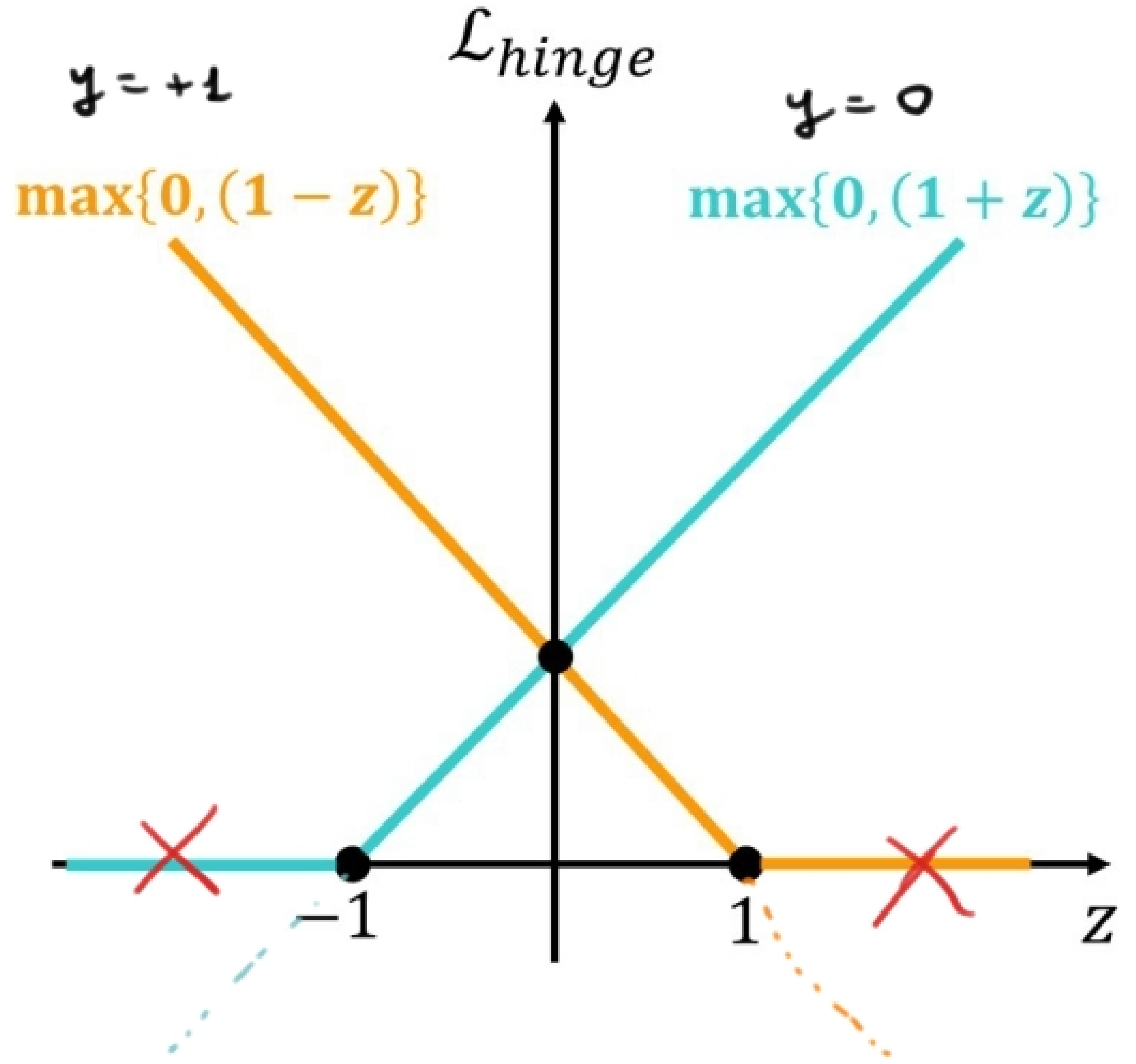
$\rightarrow y^{(i)} = a^{(i)} \rightarrow w := w$

$\rightarrow y^{(i)} = -1,\ a^{(i)} = +1 \rightarrow w := w - \alpha 2 x_{(i)}$

$\rightarrow y^{(i)} = +1,\ a^{(i)} = -1 \rightarrow w := w + \alpha 2 x_{(i)}$

Gradyenli (Dereceli) Alçalma

0 ve 1 durumu için : Tek örnekli $\rightarrow X \rightarrow [1 \times n] \rightarrow x \rightarrow [n \times 1]$



$$L(w,b) = y(1-z) + (1-y)(1+z)$$

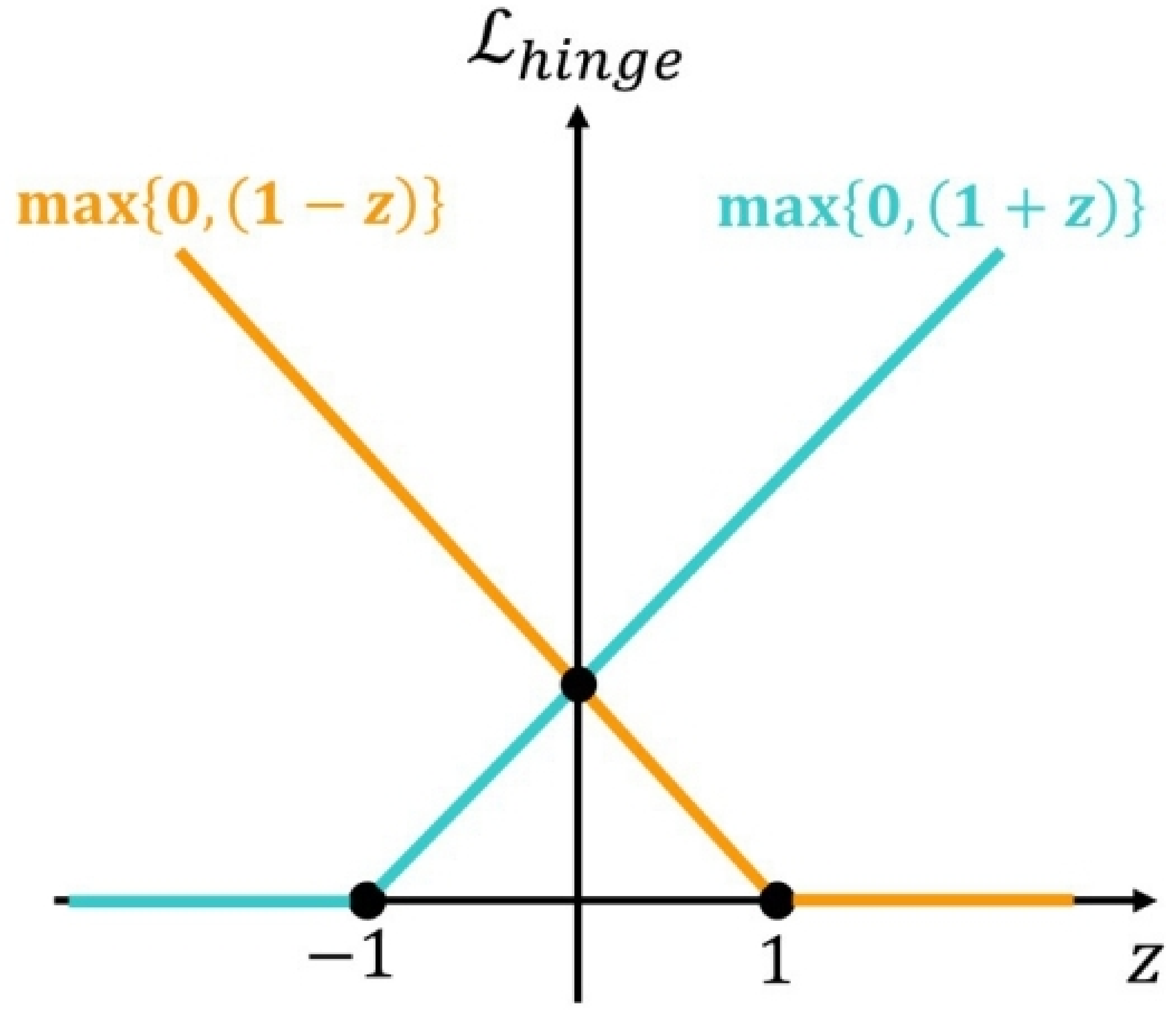
$$= \cancel{y} - yz + 1 + z - \cancel{y} - yz = 1 + z - 2yz = 1 + z(1-2y)$$

$$= 1 + (w^T x + b)(1-2y)$$

$$\frac{\partial L(w,b)}{\partial w} = x(1-2y) \qquad \frac{\partial L(w,b)}{\partial b} = 1-2y$$

Gradyenli (Dereceli) Alçalma

0 ve 1 durumu için : Tek örnekli $\rightarrow X \rightarrow [1 \times n] \rightarrow x \rightarrow [n \times 1]$



$\hat{y} = a$

Eğer $y \neq a$ ise;

$$w := w - \alpha x (1 - 2y) \begin{cases} y = 0 \longrightarrow w := w - \alpha x \\ y = 1 \longrightarrow w := w + \alpha x \end{cases}$$

$$b := b - \alpha (1 - 2y) \begin{cases} y = 0 \longrightarrow b := b - \alpha \\ y = 1 \longrightarrow b := b + \alpha \end{cases}$$

Gradyenli (Dereceli) Alçalma

0 ve 1 durumu için : Çok örnekli → $X \rightarrow [m \times n] \rightarrow x \rightarrow [n \times m]$

$$w \coloneqq w - \alpha \frac{\partial J(w,b)}{\partial w} \qquad \frac{\partial J(w,b)}{\partial w} = dw = -\frac{1}{m}\sum_{i=1}^{m} x_{(i)}\left(y^{(i)} - a^{(i)}\right) \qquad w \coloneqq w + \alpha \frac{1}{m}\sum_{i=1}^{m} x_{(i)}\left(y^{(i)} - a^{(i)}\right)$$

$$b \coloneqq b - \alpha \frac{\partial J(w,b)}{\partial b} \qquad \frac{\partial J(w,b)}{\partial b} = db = -\frac{1}{m}\sum_{i=1}^{m} \left(y^{(i)} - a^{(i)}\right) \qquad b \coloneqq b + \alpha \frac{1}{m}\sum_{i=1}^{m} \left(y^{(i)} - a^{(i)}\right)$$

$dw_{(i)} = -x_{(i)}\left(y^{(i)} - a^{(i)}\right)$ → $[n \times 1]$ kolon vektörü

$db_{(i)} = -\left(y^{(i)} - a^{(i)}\right)$ → skaler

$w := w - \alpha dw_{(i)} = w + \alpha x_{(i)}(y^{(i)} - a^{(i)})$

- $y^{(i)} = a^{(i)}$ → $w := w$
- $y^{(i)} = 0,\ a^{(i)} = 1$ → $w := w - \alpha x_{(i)}$
- $y^{(i)} = 1,\ a^{(i)} = 0$ → $w := w + \alpha x_{(i)}$

Gradyenli (Dereceli) Alçalma

Kodlama için $\{0, 1\}$ ve $\{-1, 1\}$ farketmeksizin genel prensip :

$w := np.zeros((n,1))$, $b = np.zeros(1)$

for epoch in epochs:

$dw := -\frac{1}{m} x(y-a) = -\frac{1}{m} X^T (y-a)$

$db := -\frac{1}{m} np.sum(y-a)$

$w := w - \alpha dw$

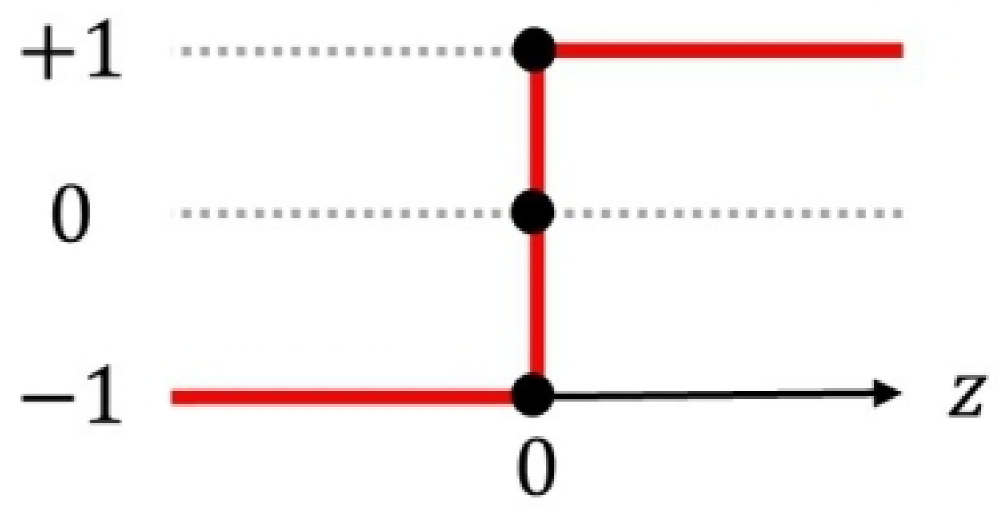
$b := b - \alpha db$

GLM

(Genelleştirilmiş Lineer Modeller)

(Generalized linear models)

Gradyenli (Dereceli) Alçalma



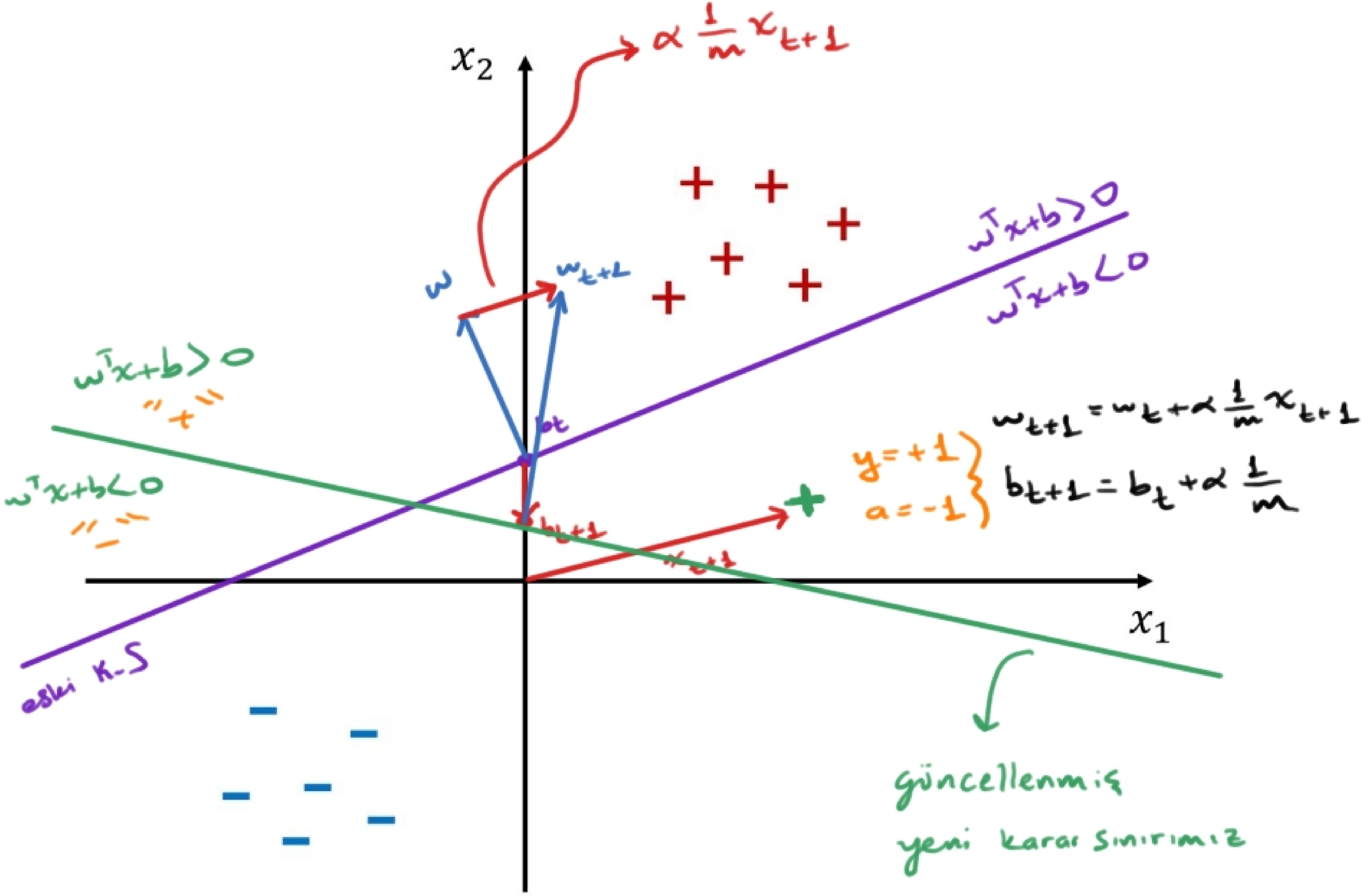
$$f(z) = \begin{cases} +1, & z \geq 0 \\ -1, & z < 0 \end{cases}$$

$$h_w(x) = \hat{y} = a = f(w^T x + b) \sim f(Xw + b)$$

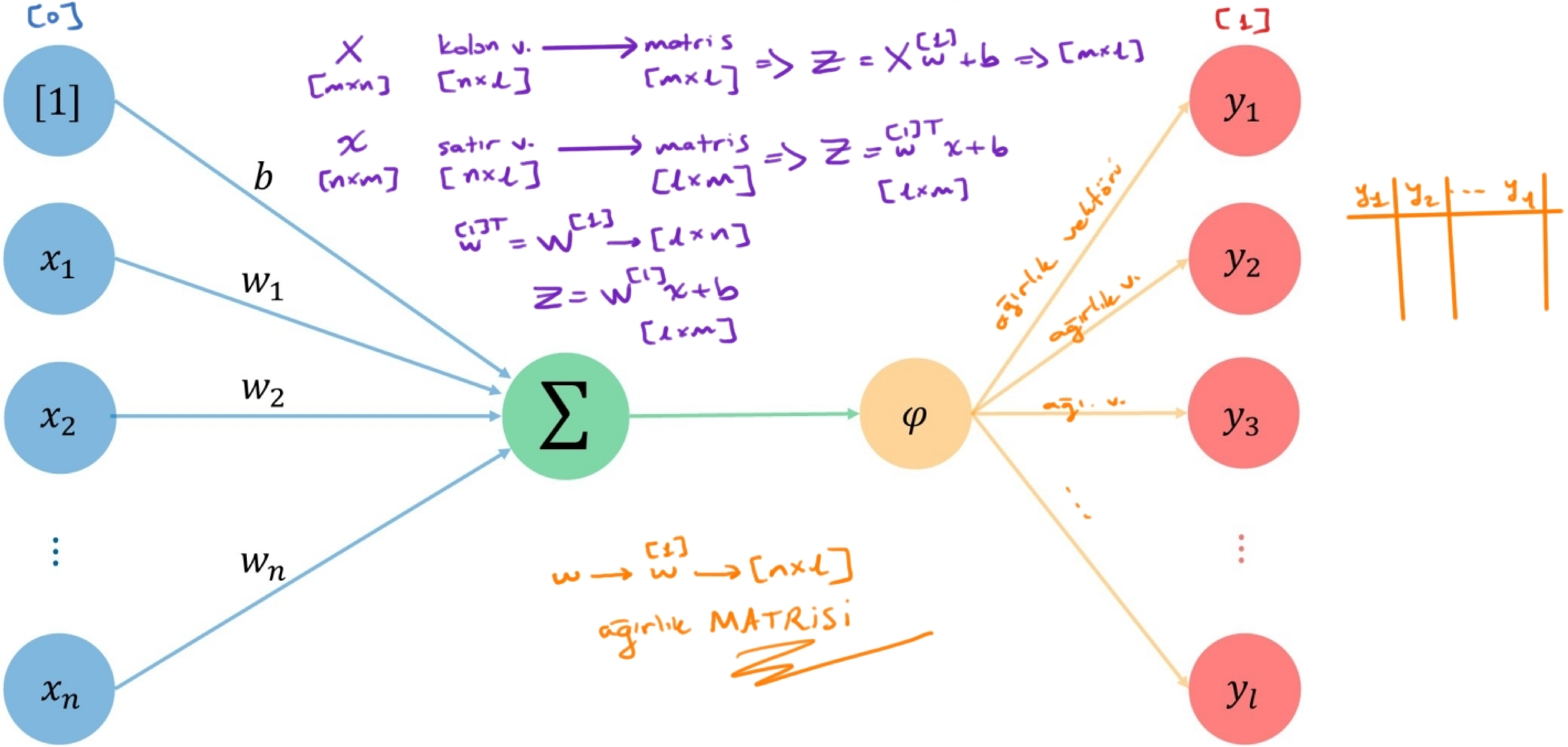
$$dw = -\frac{1}{m} x(y-a), \quad db = -\frac{1}{m} np.sum(y-a)$$

$$w := w - \alpha dw$$

$$b := b - \alpha db$$

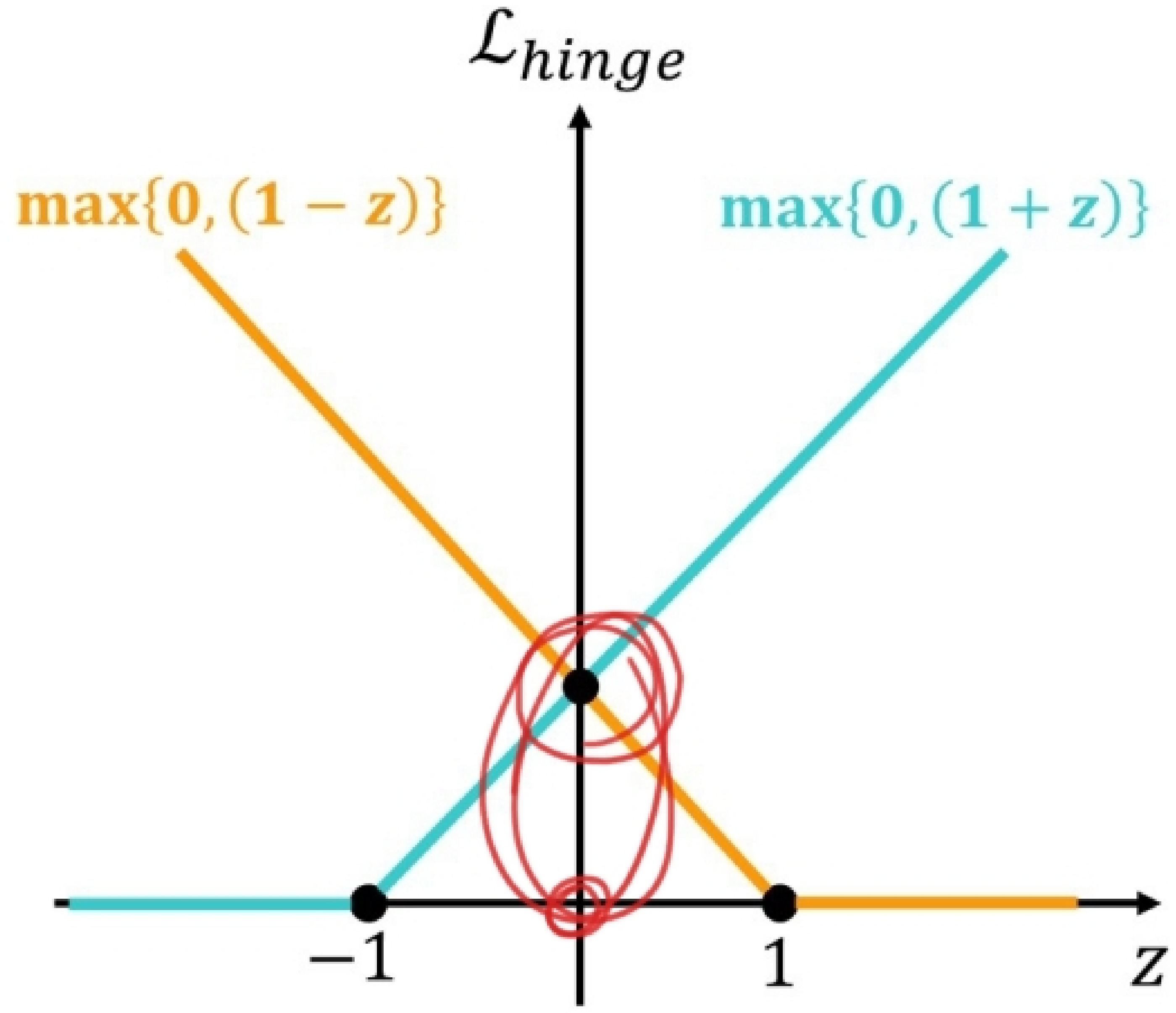


Çok Etiketli Sınıflandırma / Regresyon



Lojistik vs Perceptron - Neden Perceptron ?

Menteşe Kaybı (Hinge Loss)



$z = Xw + b$ & $z = w^T x + b$

lojistik, perceptron → $w := w - \alpha\, dw$, $b := b - \alpha\, db$

Sigmoid : $g(z) = \frac{1}{1+e^{-z}} = a = h(x)$

$w := w + \alpha \frac{1}{m} X^T (y - a)$

$b := b + \alpha \frac{1}{m}$ np.sum(y-a)

Lojistik daha iyi ✓ → YSA (ANN) ⇒ Son Katman : "sigmoid"

Perceptron → DVM (Destek Vektör Makineleri) ~ SVM

$w^T x + b > 0 \longrightarrow w^T x + b > +1 \longrightarrow$ pozitif marjin sınırı

$w^T x + b < 0 \longrightarrow w^T x + b < -1 \longrightarrow$ negatif marjin sınırı