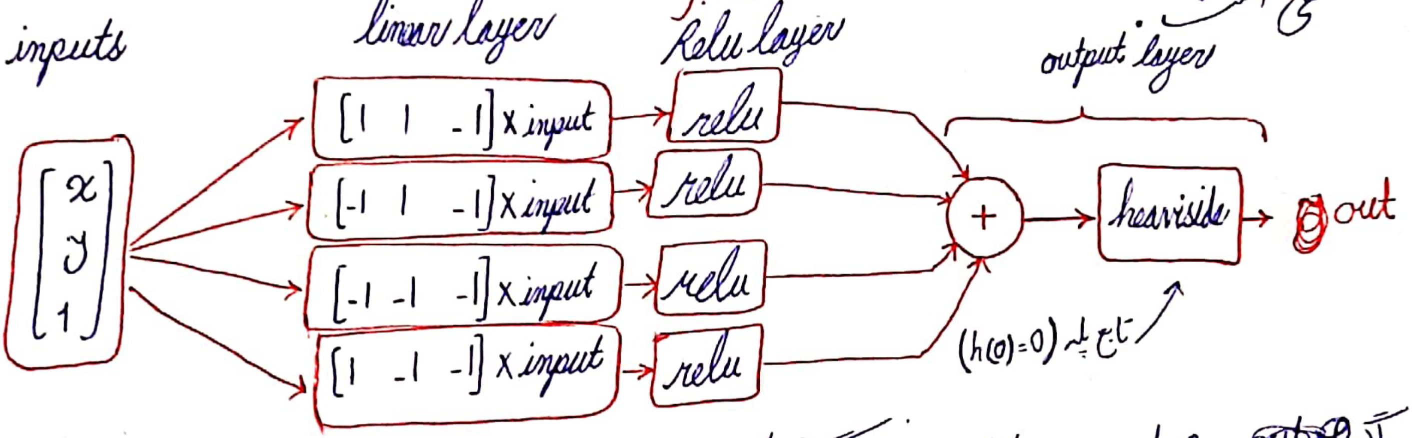
هوش مصنوعی - تمرین تئوری سری 5

مهالنسا مظاهری - 98102346

1- علت استفاده از activation function :

زیرا در صورت عدم استفاده تابع اکتیویشن غیرخطی، استفاده از شبکه عصبی بدون توجیه خواهد بود چرا که شبکه متشکل از چندین لایه خطی خواهد بود که قابل پیاده سازی توسط یک لایه خطی هم خواهد بود (ترکیب چندین لایه خطی، لایه خطی خواهد بود) و قدرت (robustness) شبکه عصبی از دست خواهد رفت؛
دلیل دیگر استفاده از تابع اکتیویشن غیرخطی این است که مسائلی که با این روش حل می‌شوند، جداپذیر خطی نیستند

2- هدف تشکیل مسئله طبقه بندی $|x|+|y|-1 \overset{>}{\underset{<}{}} 0$ (نارنجی >0، سبز <0) با استفاده از شبکه عصبی متشکل از لایه‌های خطی است



اگر ~~out>0~~ out=0 بود، لیبل نارنجی و اگر out=0 بود لیبل سبز است

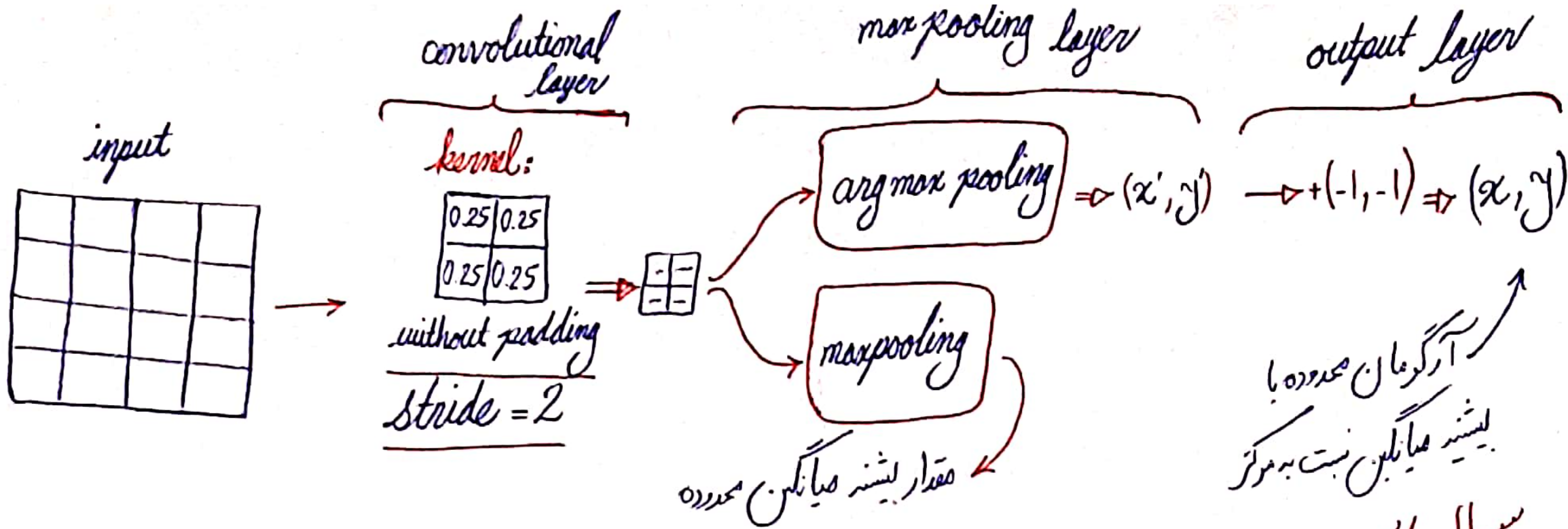
3 یک لایه برای میانگین گیری لازم خواهیم داشت که میانگین این 4 محدوده را محاسبه کند در ماتریسی قرار دهد؛ این لایه کانولوشنی خواهد بود، پس با max pooling مقدار بیشینه میانگین و با argmax pooling، آن محدوده‌ای که بیشترین مقدار را دارد را به خروجی می‌دهیم:

در لایه اول که یک لایه کانولوشنی است، کرنل 2×2 $\begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$ را بدون padding و با stride=2 با تصویر کانوالو می‌کنیم، پس خروجی این لایه یک ماتریس 2×2 است، که میانگین هر چهارتایی مد نظر در آن خانه‌ها قرار دارد.

در لایه دوم با استفاده از max pooling, argmax pooling، اندیس‌های درایه بیشینه ماتریس خروجی لایه قبل را پیدا می‌کنیم

در لایه سوم نیز با توجه به اینکه مرکز مختصات را مرکز تصویر در نظر گرفته‌ایم خروجی‌های لایه قبل را با بایاس (-1) جمع می‌کنیم

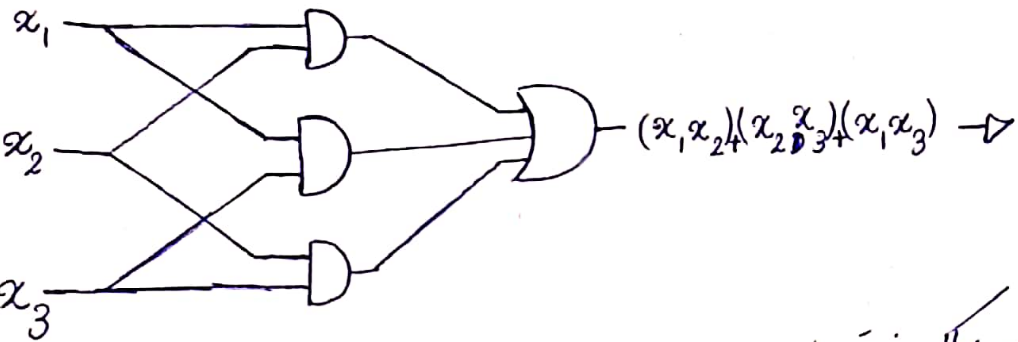
لایه سوم همان لایه خروجی است.



سوال 4

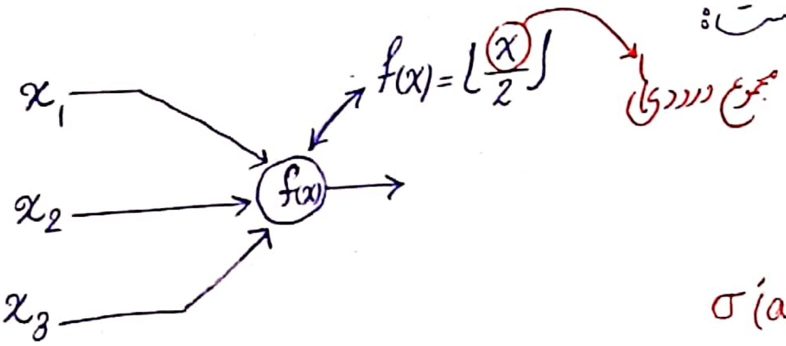
لایه‌ی میانی دو ورودی دارند خروجی آن‌ها برابر $f(x)=\lfloor \frac{x}{2} \rfloor$ است پس معادل AND است.

لایه پایانی سه ورودی دارند و خروجی آن برابر $g(x)=\lceil \frac{x}{3} \rceil$ است که معادل گیت or است.



رفتار این تابع به این شکل است که اگر بیشتر از یکی از ورودی‌ها 1 بود، خروجی 1 شود

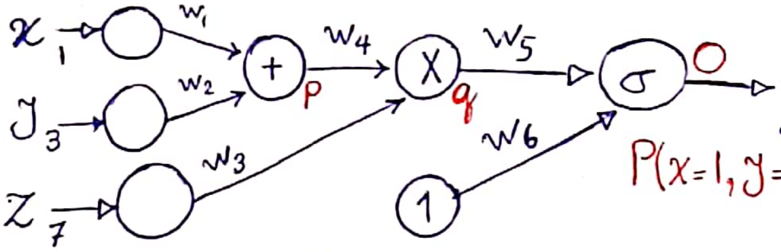
شبکه روبرو، شبکه ساده شده با همان عملکرد فوق است:



سوال 5 مشتق تابع

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

محاسبات رو به جلو



$$J_p\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} w_4 \to \begin{bmatrix} 0.25 & 0.25 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

$$P(x=1, y=3) = 2, q = .5$$

$$\frac{\partial P}{\partial x} = w_1 \qquad \frac{\partial q}{\partial P} = z w_3 w_4 \qquad \frac{\partial O}{\partial q} = w_5 \sigma(w_5 q + w_6).(1-\sigma(w_5 q + w_6))$$

$$\frac{\partial P}{\partial y} = w_2$$

$$\frac{\partial q}{\partial z} = P w_3 w_4$$

$$\nabla O = \begin{bmatrix} \frac{0.5 \times 7}{4} & \frac{0.5 \times 7}{4} & 0.5 \end{bmatrix} \times \frac{1}{2} \times \sigma(2.5)(1-\sigma(2.5))$$

محاسبات رو به عقب (back propagation)

$$f_1 : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \to \begin{bmatrix} \frac{x}{2} + \frac{y}{2} \\ z \end{bmatrix}, \quad f_2 : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to \left[\underset{w_4}{\frac{1}{2}} x_1 \times \underset{w_3}{\frac{1}{2}} x_2 \right], \quad f_3 : x_1 \to \sigma\left(\frac{x_1 + 1}{2}\right)$$

$$f(x) = f_3 \circ f_2 \circ f_1$$

$$J_f(x_q) \text{ for } x_q = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}$$

$$J_{f_1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad J_{f_2}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \left[\frac{x_2}{4} \quad \frac{x_1}{4} \right], \quad J_{f_3} = \frac{1}{2}\sigma\left(\frac{x_1+1}{2}\right)\left(1 - \sigma\left(\frac{x_1+1}{2}\right)\right)$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = f_1\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 7 \end{bmatrix} \qquad x_1 = f_2\left(\begin{bmatrix} 2 \\ 7 \end{bmatrix}\right) = 3.5$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$J_{f_2} = \left[\frac{7}{4} \quad \frac{1}{2} \right] \qquad J_{f_3} = \frac{1}{2}\sigma(2.25)\left(1 - \sigma(2.25)\right)$$

$$J_f\left(\begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}\right) = \cancel{\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}} \; J_{f_3} \times J_{f_2} \times J_{f_1} = \frac{1}{2}\sigma(2.25)\left(1 - \sigma(2.25)\right) \times \left[\frac{7}{4} \quad \frac{1}{2} \right] \times \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\sigma(2.25)\left(1 - \sigma(2.25)\right) \times \left[\frac{7}{8} \quad \frac{7}{8} \quad \frac{1}{2} \right]$$