نام و نام خانوادگی:

رضا قربانی پاجی

شماره دانشجویی:

403206565

تمرین سوم درس یادگیری ماشین

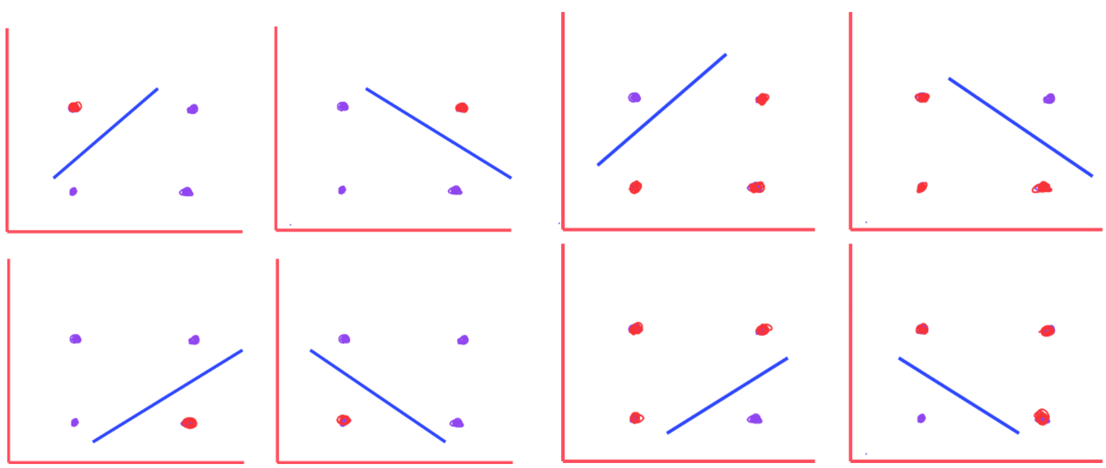
سوال 1)

الف)

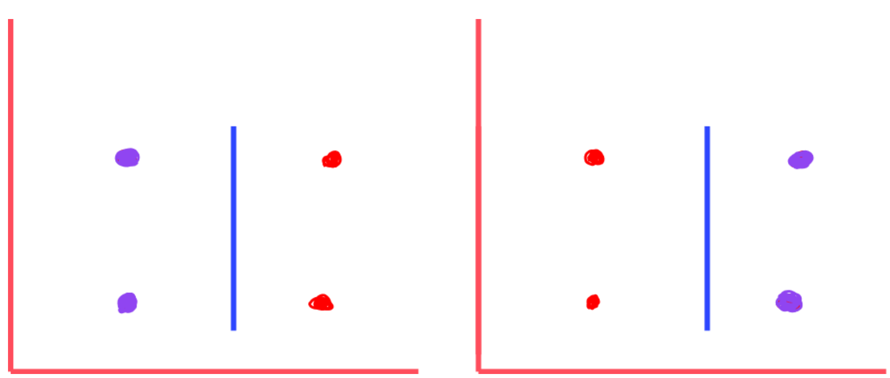
14تابع قابل نمایش هستند.
برای دو ورودی، چهار ترکیب مختلف از ورودی‌ها داریم: (1,1),(1,0),(0,1),(0,0).

رای هر یک از این چهار ترکیب، یک پرسپترون وجود دارد که فقط همان نقطه را 1 و سایر نقاط را 0 برچسب می‌زند. همچنین پرسپترونی وجود دارد که فقط همان نقطه را 0 و سایر نقاط را 0 برچسب می‌زند. این مجموعه شامل 8 حالت می‌شود.

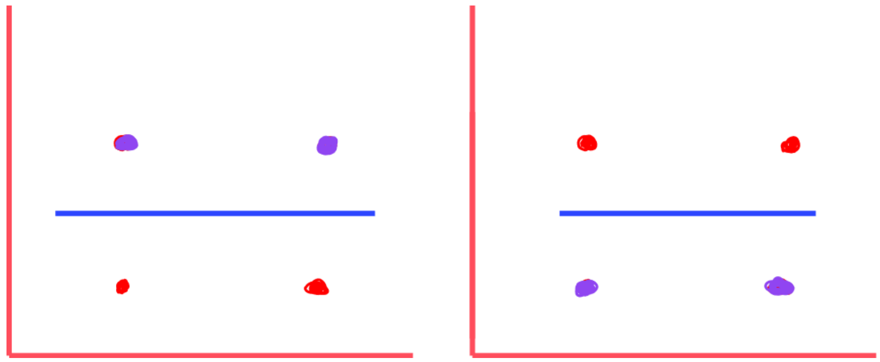
بنفش برچسب 1 و قرمز برچسب 0 می باشد



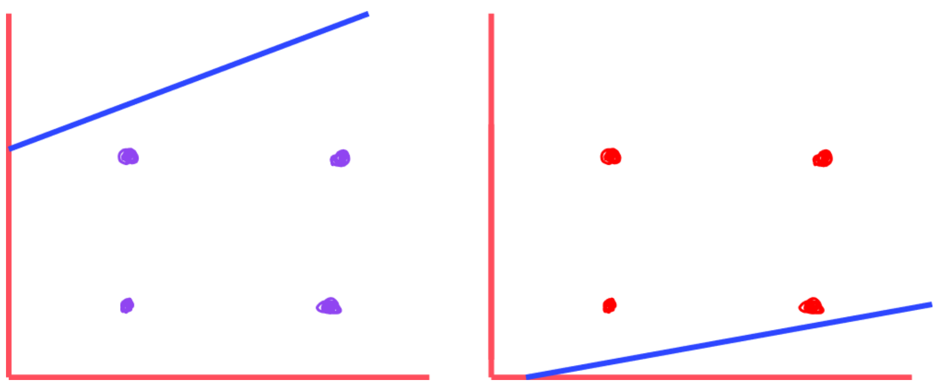
سپس، یک پرسپترون وجود دارد که ورودی دوم را نادیده می‌گیرد و به هر دو نقطه‌ای که ورودی اولشان برابر 1 است، برچسب 1 و به دو نقطه دیگر برچسب 0 می‌دهد، و یک پرسپترون دیگر نیز وجود دارد که این برچسب‌ها را به صورت معکوس اعمال می‌کند. این دو حالت 2 مورد دیگر اضافه می‌کند.



در نهایت، یک جفت پرسپترون وجود دارد که ورودی اول را نادیده می‌گیرد و دقیقاً همان کاری را انجام می‌دهد که در حالت قبلی انجام می‌شد، اما این بار با استفاده از ورودی دوم.



همچنین، موارد «ساده» وجود دارند که تمام نقاط داده برچسب 1 یا تمام نقاط داده برچسب 0 می‌گیرند، که این‌ها نیز به عنوان آستانه‌های خطی قابل نمایش هستند، زیرا تمام نقاط در یک سمت یک سطح خطی قرار می‌گیرند.



تنها توابعی که به صورت خطی جداپذیر نیستند xor و xnor می باشند

ب)

این گزاره **صحیح** است.

ج)

این گزاره **صحیح** است.

د)

این گزاره **غلط** است.

ه)

این گزاره **غلط** است.

و)

مشکل ناپدید شدن گرادیان ،زمانی است که گرادیان‌های تابع خطا نسبت به وزن‌های لایه‌های اولیه بسیار کوچک می‌شوند. این اتفاق به این دلیل رخ می‌دهد که در طی فرآیند پس‌انتشار(Backpropagation) ، گرادیان‌ها به صورت لایه به لایه ضرب می‌شوند، و اگر مشتقات تابع فعال‌سازی کوچک باشند (مانند توابع sigmoid یاtanh)، گرادیان‌ها به صورت نمایی کوچک‌تر می‌شوند

ReLU تابع فعال‌سازی است که وقتی به ورودی‌های مثبت اعمال می‌شود، خروجی خطی مثبت تولید می‌کند. اگر ورودی منفی باشد، تابع مقدار صفر را برمی‌گرداند.
مشتق تابع ReLU به این صورت تعریف می‌شود که برای ورودی‌های بزرگ‌تر از صفر برابر با 1 و برای ورودی‌های منفی برابر با 0 است.
اگر از تابع ReLU به جای تابع سیگموید برای فعال‌سازی در یک شبکه عصبی استفاده شود، مقدار مشتق جزئی تابع خطا دارای مقادیر صفر یا یک خواهد بود که مانع از ناپدید شدن گرادیان می‌شود. بنابراین استفاده از تابع ReLU از ناپدید شدن گرادیان جلوگیری می‌کند.

ز)

این گزاره **غلط** است

ی)

$$\begin{cases} m_0 = 0 \\ \mu = 0 \\ \gamma = 0.05 \\ x_0 = -2.8 \end{cases}$$

$$\frac{dy}{dx} = 1.2x^3 - 0.3x^2 - 4x - 0.8$$

محاسبه g1 از روی گرادیان تابع y:

$$g_1 = 1.2(-2.8)^3 - 0.3(-2.8)^2 - 4(-2.8) - 0.8 = -15.1$$

بروزرسانی momentum با گرادیان بدست امده و momentum قبلی:

$$m_1 = \mu m_0 + (1-\mu)g_1 = 0.7(0) + 0.3(-15.1) = -4 \cdot 53$$

بروزرسانی پارامتر x1:

$$x_1 = x_0 - \gamma m_1 = -2.8 - 0.05(-4 \cdot 53) = -2.5256$$

محاسبه y با x بروزرسانی شده:

$$y_1 = 0.3x_1^4 - 0.1x_1^3 - 2x_1^2 - 0.8x_1$$

$$y_1 = 0.3(-2.5256)^4 - 0.1(-2.5256)^3 - 2(-2.5256)^2 - 0.8(-2.5256) = 5.12$$

موارد فوق را برای iteration 2 نیز تکرار می کنیم

$$g_2 = 1.2(-2.5256)^3 - 0.3(-2.5256)^2 - 4(-2.5256) - 0.8 = -11.9428$$

$$m_2 = \mu m_1 + (1-\mu)g_2 = 0.7(-4.53) + 0.3(-11.9428) = -7.4247$$

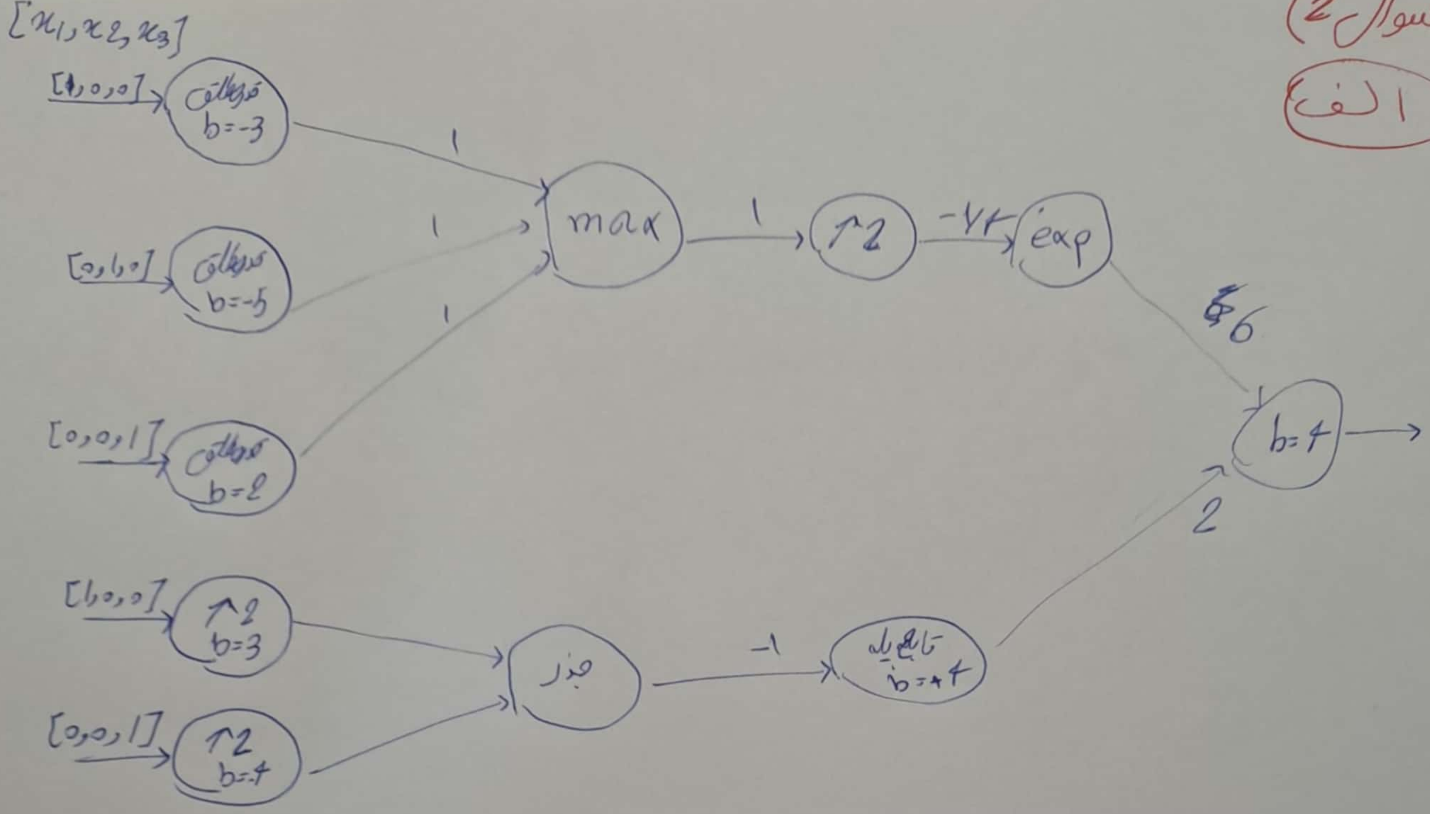
$$x_2 = x_1 - \gamma m_2 = -2.5256 - 0.05(-7.4247) = -2.1544$$

$$y_2 = 0.3x_2^4 - 0.1x_2^3 - 2x_2^2 - 0.8x_2$$

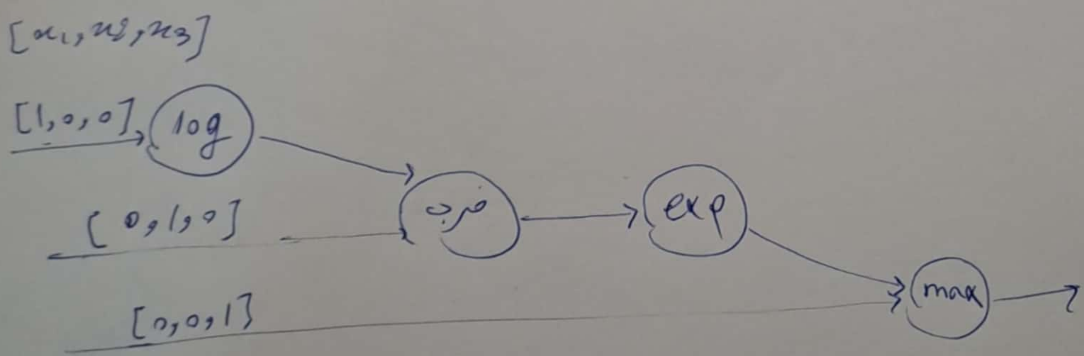
$$y_2 = 0.3(-2.1544)^4 - 0.1(-2.1544)^3 - 2(-2.1544)^2 - 0.8(-2.1544) = 3.44$$

سوال 2

الف


[x1, x2, x3]
[1,0,0]
قدرمطلق b=-3
[0,1,0]
قدرمطلق b=-5
[0,0,1]
قدرمطلق b=2
max
^2
exp
b=4
[1,0,0]
^2 b=3
[0,0,1]
^2 b=4
جذر
تابع پله b=4
1
1
1
1
-1
2
6

ب


[x1, x2, x3]
[1,0,0]
log
[0,1,0]
ضرب
exp
[0,0,1]
max

سوال 3

① $\begin{cases} W_1 : D_{a_1} \times D_x \\ b_1 : D_{a_1} \times 1 \end{cases}$ $\qquad$ $\begin{cases} w_2 : 1 \times D_{a_1} \\ b_2 : 1 \times 1 \end{cases}$

اگه vectorize کنیم w و b تغییری نمی‌کنه فقط X و Y عوض می‌شوند ← $\begin{cases} X : D_x \times D_m \\ Y : 1 \times m \end{cases}$

---

② $$\frac{\partial J}{\partial \hat{y}^{(i)}} = \frac{\partial J}{\partial L^{(i)}} \times \frac{\partial L^{(i)}}{\partial \hat{y}^{(i)}} = -\frac{1}{m}\left(\frac{y^{(i)}}{\hat{y}^{(i)}} - \frac{1-y^{(i)}}{1-\hat{y}^{(i)}}\right) = \boxed{-\frac{1}{m}\left(\frac{y^{(i)} - \hat{y}^{(i)}}{\hat{y}^{(i)}(1-\hat{y}^{(i)})}\right) = \delta_1^{(i)}}$$

$$\frac{\partial J}{\partial \hat{Y}} = -\frac{1}{m}\left(\frac{Y-\hat{Y}}{\hat{Y}(1-\hat{Y})}\right)$$

---

③ $$\frac{\partial \hat{y}^{(i)}}{\partial Z_2} = \sigma(Z_2)\left(1-\sigma(Z_2)\right) = \delta_2^{(i)}$$

---

④ $$\frac{\partial Z_2}{\partial a_1} = W_2 = \delta_3^{(i)}$$

---

⑤ $$\frac{\partial a_1^{\to \text{Relu}}}{\partial Z_1} = \begin{cases} 1 & Z_1 > 0 \\ 0 & Z_1 < 0 \end{cases} = \delta_4^{(i)}$$
$\hookrightarrow I(Z_1 > 0)$

---

⑥ $$\frac{\partial Z_1}{\partial W_1} = (x^{(i)})^T = \delta_5^{(i)}$$

---

⑦ $$\frac{\partial J}{\partial W_1} = \prod_{k=1}^{k=5} \delta_k^{(i)} = -\frac{1}{m}\left(\frac{y^{(i)} - \hat{y}^{(i)}}{\hat{y}^{(i)}(1-\hat{y}^{(i)})}\right) \times \sigma(Z_2)\left(1-\sigma(Z_2)\right) \times W_2 \times I(Z_1 > 0) \times (x^{(i)})^T$$

سوال (4)

الف) $z^{(1)} = w^{(1)}x + b^{(1)} = \begin{bmatrix} a \\ b \end{bmatrix}$

$z^{(2)} = w^{(2)} \times h^{(1)} + b^{(2)}$

$h_1^{(1)} = \sigma(z^{(1)}) = \begin{bmatrix} \frac{1}{1+e^{-a}} \\ \frac{1}{1+e^{-b}} \end{bmatrix} \Rightarrow z^{(2)} = \begin{bmatrix} \frac{a}{1+e^{-a}} + \frac{b}{1+e^{-b}} \\ \frac{-b}{1+e^{-a}} - \frac{a}{1+e^{-b}} \end{bmatrix}$

$z^{(3)} = w^{(3)} \times h^{(2)} + b^{(3)}$

$h^{(2)} = Relu(z^{(2)}) = \begin{bmatrix} \max(0, \frac{a}{1+e^{-a}} + \frac{b}{1+e^{-b}}) \rightarrow h_1^{(2)} \\ \max(0, \frac{-b}{1+e^{-a}} - \frac{a}{1+e^{-b}}) \rightarrow h_2^{(2)} \end{bmatrix} \Rightarrow z^{(3)} = \begin{bmatrix} a h_1^{(2)} - b h_2^{(2)} \\ -a h_1^{(2)} + b h_2^{(2)} \end{bmatrix}$

$z^{(4)} = w^{(4)} h^{(3)} + b^{(4)}$

$h^{(3)} = \text{softmax}(z^{(3)}) = \begin{bmatrix} \frac{e^{z_1^{(3)}}}{e^{z_1^{(3)}} + e^{z_2^{(3)}}} \\ \frac{e^{z_2^{(3)}}}{e^{z_1^{(3)}} + e^{z_2^{(3)}}} \end{bmatrix} \Rightarrow z^{(4)} = \begin{bmatrix} a h_1^{(3)} - b h_2^{(3)} \\ 0.5 a h_1^{(3)} + a h_2^{(3)} \end{bmatrix}$

$h^{(4)} = I(z^{(4)}) = z^{(4)}$

ب) $E = \frac{1}{2}\left[\left((a h_1^{(3)} - b h_2^{(3)}) - 0 \right)^2 + \left((0.5 a h_1^{(3)} + a h_2^{(3)}) - 1 \right)^2 \right]$

2) $\frac{dE}{dh^{(4)}} = h^{(4)} - t \Rightarrow \begin{bmatrix} z_1^{(4)} \\ z_2^{(4)} - 1 \end{bmatrix} = \delta_4$

layer 4, $\frac{dE}{dh^{(4)}} = \frac{dE}{dz^{(4)}} = \delta_4$

$\frac{dE}{dw_4} = \frac{dE}{dz^{(4)}} \times \frac{dz^{(4)}}{dw^{(4)}} = \delta_4 (h^{(3)})^T = \begin{bmatrix} z_1^{(4)} h_1^{(3)} & z_1^{(4)} h_2^{(3)} \\ (z_2^{(4)} - 1) h_1^{(3)} & (z_2^{(4)} - 1) h_2^{(3)} \end{bmatrix}$

$\frac{dE}{db_4} = \delta_4$

layer3 $\delta_3 = w_4^T \times \delta_4$ $\quad \delta_3 = \begin{bmatrix} a.z_1^{(4)} + 0.5a(z_2^{(4)} - 1) \\ -b.z_1^{(4)} + a(z_2^{(4)} - 1) \end{bmatrix}$

$\frac{dE}{dz_i^3} = \frac{dE}{dh_i^{(3)}} \times \frac{dh_i^{(3)}}{dz_i^{(3)}}$ مشتق softmax

$$\delta_3 = \begin{bmatrix} [a(z_1^{(4)}) + 0.5a(z_2^{(4)} - 1)] \times h_1^{(3)} \times (1 - h_1^{(3)}) \\ [-b(z_1^{(4)}) + a(z_2^{(4)} - 1)] \times h_2^{(3)} (1 - h_2^{(3)}) \end{bmatrix}$$

$$\frac{dE}{dw^{(3)}} = \delta_3 \times h^{(2)T} = \begin{bmatrix} \delta_1^{(3)} h_1^{(2)} & \delta_1^{(3)} h_2^{(2)} \\ \delta_2^{(3)} h_1^{(2)} & \delta_2^{3} h_2^{(2)} \end{bmatrix}$$

$$\frac{dE}{db_3} = \delta_3$$

layer2: $\delta_2 = w_3^T \delta_3 = \begin{bmatrix} a\delta_1^{(3)} - a\delta_2^{(3)} \\ -b\delta_1^{(3)} + b\delta_2^{(3)} \end{bmatrix}$

مشتق Relu

$$\delta_2 = \delta_2 \times (I(z_i^{(2)} > 0))$$

$$\frac{dE}{dw_2} = \delta_2 \times h^{(1)T} = \begin{bmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \end{bmatrix} [h_1^{(1)} \; h_2^{(2)}]$$

layer1: $\delta_1 = w_2^T \delta_2 = \begin{bmatrix} a\delta_1^{(2)} - b\delta_2^{(2)} \\ b\delta_1^{(2)} - a\delta_2^{(2)} \end{bmatrix}$

مشتق sigmoid

$$\frac{dE}{dw_1} = \delta_1 \times (\text{input})^T = \begin{bmatrix} \delta_1^{(1)} & 0 \\ \delta_2^{(1)} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} (a\delta_1^{(2)} - b\delta_2^{(2)}) \times \sigma(a)(1 - \sigma(a)) & 0 \\ (b\delta_1^{(2)} - a\delta_2^{(2)}) \times \sigma(b)(1 - \sigma(b)) & 0 \end{bmatrix}$$

$$\frac{dE}{b_1} = \delta_1$$

د) $w_4 = w_4 - \eta \frac{dE}{dw_4}$

$$w_4 = \begin{bmatrix} a & -b \\ a.5a & a \end{bmatrix} - \begin{bmatrix} \eta\left(\frac{dE}{dw_4}\right)_1 & \eta\left(\frac{dE}{dw_4}\right)_2 \\ \eta\left(\frac{dE}{dw_4}\right)_3 & \eta\left(\frac{dE}{dw_4}\right)_4 \end{bmatrix}$$

$b_4 = b_4 - \eta \delta_4$

این مراحل را برای تمامی لایه‌ها تا بدست همه ضرایب و بایاس‌ها انجام می‌دهیم

سوال 5

ربع اول محور مختصات

$$\begin{cases} x_2+x_1\leqslant 5 \longrightarrow -x_2-x_1+5\geqslant 0 \\ -x_2+x_1\leqslant 1 \longrightarrow x_2-x_1+1\geqslant 0 \\ x_2-x_1\leqslant 1 \longrightarrow -x_2+x_1+1\geqslant 0 \\ -x_2-x_1\leqslant -3 \longrightarrow x_2+x_1-3\geqslant 0 \end{cases}$$

ربع دوم محور مختصات

$$\begin{cases} x_2+x_1\leqslant 1 \longrightarrow -x_2-x_1+1\geqslant 0 \\ x_2-x_1\leqslant 5 \longrightarrow -x_2+x_1+5\geqslant 0 \\ -x_2-x_1\leqslant 1 \longrightarrow x_2+x_1+1\geqslant 0 \\ -x_2+x_1\leqslant -3 \longrightarrow x_2-x_1-3\geqslant 0 \end{cases}$$

ربع سوم محور مختصات

$$\begin{cases} x_2+x_1\leqslant -3 \longrightarrow -x_2-x_1-3\geqslant 0 \\ x_2-x_1\leqslant 1 \longrightarrow -x_2+x_1+1\geqslant 0 \\ -x_2+x_1\leqslant 1 \longrightarrow x_2-x_1+1\geqslant 0 \\ -x_2-x_1\leqslant 5 \longrightarrow x_2+x_1+5\geqslant 0 \end{cases}$$

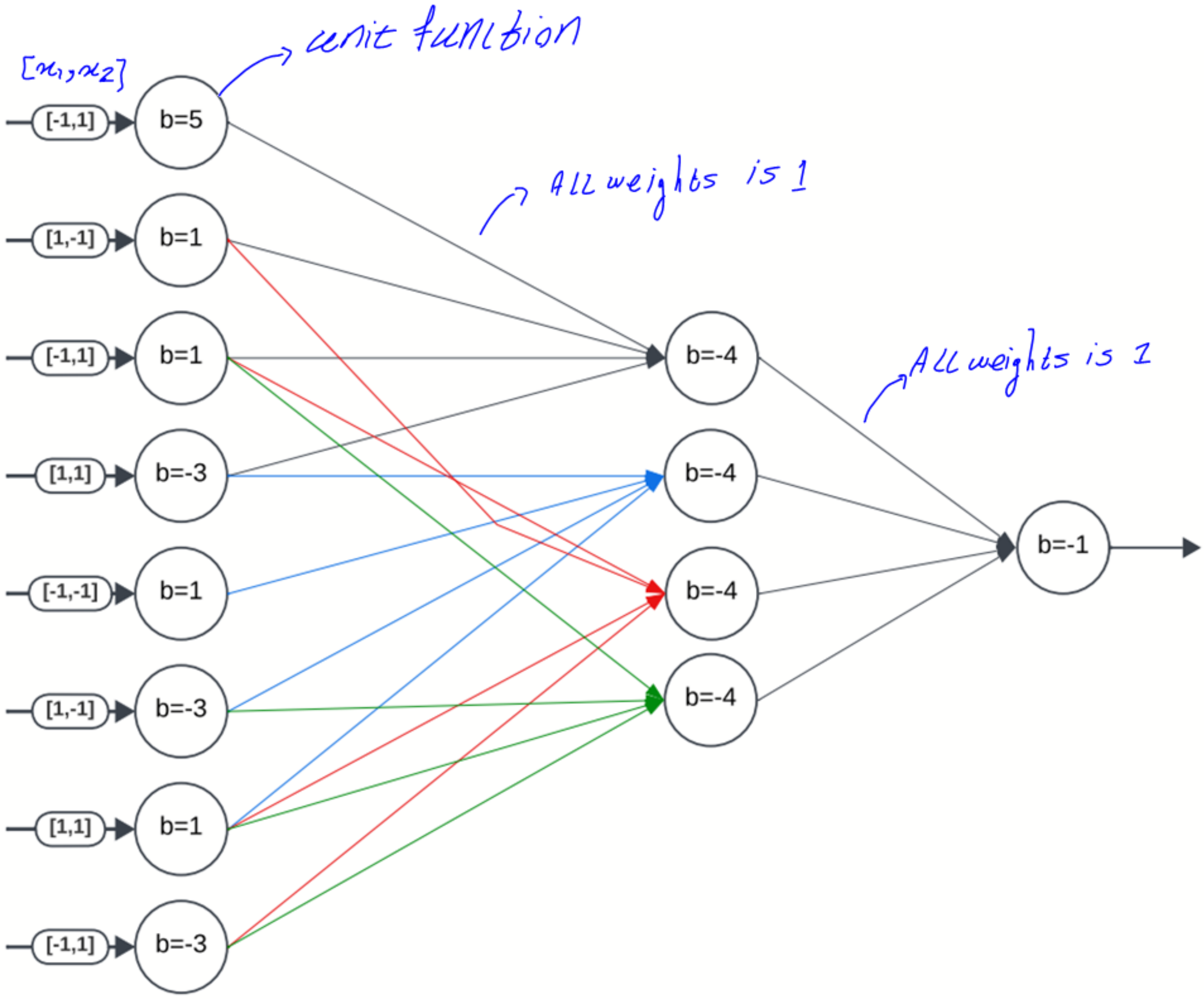
ربع چهارم محور مختصات

$$\begin{cases} x_2-x_1\leqslant -3 \longrightarrow -x_2+x_1-3\geqslant 0 \\ x_2+x_1\leqslant 1 \longrightarrow -x_2-x_1+1\geqslant 0 \\ -x_2-x_1\leqslant 1 \longrightarrow x_2+x_1+1\geqslant 0 \\ -x_2+x_1\leqslant 5 \longrightarrow x_2-x_1+5\geqslant 0 \end{cases}$$

با گرفتن اشتراک این معادلات مرزی به معادلات زیر میرسیم

$$\begin{cases} -x_2-x_1+5\geqslant 0 \\ -x_2+x_1+1\geqslant 0 \\ x_2-x_1+1\geqslant 0 \\ x_2+x_1-3\geqslant 0 \\ -x_2-x_1+1\geqslant 0 \\ -x_2+x_1-3\geqslant 0 \\ x_2+x_1+1\geqslant 0 \\ -x_2-x_1-3\geqslant 0 \end{cases} \Longrightarrow$$

این نواحی مرزی را می‌توان با شبکه عصبی به همراه تابع فعالسازی پله نمایش داد.



ب) برای این قسمت نیز مشابه قسمت قبل نواحی را مشخص و معادلات مرزی را مشخص می‌کنیم

$-x_2 - \frac{1}{3} + 2 \geq 0$

$x_2 + 3x + 6 \geq 0$

$-x_2 + 3x_1 + 6 \geq 0$

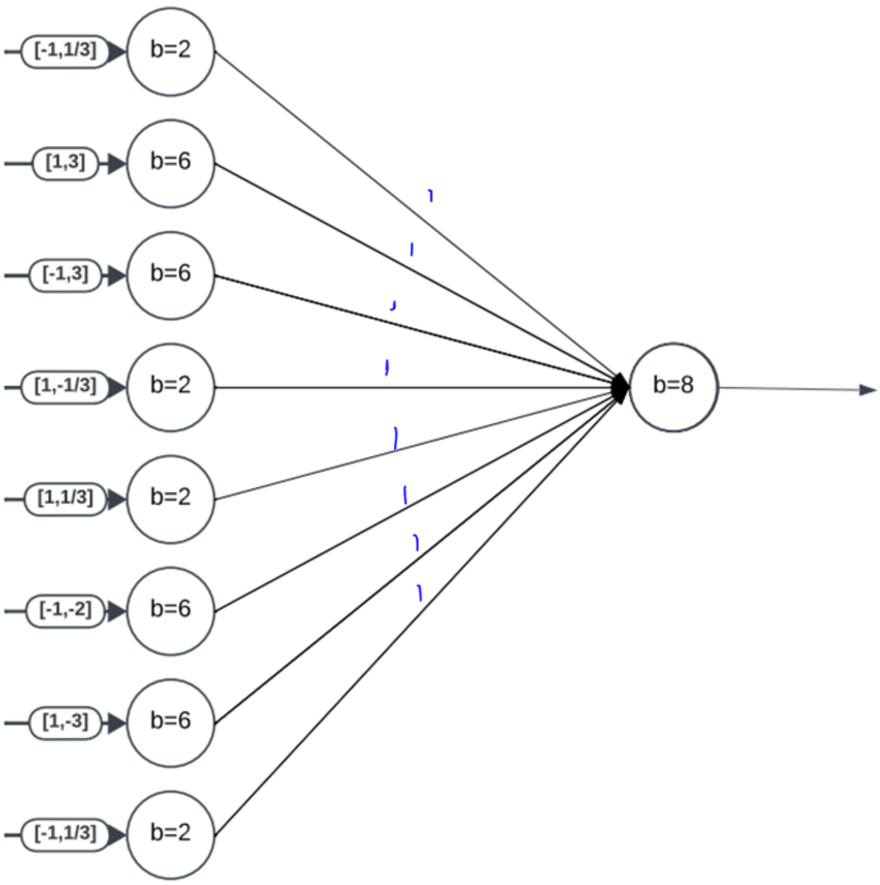
$x_2 - \frac{1}{3}x_1 + 2 \geq 0$

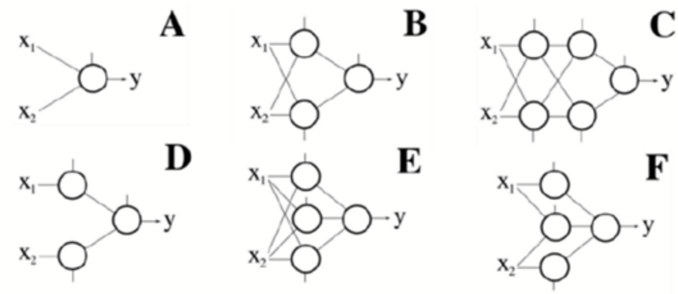
$x_2 + \frac{1}{3}x_1 + 2 \geq 0$

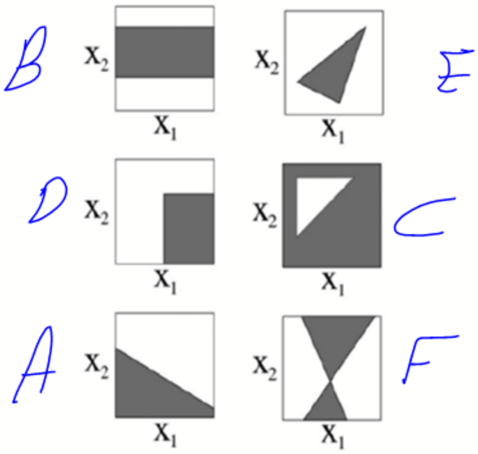
$-x_2 - 3x_1 + 6 \geq 0$

$x_2 - 3x_1 + 6 \geq 0$

$-x_2 + \frac{1}{3}x_1 + 2 \geq 0$







A: ما به کمک یک نورون می توانیم نواحی مرزی خطی داشته باشیم

B: به کمک این شبکه می توانیم به طور مثال ناحیه اشتراکی دو ناحیه را بگیریم

C: چون لایه هاش زیاده توانایی اشتراک گذاری بین مرزها رو دارد

D: به کمک این شبکه می توانیم ناحیه اشتراکی 2 خط $x_1=a$ و $x_2=b$ را مشخص کنیم که در شکل مشخص است

E: به کمک 3 نورون لایه مخفی اول می توانیم 3 خط مرزی مشخص کنیم و در نهایت اشتراکی نواحی مورد نظر رو بگیریم

سوال ۶) الف)

$P = \{x \in R^n \mid Ax \leq b\} \longrightarrow$ Polytope

و دلیل اینکه ثابت کنیم تابع convex است:

$$f(\lambda z_1 + (1-\lambda) z_2) \leq \lambda f(z_1) + (1-\lambda) f(z_2)$$

حال می‌خواهیم ثابت کنیم Relu convex است

$$Relu(\lambda z_1 + (1-\lambda) z_2) = \begin{cases} \lambda z_1 + (1-\lambda) z_1 & z_1 \geq 0 , z_2 \geq 0 \\ 0 & z_1 \leq 0 , z_2 \leq 0 \\ \lambda z_1 & z_1 \geq 0 \quad z_2 \leq 0 \\ (1-\lambda) z_2 & z_1 \leq 0 \quad z_2 \geq 0 \end{cases} = \lambda Relu(z_1) + (1-\lambda) Relu(z_2)$$

پس Relu یک تابع convex است.

$w^T x + b \leftarrow$ خودش که خطی است Relu همانطور که قبلاً گفتیم است چون خطی پس $Relu(w^T x + b)$ ، convex است

$$P = \{x \in R^m \mid w^T x + b \geq 0\} \Rightarrow P = \{x \in R^m \mid -w^T x \leq -b\}$$

این حاصل polytope است

اگر بخواهیم ساختار NN را بصورت یک تعداد polytope است. پس convex است