Deep Learning HW1: Zahra Mohammadi

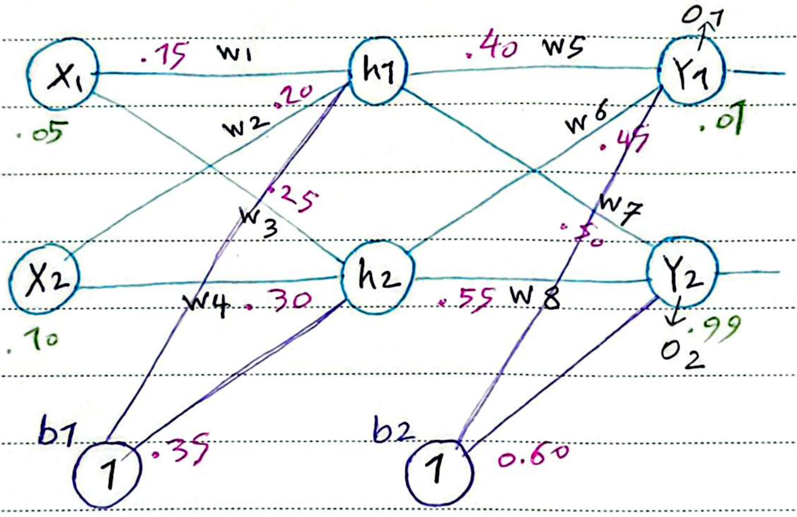
Subject:
Date

۱_ شبکه عصبی زیر را در نظر بگیرید. این شبکه دو خروجی Y_1 و Y_2 دارد که بعد از train مقادیر آن ها باید به ترتیب ۰۱. و ۰۲. باشد.

تابع loss را به صورت زیر تعریف می‌کنیم:

$$Loss = \frac{1}{2}\sum_{i=1}^{2}(Y_i-\hat{Y}_i)^2$$

تمام activation function ها sigmoid می‌باشند. مقدار learning rate را 0.5 فرض کنید. برای یک تکرار، مقادیر w و b را آپدیت کنید.



ورودی ها $\begin{cases} x_1 = 0.05 \\ x_2 = 0.10 \end{cases}$

خروجی های مطلوب یا Label $\begin{cases} y_1 = 0.01 \\ y_2 = 0.02 \end{cases}$

$\eta = 0.5$

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

$$L = \frac{1}{2}\left[(y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2\right]$$

$$z_{h_1} = x_1 w_1 + x_2 w_2 + b_1 = 0.05 \times 0.15 + 0.1 \times 0.2 + 0.35$$

$$z_{h_1} = 0.0075 + 0.02 + 0.35 = 0.377$$

$$e^{-0.378} \simeq 0.686$$

$$h_1 = \sigma(z_{h_1}) = \frac{1}{1+e^{-0.378}} = \frac{1}{1.686} \simeq 0.593$$

Subject:
Date

o_1, o_2 } مجموع وزن دار قبل از سیگموید در لایهٔ خروجی

y_1, $\hat{y}_2$ } خروجی شبکه بعد از سیگموید

$$z_{h2} = x_1 w_3 + x_2 w_4 + b_1 = 0.05 \times 0.25 + 0.1 \times 0.3 + 0.35$$

$$z_{h2} = 0.0125 + 0.03 + 0.35 \simeq 0.392$$

$$e^{-0.392} \simeq 0.676$$

$$h_2 = \sigma(z_{h2}) = \frac{1}{1 + e^{-0.392}} = \frac{1}{1.676} = 0.597$$

$$o_1 = h_1 w_5 + h_2 w_6 + b_2 = 0.593 \times 0.4 + 0.597 \times 0.45 + 0.6$$

$$o_1 = 0.237 + 0.269 + 0.6 = 1.106$$

$$e^{-1.106} \simeq 0.331$$

$$\hat{Y}_1 = \sigma(o_1) = \frac{1}{1 + e^{-1.106}} = \frac{1}{1.331} \simeq 0.751$$

$$o_2 = h_1 w_7 + h_2 w_8 + b_2 = 0.593 \times 0.5 + 0.597 \times 0.55 + 0.6$$

$$o_2 = 0.297 + 0.328 + 0.6 = 1.225$$

$$e^{-1.225} \simeq 0.294$$

$$\hat{Y}_2 = \sigma(o_2) = \frac{1}{1 + e^{-1.225}} = \frac{1}{1.294} = 0.773$$

$$y_1 - \hat{y}_1 = 0.01 - 0.751 = -0.741$$

$$y_2 - \hat{y}_2 = 0.02 - 0.773 = -0.753$$

$$L = \frac{1}{2}\left[(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\right] = \frac{1}{2}\left[(-0.741)^2 + (-0.753)^2\right]$$

$$L = \frac{1}{2}(0.549 + 0.567) = \frac{1}{2} \times 1.116 \simeq 0.558$$

$L = \frac{1}{2}(y_1-\hat{y}_1)^2 + \frac{1}{2}(y_2-\hat{y}_2)^2$ $\qquad L_1 = \frac{1}{2}(y_1-\hat{y}_1)^2$

$\frac{\partial L_1}{\partial \hat{y}_1} = \frac{1}{2} \times 2 \times (y_1-\hat{y}_1) \times -1 = \hat{y}_1 - y_1 \rightarrow \frac{\partial L}{\partial \hat{y}_1} = \hat{y}_1 - y_1$

$\frac{\partial L}{\partial \hat{y}_1} = 0.751 - 0.01 = 0.741$

$\frac{\partial L}{\partial \hat{y}_2} = \hat{y}_2 - y_2 = 0.773 - 0.02 = 0.753$

$\hat{y}_1 = \sigma(o_1)$ $\qquad \hat{y}_2 = \sigma(o_2)$ $\qquad \sigma(z) = \frac{1}{1+e^{-z}}$

$\frac{d}{dz}\sigma(z) = \sigma(z).(1-\sigma(z))$

$\frac{\partial \hat{y}_1}{\partial o_1} = \hat{y}_1(1-\hat{y}_1) = 0.751 \times (1-0.751) = 0.751 \times 0.249 \simeq 0.187$

$\frac{\partial \hat{y}_2}{\partial o_2} = \hat{y}_2(1-\hat{y}_2) = 0.773 \times (1-0.773) = 0.773 \times 0.227 \simeq 0.176$

$o_1 \rightarrow \hat{y}_1 \rightarrow L$ $\qquad dz_{o_1} = \frac{\partial L}{\partial o_1} = \frac{\partial L}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial o_1}$

$dz_{o_1} = (\hat{y}_1 - y)\cdot \hat{y}_1(1-\hat{y}_1) = 0.741 \times 0.187 \simeq 0.138$

$dz_{o_2} = \frac{\partial L}{\partial o_2} = \frac{\partial L}{\partial \hat{y}_2} \cdot \frac{\partial \hat{y}_2}{\partial o_2} = (\hat{y}_2 - y)\cdot \hat{y}_2(1-\hat{y}_2)$

$dz_{o_2} = 0.753 \times 0.176 = 0.132$

$\boxed{o_1 = h_1 w_5 + h_2 w_6 + b_2} \longrightarrow \frac{\partial o_1}{\partial w_5} = h_1$

$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial w_5} = dz_{o_1} \cdot h_1 = 0.138 \times 0.593 = 0.082$

Subject:
Date

$\frac{\partial o_1}{w_6} = h_2$ $\qquad$ $\frac{\partial L}{\partial w_6} = dz_{o_1} \cdot h_2 = 0.138 \times 0.597 = 0.083$

$\boxed{o_2 = h_1 w_7 + h_2 w_8 + b_2}$ $\qquad$ $\frac{\partial o_2}{\partial w_7} = h_1$, $\frac{\partial o_2}{\partial w_8} = h_2$

$$\frac{\partial L}{\partial w_7} = \frac{\partial L}{\partial o_2} \cdot \frac{\partial o_2}{\partial w_7} = dz_{o_2} \cdot h_1 = 0.132 \times 0.593 \simeq 0.078$$

$$\frac{\partial L}{\partial w_8} = \frac{\partial L}{\partial o_2} \cdot \frac{\partial o_2}{\partial w_8} = dz_{o_2} \cdot h_2 = 0.132 \times 0.597 \simeq 0.079$$

$\frac{\partial o_1}{\partial b_2} = 1$, $\frac{\partial o_2}{\partial b_2} = 1$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial b_2} + \frac{\partial L}{\partial o_2} \cdot \frac{\partial o_2}{\partial b_2} = \frac{\partial L}{\partial o_1} + \frac{\partial L}{\partial o_2} = dz_{o_1} + dz_{o}$$

$$\frac{\partial L}{\partial b_2} = 0.138 + 0.132 = 0.270$$

$h_1 \rightarrow o_1 \rightarrow \hat{y}_1 \rightarrow L$ $\qquad$ $h_1 \rightarrow o_2 \rightarrow \hat{y}_2 \rightarrow L$

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial o_1} \cdot \overset{w_5}{\frac{\partial o_1}{\partial h_1}} + \frac{\partial L}{\partial o_2} \cdot \overset{w_7}{\frac{\partial o_2}{\partial h_1}}$$

$$\frac{\partial L}{\partial h_1} = 0.138 \times 0.4 + 0.132 \times 0.5 = 0.055 + 0.066 = 0.121$$

$$\frac{\partial L}{\partial h_2} = \frac{\partial L}{\partial o_1} \cdot \overset{w_6}{\frac{\partial o_1}{\partial h_2}} + \frac{\partial L}{\partial o_2} \cdot \overset{w_8}{\frac{\partial o_2}{\partial h_2}}$$

$$\frac{\partial L}{\partial h_2} = 0.138 \times 0.45 + 0.132 \times 0.55 = 0.062 + 0.073 = 0.135$$

$$h_1 = \sigma(z_{h1}) \longrightarrow \frac{\partial h_1}{\partial z_{h1}} = h_1(1-h_1)$$

$$dz_{h1} = \frac{\partial L}{\partial z_{h1}} = \frac{\partial L}{\partial h_1} \cdot h_1(1-h_1) = 0.593 \times \overset{0.407}{(1-0.593)} = 0.241$$

$$dz_{h1} = 0.121 \times 0.241 \simeq 0.029$$

$$dz_{h2} = \frac{\partial L}{\partial h_2} \cdot h_2(1-h_2) = 0.597 \cdot \overset{0.403}{(1-0.597)} = 0.241$$

$$dz_{h2} = 0.135 \times 0.241 = 0.032$$

$$z_{h1} = x_1 w_1 + x_2 w_2 + b_1 \longrightarrow \frac{\partial z_{h1}}{\partial w_1} = x_1 \quad , \quad \frac{\partial z_{h1}}{\partial w_2} = x_2$$

$$\frac{\partial L}{\partial w_1} = dz_{h1} \cdot x_1 = 0.029 \times 0.05 = 0.001$$

$$\frac{\partial L}{\partial w_2} = dz_{h1} \cdot x_2 = 0.029 \times 0.1 = 0.003$$

$$z_{h2} = x_1 w_3 + x_2 w_4 + b_1 \longrightarrow \frac{\partial z_{h2}}{\partial w_3} = x_1 \quad , \quad \frac{\partial z_{h2}}{\partial w_4} = x_2$$

$$\frac{\partial L}{\partial w_3} = dz_{h2} \cdot x_1 = 0.032 \times 0.05 = 0.002$$

$$\frac{\partial L}{\partial w_4} = dz_{h2} \cdot x_2 = 0.032 \times 0.1 = 0.003$$

$$\frac{\partial L}{\partial b_1} = dz_{h1} + dz_{h2} = 0.029 + 0.032 = 0.061$$

آپدیت وزن ها و بایاس ها :

پارامتر جدید = پارامتر قدیم − $\eta\left(\frac{\partial Loss}{\partial \text{پارامتر}}\right)$

$w'_1 = 0.15 - 0.5 \times 0.001 = 0.149$

$w'_2 = 0.2 - 0.5 \times 0.003 = 0.199$

$w'_3 = 0.25 - 0.5 \times 0.002 = 0.249$

$w'_4 = 0.3 - 0.5 \times 0.003 = 0.298$

$b'_1 = 0.35 - 0.5 \times 0.061 = 0.319$

$w'_5 = 0.4 - 0.5 \times 0.082 = 0.359$

$w'_6 = 0.45 - 0.5 \times 0.083 = 0.409$

$w'_7 = 0.5 - 0.5 \times 0.078 = 0.5 - 0.039 = 0.461$

$w'_8 = 0.55 - 0.5 \times 0.079 = 0.55 - 0.04 = 0.511$

$b'_2 = 0.6 - 0.5 \times 0.271 = 0.6 - 0.136 = 0.465$

2_ شبکه عصبی زیر را در نظر بگیرید و فرض کنید:

$$f = w_1^{[3]} a_1^{[2]} + w_2^{[3]} a_2^{[2]}$$

مقادیر زیر را محاسبه کنید:

الف) $\delta_1 = \frac{\partial f(x)}{\partial z_1^{[2]}}$

ب) $\delta_2 = \frac{\partial f(x)}{\partial z_2^{[2]}}$

ج) $\delta_3 = \frac{\partial f(x)}{\partial z_1^{[1]}}$

د) $\delta_4 = \frac{\partial f(x)}{\partial w_{11}^{[1]}}$

(شکل و بقیه اطلاعات در فایل)

$$f(x) = w_1^{[3]} a_1^{[2]} + w_2^{[3]} a_2^{[2]}$$ (الف

$$\frac{\partial f}{\partial a_1^{[2]}} = w_1^{[3]} \cdot \frac{\partial a_1^{[2]}}{\partial a_1^{[2]}} + w_2^{[3]} \cdot 0 = w_1^{[3]}$$ ← $a_2^{[2]}$ مستقل از $a_1^{[2]}$ است

$$\delta_1 = w_1^{[3]}$$

(ب

$$z_1^{[2]} \longrightarrow a_1^{[2]} \longrightarrow f$$

$$\frac{\partial f}{\partial z_1^{[2]}} = \underbrace{\frac{\partial f}{\partial a_1^{[2]}}}_{w_1^{[2]}} \cdot \frac{\partial a_1^{[2]}}{\partial z_1^{[2]}}$$

$$\frac{\partial a_1^{[2]}}{\partial z_1^{[2]}} = \sigma(z_1^{[2]}) = \sigma(z_1^{[2]})(1 - \sigma(z_1^{[2]})) = a_1^{[2]}(1 - a_1^{[2]})$$

$$\delta_2 = \frac{\partial f}{\partial z_1^{[2]}} = w_1^{[3]} \cdot a_1^{[2]}(1 - a_1^{[2]})$$

(ج

$$z_1^{[1]} \longrightarrow a_1^{[1]} \longrightarrow z_1^{[2]} \longrightarrow a_1^{[2]} \longrightarrow f$$

$$z_1^{[1]} \longrightarrow a_1^{[1]} \longrightarrow z_2^{[2]} \longrightarrow a_2^{[2]} \longrightarrow f$$

$$\frac{\partial f}{\partial z_1^{[1]}} = \frac{\partial f}{\partial a_1^{[1]}} \cdot \frac{\partial a_1^{[1]}}{\partial z_1^{[1]}} \qquad \delta_3 = \frac{\partial f}{\partial z_1^{[1]}}$$

$$z_1^{[2]} = w_{11}^{[2]} a_1^{[1]} + w_{12}^{[2]} a_2^{[1]} \longrightarrow \frac{\partial z_1^{[2]}}{\partial a_1^{[1]}} = w_{11}^{[2]}$$

$$z_2^{[2]} = w_{21}^{[2]} a_1^{[1]} + w_{22}^{[2]} a_2^{[1]} \longrightarrow \frac{\partial z_2^{[2]}}{\partial a_2^{[1]}} = w_{21}^{[2]}$$

Subject:
Date

$$\frac{\partial f}{\partial a_1^{[1]}} = \frac{\partial f}{\partial z_1^{[2]}} \frac{\partial z_1^{[2]}}{\partial a_1^{[1]}} + \frac{\partial f}{\partial z_2^{[2]}} \frac{\partial z_2^{[2]}}{\partial a_1^{[1]}} \qquad \frac{\partial f}{\partial z_1^{[2]}} = \delta_2$$

$$\frac{\partial f}{\partial a_2^{[2]}} = w_2^{[3]} \qquad \frac{\partial a_2^{[2]}}{\partial z_2^{[2]}} = a_2^{[2]} (1 - a_2^{[2]})$$

$$\frac{\partial f}{\partial z_2^{[2]}} = \frac{\partial f}{\partial a_2^{[2]}} \cdot \frac{\partial a_2^{[2]}}{\partial z_2^{[2]}} = w_2^{[3]} \cdot a_2^{[2]} \cdot (1 - a_2^{[2]})$$

$$\frac{\partial f}{\partial a_1^{[1]}} = \delta_2 . w_{11}^{[2]} + (w_2^{[3]} a_2^{[2]} (1 - a_2^{[2]})) . w_{21}^{[2]}$$

$$\frac{\partial a_1^{[1]}}{\partial z_1^{[1]}} = a_1^{[1]} (1 - a_1^{[1]})$$

$$\delta_3 = \frac{\partial f}{\partial z_1^{[1]}} = \frac{\partial f}{\partial a_1^{[1]}} \cdot \frac{\partial a_1^{[1]}}{\partial z_1^{[1]}}$$

$$\delta_3 = [\delta_2 w_{11}^{[2]} + w_2^{[3]} a_2^{[2]} (1 - a_2^{[2]}) w_{21}^{[2]}] . a_1^{[1]} (1 - a_1^{[1]})$$

$$\delta_2 = w_1^{[3]} . a_1^{[2]} (1 - a_1^{[2]})$$

$$\delta_3 = a_1^{[1]} (1 - a_1^{[1]}) (w_{11}^{[2]} w_1^{[3]} a_1^{[2]} (1 - a_1^{[2]}) + w_{21}^{[2]} w_2^{[3]} a_2^{[2]} (1 - a_2^{[2]}))$$

$$z_1^{[1]} = w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2 \qquad \frac{\partial z_1^{[1]}}{\partial w_{11}^{[1]}} = x_1$$

$$\frac{\partial f}{\partial z_1^{[1]}} = \delta_3 \qquad \delta_4 = \frac{\partial f}{\partial w_{11}^{[1]}} = \frac{\partial f}{\partial z_1^{[1]}} \frac{\partial z_1^{[1]}}{\partial w_{11}^{[1]}} = \delta_3 . x_1$$

$$\delta_4 = x_1 \delta_3$$

$$\delta_4 = x_1 a_1^{[1]} (1 - a_1^{[1]}) (w_{11}^{[2]} w_1^{[3]} a_1^{[2]} (1 - a_1^{[2]}) + w_{21}^{[2]} w_2^{[3]} a_2^{[2]} (1 - a_2^{[2]}))$$

3) یک مسئله طبقه بندی دودویی را می توان با دو رویکرد زیر حل کرد:

رویکرد 1: رگرسیون لجستیک ساده (یک نورون)

خروجی شما بصورت زیر است: $\hat{y} = \sigma(w_1 x + b_1)$

اگر $\hat{y} \leq 0.5$ باشد ، به عنوان 0 طبقه بندی می‌شود ، و در غیر اینصورت به عنوان 1.

رویکرد ۲: رگرسیون softmax ساده (دو نورون)

خروجی شما بصورت زیر است:

$$\hat{y} = \mathrm{softmax}(w_s x + b_s) = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$$

اگر $\hat{y}_1 > \hat{y}_2$ باشد ، به عنوان 0 طبقه بندی می‌شود و در غیر اینصورت به عنوان 1.

رویکرد 1: رگرسیون لجستیک ساده

$$\hat{y} = \sigma(w_1 x + b_1)$$

$z_1 = w_1 x + b_1$ $\qquad \hat{y} = \sigma(z_1) = \frac{1}{1+e^{-z_1}}$

اگر $\hat{y} \leq 0.5 \longrightarrow$ کلاس 0 ، در غیر اینصورت $\longrightarrow$ کلاس 1

$\sigma(0) = 0.5$ تابع سیگموید صعودی است

$$\hat{y} \leq 0.5 \longleftrightarrow z_1 \leq 0$$

$$\hat{y} > 0.5 \longleftrightarrow z_1 > 0$$

اگر $w_1 x + b_1 \leq 0 \longrightarrow$ کلاس 0

اگر $w_1 x + b_1 > 0$ کلاس 1

$w_1 x + b_1 = 0 \longleftarrow$ مرز تقسیم

2

رویکرد ۲ : رگرسیون softmax با دو نورون

$$\hat{y} = \text{softmax}(W_s X + b_s) = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$$

$$\begin{cases} z_1 = w_1^T x + b_1 \\ z_2 = w_2^T x + b_2 \end{cases} \Leftarrow$$

سطر اول W_s : بردار وزن نورون اول : w_1

سطر دوم W_s : بردار وزن نورون دوم : w_2

بایاس ها : نورن ۱ : b_1 ، نورون ۲ : b_2

softmax :

$$\hat{y}_1 = \frac{e^{z_1}}{e^{z_1} + e^{z_2}} \quad , \quad \hat{y}_2 = \frac{e^{z_2}}{e^{z_1} + e^{z_2}}$$

اگر $\hat{y}_1 \geq \hat{y}_2$ → کلاس ۰ ، در غیر اینصورت → کلاس ۱

$$\hat{y}_1 \geq \hat{y}_2 \longrightarrow \frac{e^{z_1}}{e^{z_1} + e^{z_2}} \geq \frac{e^{z_2}}{e^{z_1} + e^{z_2}} \longrightarrow e^{z_1} \geq e^{z_2}$$

$$\longrightarrow z_1 \geq z_2$$

$$w_1^T x + b_1 \geq w_2^T x + b_2 \longrightarrow (w_1^T - w_2^T) x + (b_1 - b_2) \geq 0$$

$$\rightarrow (w_1 - w_2)^T x + (b_1 - b_2) \geq 0$$

اگر $(w_1 - w_2)^T x + (b_1 - b_2) \geq 0$ → کلاس ۰ ، در غیر اینصورت → کلاس ۱

$(w_1 - w_2)^T x + (b_1 - b_2) = 0$ ← مرز تصمیم

رویکرد ۲ مدل پیچیده‌تری یاد می‌گیرد، هر مدلی که با softmax (دو نورون) ساخته می‌شود، با یک لجستیک تک نورونی هم قابل پیاده‌سازی است. یعنی از لحاظ مرز تصمیم هر دو فقط یک مرز خطی دارند، فقط در رویکرد ۲ پارامتر اضافه و تکراری داریم.

معادل‌سازی دو رویکرد: می‌خواهیم رویکرد ۲ و ۱ همان مرز تصمیم را ایجاد کنند، یعنی مجموعه نقاطی که در رویکرد ۲ کلاس ۰ می‌شوند، دقیقاً همان‌هایی باشند که در رویکرد ۱ کلاس ۰ می‌شوند.

$$\begin{cases} (w_1-w_2)^T x+(b_1-b_2)\geq 0 & \text{کلاس ۰ از رویکرد ۲:} \\ w_1 x+b_1\leq 0 & \text{کلاس ۰ از رویکرد ۱:} \end{cases}$$

این دوتا را برابر هم می‌خواهیم $\longrightarrow w_1 x+b_1=-\left[(w_1-w_2)^T x+(b_1-b_2)\right]$

$$w_1=-(w_1-w_2)=w_2-w_1$$
$$b_1=-(b_1-b_2)=b_2-b_1$$
$$\longrightarrow \begin{cases} w_1=w_2-w_1 \\ b_1=b_2-b_1 \end{cases}$$

هر انتخابی از bs ، ws (یعنی w_1 ، w_2 ، b_1 ، b_2) را می‌توان با انتخاب مناسب $w_1=w_2-w_1$ و $b_1=b_2-b_1$ به یک مدل لجستیک تک نورونی معادل تبدیل کرد به‌طوری که قانون تصمیم زیر در دو رویکرد دقیقاً یکی باشد:

$$\text{sofmax}: \quad \hat{y}_1\geq \hat{y}_2 \longleftrightarrow (w_1-w_2)^T x+(b_1-b_2)\geq 0$$

$$\text{Logistic}: \quad \hat{y}\leq 0.5 \longleftrightarrow (w_2-w_1)^T x+(b_2-b_1)\leq 0$$

این دو ناحیه تصمیم یکسان هستند. بنابراین رویکرد ۲ فقط پارامتر اضافه و هم‌ارز دارد ولی قدرت مدلش برای طبقه‌بندی دودویی از رویکرد ۱ بیشتر نیست.