HW05_ Artificial Intelijence

Saeed Razavi

Student ID: 98106542

1

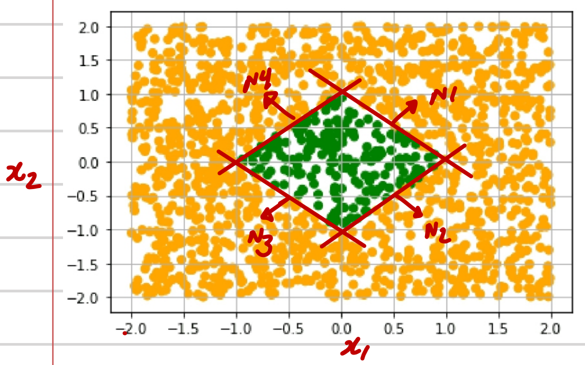
۱. (۱۶ نمره) چرا در شبکه‌های عصبی از تابع‌های اکتیویشن غیرخطی استفاده می‌کنیم؟ (به طور خاص اهمیت این موضوع برای شبکه‌های عصبی چند لایه چیست؟)

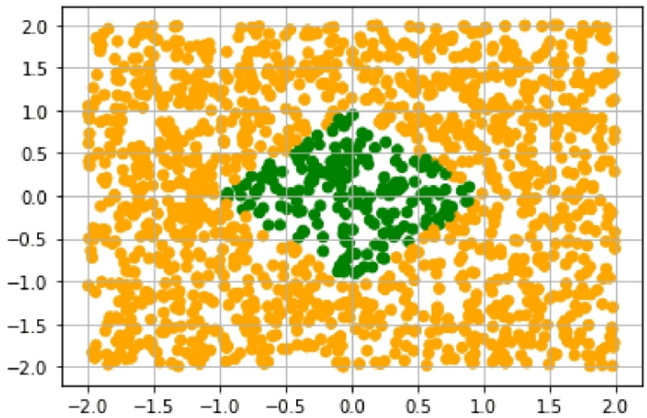
← اگر از تابع activation خطی استفاده کنیم، کل شبکه عصبی را می‌توان به صورت یک $\hat{y} = W_{total}^T x$ مدل کرد.

← شبکه‌ی عصبی دیگر قادر نخواهد بود داده‌های غیرخطی که Linearly seperatable نیستند را جداسازی کند.

← در حقیقت تابع‌های activation غیرخطی نیاز است تا بتوان داده‌ها را به یک فضای جدید برد تا بتوان در آن فضای جدید، آنها را به صورت خطی جدا کرد.

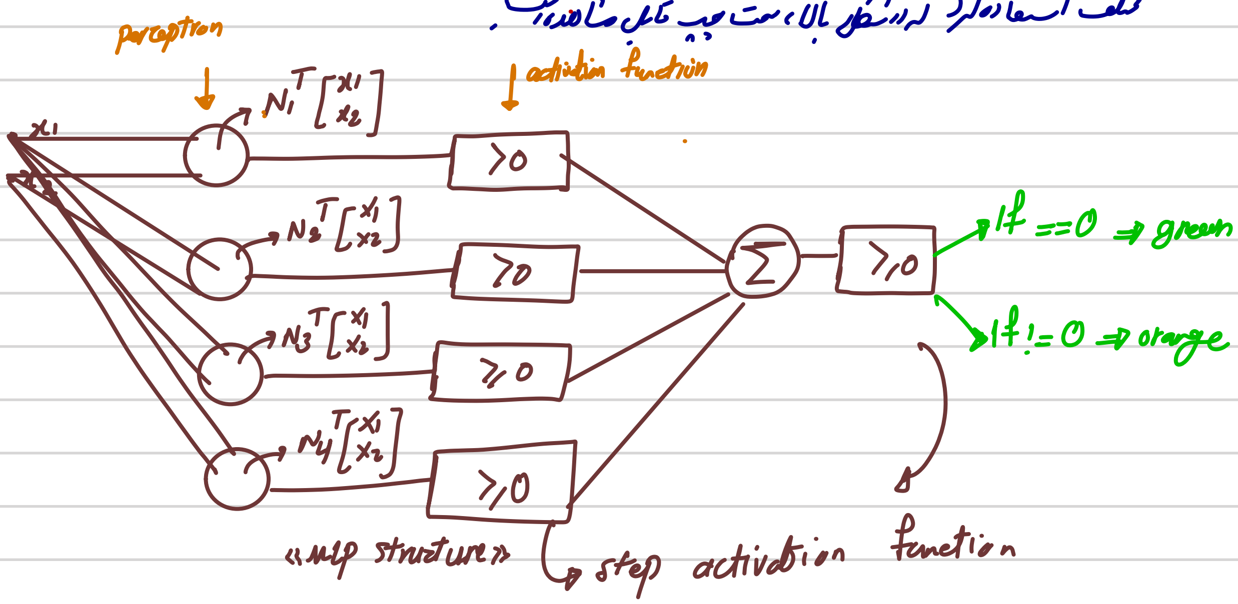
۲. (۲۱ نمره) برای تشخیص داده‌های سبز از نارنجی در شکل زیر یک شبکه‌ی عصبی طراحی کنید.



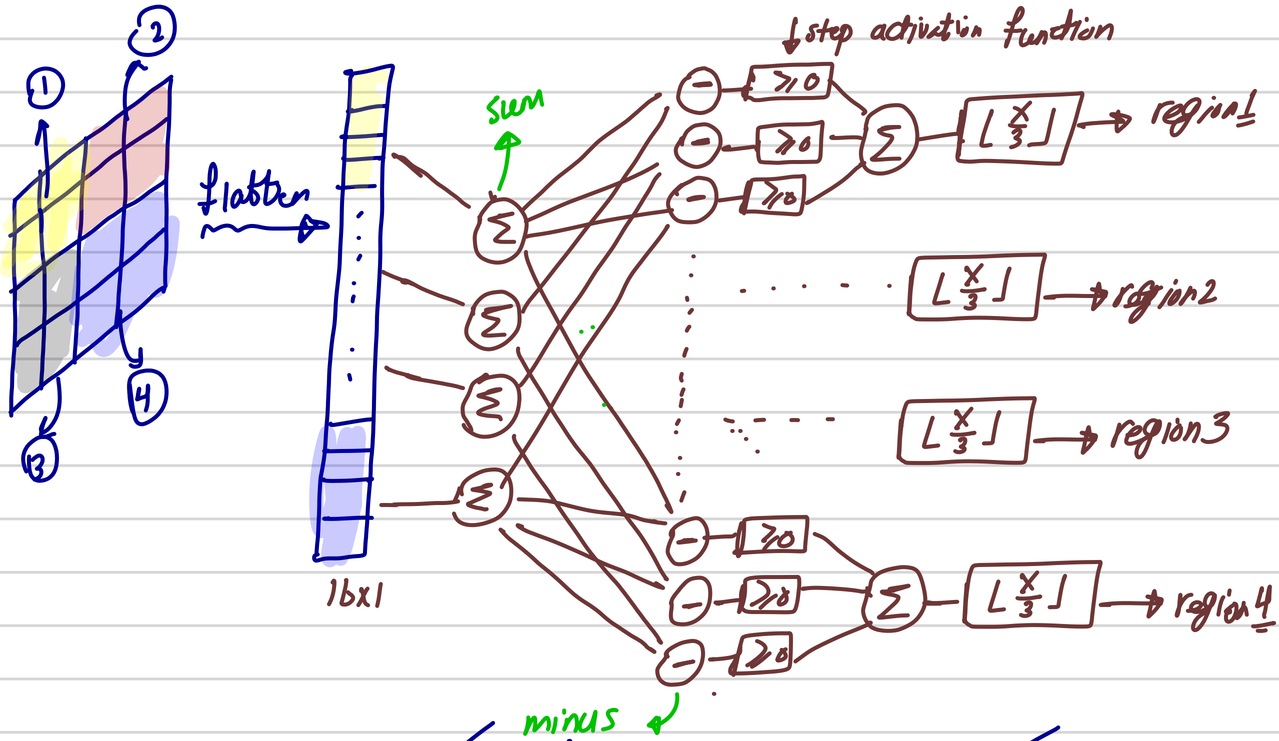


$N_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, N_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, N_3 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, N_4 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

برای جداسازی داده‌ها می‌توان از ۴، Hyperplane مختلف استفاده کرد که در شکل بالا، سمت چپ قابل مشاهده است.



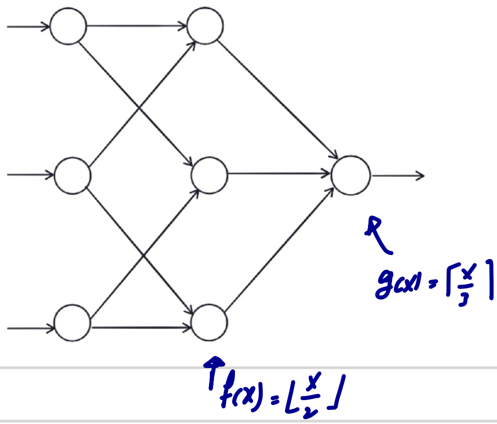
۳. (۲۱ نمره) یک نمایشگر با ابعاد ۴ پیکسل در ۴ پیکسل را در نظر بگیرید. فرض کنید مبدا مختصات را در وسط این نمایشگر قرار دهیم و آن را به چهار محدوده‌ی چهار پیکسلی تقسیم کنیم. مقدار روشنایی هر پیکسل می‌تواند بین ۰ و ۱ داشته باشد. یک شبکه‌ی عصبی طراحی کنید که محدوده‌ی با بیش‌ترین میانگین روشنایی پیکسل‌ها را مشخص کند (نیازی به در نظر گرفتن حالتی که تساوی پیش بیاید نیست).



ابتدا تصویر ۴x۴ را flat می‌کنیم. سپس pixel های هر region را به یک جمع‌کننده می‌دهیم. حال برای در region دلخواه، حاصل جمع روشنایی آن را از بقیه‌ی region ها، یکی یکی کم می‌کنیم. در نهایت با استفاده از فیلتر $\lfloor \frac{x}{3} \rfloor$ روشن‌ترین ناحیه پیدا می‌شود.

۴. (۲۱ نمره) شبکه‌ی زیر یک تابع بولین را (که هر ورودی آن می‌تواند مقدار ۰ یا ۱ را داشته باشد) شبیه‌سازی کرده است. تمامی وزن‌ها برابر ۱ هستند. خروجی نورون‌های لایه‌ی میانی و پایانی به ترتیب از اعمال تابع‌های $f(x) = \lfloor \frac{x}{2} \rfloor$ و $g(x) = \lceil \frac{x}{3} \rceil$ روی مجموع ورودی‌ها به دست می‌آید. تابع بولین پیاده‌سازی شده را به دست بیاورید و در صورت امکان شبکه‌ی ساده‌تری برای آن ارائه کنید.

Write truth table

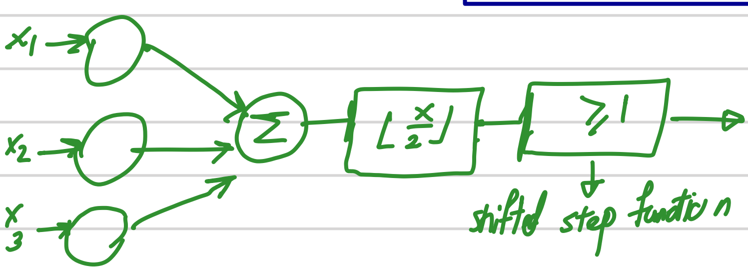


y	x_1	x_2	x_3
1	1	1	0
1	0	1	1
1	1	0	1
0	0	0	1
0	1	0	0
0	0	1	0
0	0	0	0
1	1	1	1

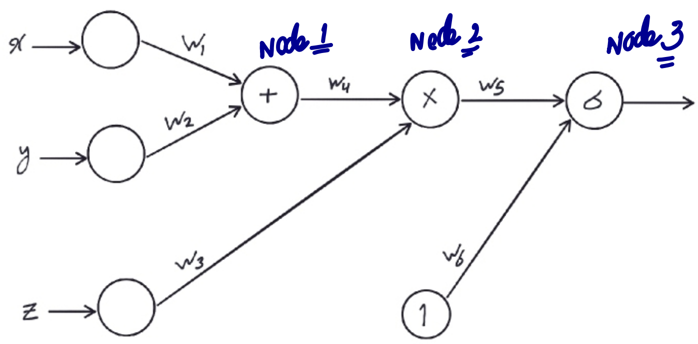
$\boxed{\text{boolean function}: (x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_2 \wedge x_3)}$

تابع boolean بدست آمده!

شبکه‌ی ساده‌تر



۵. (۲۱ نمره) یک دور از محاسبات رو به جلو (feed forward) و رو به عقب (backpropagation) را برای شبکه‌ی زیر انجام دهید. در ابتدا همه‌ی وزن‌ها برابر $\frac{1}{2}$ هستند و داریم $x = 1, y = 3, z = 7$. تابع موجود در نورون نهایی تابع سیگموید است. (پیشنهاد می‌شود درباره‌ی مشتق تابع سیگموید جست‌وجو کنید.)



Proof:

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

$$\Rightarrow \frac{\partial}{\partial x}\sigma(x) = -1(1+e^{-x})^{-2}(-e^{-x})$$

$$\Rightarrow \frac{\partial}{\partial x}\sigma(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\Rightarrow \frac{\partial}{\partial x}\sigma(x) = \frac{1}{1+e^{-x}} \times \frac{e^{-x}}{1+e^{-x}} = \boxed{\sigma(x)(1-\sigma(x))}$$

$$\sigma'(x) = \sigma(x)(1-\sigma(x)) \quad \text{(I)}$$

feed forward part

→ feed-forward ⇒ $out(Node\,1) = \frac{1}{2}[1+3] = 2$, $out(Node\,2) = \frac{7}{2}$

$out(Node\,3) = \sigma\left(\frac{1}{2} + \frac{7}{4}\right) = \sigma\left(\frac{9}{4}\right) = 0.904$

→ before go to backpropagation, It's better to have proper formulation

for each Node ⇒
$$\begin{cases} out(Node\,1) = W_1 x + W_2 y \\ out(Node\,2) = W_3(z) \times out(Node\,1) W_4 = W_3 z + W_4[W_1 x + W_2 y] \\ out(Node\,3) = \sigma(W_6 + W_5(out(Node\,2))) = \sigma(\underbrace{W_6 + W_5[W_3 z + W_4(W_1 x + W_2 y)]}_{u}) \end{cases}$$

→ $\boxed{\text{suppose cost function is } C = (y - out(Node\,3))^2}$

$$\Rightarrow \frac{\partial C}{\partial W_6} = \frac{\partial C}{\partial out(Node\,3)} \times \underbrace{\frac{\partial out(Node\,3)}{\partial W_6}} = -2[out(Node\,3)][y - out(Node\,3)]$$

$$\frac{\partial \sigma}{\partial u} \times \frac{\partial u}{\partial W_6} \rightarrow \sigma(u)(1-\sigma(u))(1) \Rightarrow \frac{\partial C}{\partial W_6} = (\quad) \times \sigma(u)(1-\sigma(u))$$

$$u = W_6 + W_5[W_3 z + W_4(W_1 x + W_2 y)]$$

$$\boxed{\Rightarrow W_6^{new} = W_6 - \gamma \frac{\partial C}{\partial W_6}}$$

$$\frac{\partial c}{\partial w_5} = \frac{\partial c}{\partial(\text{out}(\text{Node }3))} \times \frac{\partial(\text{out}(\text{Node }3))}{\partial w_5}$$

نحوه‌ی محاسبه‌ی آن تقریباً شبیه به W6 می‌باشد.

$$-2[\text{out}(\text{Node }3)][y-\text{out}(\text{Node }3)]$$

$$[w_3 z + w_4(w_1 x + w_2 y)] \times \sigma(u)(1-\sigma(u))\Big|_{u = w_6 + w_5[w_3 z + w_4(w_1 x + w_2 y)]}$$

$$\Rightarrow \boxed{w_5^{new} = w_5 - \gamma \frac{\partial c}{\partial w_5}}$$

$$\frac{\partial c}{\partial w_4} = \frac{\partial c}{\partial(\text{out}(\text{Node }3))} \times \frac{\partial(\text{out}(\text{Node }3))}{\partial(\text{out}(\text{Node }2))} \times \frac{\partial\, \text{out}(\text{Node }2)}{\partial w_4}$$

$$\text{III} \rightarrow \boxed{w_1 x + w_2 y}$$

$$\text{I} \rightarrow \boxed{-2[\text{out}(\text{Node }3)][y-\text{out}(\text{Node }3)]}$$

$$\text{II} \rightarrow \frac{\partial \sigma(w_6 + w_5(\text{out}(\text{Node }2)))}{\partial(\text{out}(\text{Node }2))} = \boxed{w_5\, \sigma(u)(1-\sigma(u))\Big|_{u = w_6 + w_5[\text{out}(\text{Node }2)]}}$$

$$\Rightarrow \frac{\partial c}{\partial w_4} = \text{I} \times \text{II} \times \text{III}$$

$$\boxed{w_4^{new} = w_4 - \gamma \frac{\partial c_4}{\partial w_4}}$$

$$\frac{\partial c}{\partial w_3} = \frac{\partial c}{\partial(\text{out}(\text{Node }3))} \times \frac{\partial(\text{out}(\text{Node }3))}{\partial(\text{out}(\text{Node }2))} \times \frac{\partial(\text{out}\ \text{Node }2)}{\partial w_3}$$

$$\text{III} \rightarrow z$$

$$\text{I} \rightarrow \boxed{-2[\text{out}(\text{Node }3)][y-\text{out}(\text{Node }3)]}$$

$$\text{II} \rightarrow \boxed{w_5\, \sigma(u)(1-\sigma(u))\Big|_{u = w_6 + w_5[\text{out}(\text{Node }2)]}}$$

$$\frac{\partial c}{\partial w_3} = \text{I} \times \text{II} \times \text{III}$$

$$\boxed{w_3^{new} = w_3 - \gamma \frac{\partial c_3}{\partial w_3}}$$

$$\frac{\partial z}{\partial w_2} = \frac{\partial z}{\partial (out(Node\,3))} \times \frac{\partial (out(Node\,3))}{\partial (out(Node\,2))} \times \frac{\partial\, out(Node\,2)}{\partial\, out(Node\,1)} \times \frac{\partial\, out(Node\,1)}{\partial w_2}$$

I: $\boxed{-2[out(Node\,3)][y - out(Node\,3)]}$

II: $\boxed{w_5\, \sigma(u)(1-\sigma(u)) \quad u = w_6 + w_5[out(Node\,2)]}$

III: $\boxed{w_4}$

IV: $\boxed{y}$

$$\rightarrow \frac{\partial z}{\partial w_2} = \text{I} \times \text{II} \times \text{III} \times \text{IV} \rightarrow w_2^{new} = w_2 - \gamma \frac{\partial z}{\partial w_2}$$

$$\frac{\partial z}{\partial w_1} = \frac{\partial z}{\partial (out(Node\,3))} \times \frac{\partial (out(Node\,3))}{\partial (out(Node\,2))} \times \frac{\partial\, out(Node\,2)}{\partial\, out(Node\,1)} \times \frac{\partial\, out(Node\,1)}{\partial w_1}$$

I: $\boxed{-2[out(Node\,3)][y - out(Node\,3)]}$

II: $\boxed{w_5\, \sigma(u)(1-\sigma(u)) \quad u = w_6 + w_5[out(Node\,2)]}$

III: $\boxed{w_4}$

IV: $\boxed{x}$

$$\rightarrow \frac{\partial z}{\partial w_1} = \text{I} \times \text{II} \times \text{III} \times \text{IV} \rightarrow w_1^{new} = w_1 - \gamma \frac{\partial z}{\partial w_1}$$