به‌نام خدا

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

هوش مصنوعی

تمرین ۵: Neural Networks

مهراد میلانلو

۹۹۱۰۵۷۷۵

سوال ۱

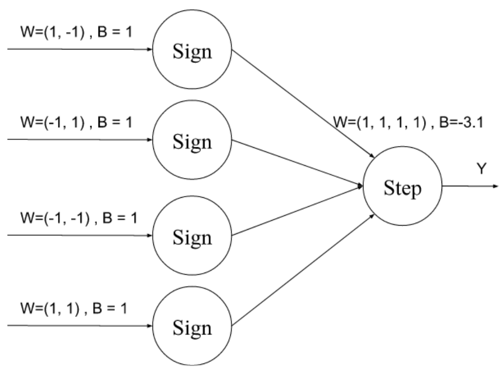
فرض کنید که هیچ تابع Activation غیرخطی نداشته باشیم. و همگی خطی باشند. آن‌گاه نشان می‌دهیم خروجی شبکه‌ی عصبی چندلایه، صرفا یک تابع خطی از ورودی‌ها خواهد بود.

$$f_i(X) = W_i^T X + B_i$$
$$\Rightarrow y = f_n(f_{n-1}(f_{n-2}(...) + B_{n-2}) + B_{n-1}) = \left(\prod_{i=1}^{n} W_i\right) X + \sum_{i=1}^{n} \left(\prod_{j=i+1}^{n} W_j\right) B_i$$
$$\Rightarrow y = WX + B$$

بنابراین خروجی این شبکه‌ی عصبی چندلایه یک تبدیل خطی‌است و تنها می‌تواند مدل‌هایی که رفتار خطی دارند را یاد بگیرد. در صورتی که هدف استفاده از شبکه‌ی عصبی شامل لایه‌ی های مخفی آن است که بتوانیم داده‌ها با اشکال مختلف را یاد بگیریم و عملا با این کار استفاده از شبکه‌ی عصبی چند لایه بی تاثیر می‌شود.

سوال ۲

برای تشخیص نقاط سبز از نقاط نارنجی کافی‌ست نامعادله‌ی $|x_1| + |x_2| \leqslant 1$ را حل کنیم. برای حل این نامعادله نیز با استفاده از شبکه‌ی عصبی، آن را ابتدا به چهار بخش تقسیم می‌کنیم.



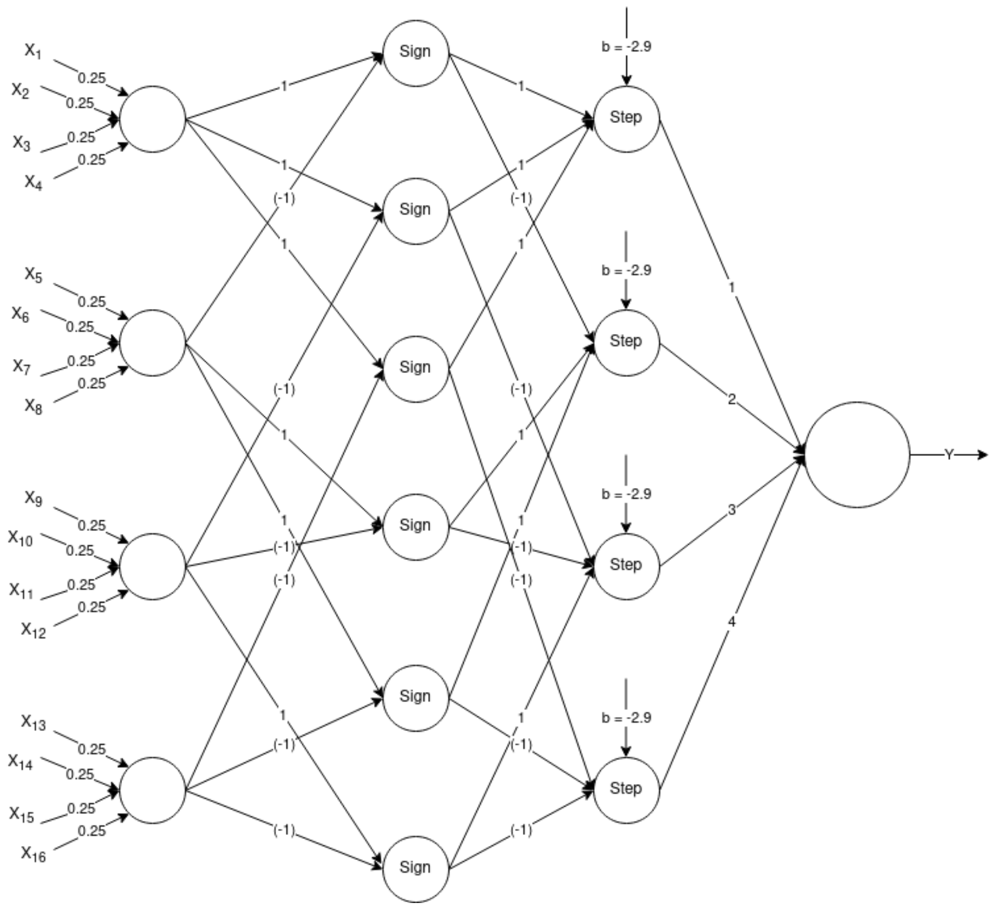
در واقع نامعادله را به چهار قسمت زیر تقسیم می‌کنیم:

$$\begin{cases} x_1 - x_2 \leqslant 1 \\ -x_1 + x_2 \leqslant 1 \\ -x_1 - x_2 \leqslant 1 \\ x_1 + x_2 \leqslant 1 \end{cases}$$

و در نهایت اگر ورودی در تمام این نامعادلات صدق کند، خروجی شبکه ۱ و در غیر این‌صورت ۰ خواهد بود. ورودی‌های شبکه به‌ترتیب x_1, x_2 هستند. همین‌طور در لایه‌ی اول تابع فعال‌سازی تابع Sign است. $Sign(x) = \begin{cases} 1 & if\ x \geqslant 0 \\ -1 & if\ x < 0 \end{cases}$. در لایه‌ی آخر نیز تابع فعال‌سازی تابع Step است. $Step(x) = \begin{cases} 1 & if\ x \geqslant 0 \\ 0 & if\ x < 0 \end{cases}$. بایاس لایه‌ی آخر نیز طوری قرار دادیم که اگر هر ۴ خروجی لایه‌ی قبلی ۱ بود، خروجی ۱ بدهد.

سوال ۳

شبکه‌ی عصبی را به‌گونه‌ی زیر طراحی می‌کنیم:



مقدار روشنایی هر پیکسل را به‌عنوان ورودی‌های شبکه‌ی عصبی از $x_۱$ تا $x_۶$ در نظر می‌گیریم. سپس با استفاده از ضریب یکسان ۰/۲۵، میانگین روشنایی پیکسل‌های هر ناحیه را محاسبه کرده و به لایه‌ی بعدی شبکه می‌رویم. اکنون می‌خواهیم بین میانگین‌های محاسبه شده که هرکدام مربوط به یک ناحیه هستند، مقدار ماکسیمم را مشخص کنیم تا خروجی شبکه باشد. در این لایه، تابع فعال‌سازی را تابع Sign در نظر می‌گیریم: $Sign(x) = \begin{cases} ۱ & if\ x \geqslant ۰ \\ -۱ & if\ x < ۰ \end{cases}$. ضریب‌ها را طوری تعیین می‌کنیم که به‌ترتیب از بالا به پایین، مقدار ورودی به تابع فعالساز، $z_۱ - z_۲, z_۱ - z_۳, z_۱ - z_۴, z_۲ - z_۳, z_۲ - z_۴, z_۳ - z_۴$ باشند. که z_i مقدار میانگین روشنایی ناحیه‌ی iام باشد. در لایه‌ی بعدی، باید مشخص کنیم کدام z_i از بقیه بیش‌تر است و در واقع حاصل تفاضل بقیه‌ی z_jها از آن مثبت می‌شود. که این را نیز با مشخص کردن ضرایب مناسب و تابع فعالسازی Sign انجام می‌دهیم. دقت کنید که مقدار بایاس $b = -۲/۹$ مشخص می‌کنیم که z_i باید لزوما از روشنایی هر سه ناحیه‌ی دیگر بیش‌تر باشد. در لایه‌ی نهایی نیز به خروجی روشنایی هر ناحیه، ضریب شماره‌ی همان ناحیه را می‌دهیم. تابع فعالسازی این لایه را نیز تابع Step در نظر می‌گیریم: $Step(x) = \begin{cases} ۱ & if\ x \geqslant ۰ \\ ۰ & if\ x < ۰ \end{cases}$. چون فقط خروجی یک ناحیه ۱ و بقیه ۰ هستند، مقدار آن در ضریب مربوطه ضرب شده و خروجی نهایی این شبکه‌ی عصبی، شماره‌ی ناحیه‌ای است که میانگین روشنایی پیکسل‌های آن ماکسیمم است.

سوال ۴

فرض کنید ورودی‌های شبکه $X_۱, X_۲, X_۳$ باشند. در لایه‌ی میانی شبکه، هر نورون دو ورودی دارد و تابع فعالسازی آن $f(X) = \lfloor \frac{X}{۲} \rfloor$ است. بنابراین با توجه به ورودی‌های این نورون، خروجی آن این چنین است:

$$f(X) = \lfloor \frac{X}{۲} \rfloor = \begin{cases} ۰ & if\ X_۱ + X_۲ = ۰ \\ ۰ & if\ X_۱ + X_۲ = ۱ \\ ۱ & if\ X_۱ + X_۲ = ۲ \end{cases}$$

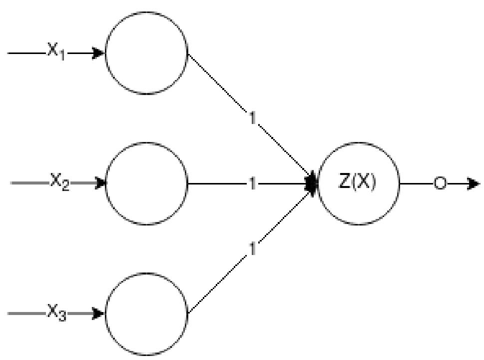
بنابراین واضح است که عملکرد این تابع مشابه تابع AND است. به‌طریق مشابه برای تابع فعالساز لایه‌ی خروجی داریم:

$$g(Y) = \lceil \frac{Y}{۳} \rceil = \begin{cases} ۰ & if\ Y_۱ + Y_۲ + Y_۳ = ۰ \\ ۱ & if\ Y_۱ + Y_۲ + Y_۳ = ۱ \\ ۱ & if\ Y_۱ + Y_۲ + Y_۳ = ۲ \\ ۱ & if\ Y_۱ + Y_۲ + Y_۳ = ۳ \end{cases}$$

که عملکرد این تابع مشابه تابع OR با سه ورودی‌است. بنابراین واضح است که خروجی این شبکه‌ی عصبی برحسب ورودی‌ها برابر است با:

$$O(X_۱, X_۲, X_۳) = (X_۱ \wedge X_۲) \vee (X_۱ \wedge X_۳) \vee (X_۲ \wedge X_۳)$$

که این رابطه بیانگر آن است که اگر دو ورودی مقدار ۱ داشته باشند، مقدار خروجی ۱ و در غیر این‌صورت مقدار خروجی ۰ می‌شود. پس می‌توان این شبکه‌ی عصبی را به شکل ساده‌تر زیر پیاده‌سازی کرد.



که تابع فعالسازی $Z(X) = \lfloor \frac{X}{۲} \rfloor$ است. واضح است که این شبکه‌ی عصبی خروجی مشابه دارد.

سوال ۵

توابع شبکه‌ی عصبی سوال را، از چپ به راست f, g, o می‌نامیم. بنابراین خواهیم داشت:

$$f = w_۱ x + w_۲ y$$

$$g = w_۴ f \times w_۳ z = w_۳ w_۴ z f$$

$$o = \sigma(w_۶ + w_۵ g)$$

اکنون با توجه به مقادیر داده شده و این‌که نمام وزن ها برابر $\frac{۱}{۲}$ هستند، عملیات feed forward را انجام می‌دهیم:

$$f = w_۱x + w_۲y = \frac{۱}{۲}۱ + \frac{۱}{۲}۳ = ۲$$

$$g = w_۴f \times w_۳z = w_۳w_۴zf = \frac{۱}{۲}\frac{۱}{۲}۷ \times ۲ = ۳/۵$$

$$o = \sigma(w_۶ + w_۵g) = \sigma(\frac{۱}{۲} + \frac{۱}{۲}\frac{۷}{۲}) \approx ۰/۹۰۵$$

حال به سراغ عملیات backpropagation می‌رویم. در حالت عادی، باید تابع cost function در اختیارمان قرار داده می‌شد تا با محاسبه‌ی مشتق جزئی آن نسبت به ضرایب w یعنی $\frac{\partial C}{\partial w}$ بتوانیم تاثیر تغییر هر یک را روی کم یا زیاد شدن cost function بسنجیم. اما در این سوال، مشتق خروجی شبکه یعنی o را نسبت به ضرایب w محاسبه می‌کنیم. بنابراین خواهیم داشت:

$$\sigma(x) = \frac{۱}{۱ + e^{-x}} \Rightarrow \frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(۱ + e^{-x})^۲} = \frac{۱}{۱ + e^{-x}}(۱ - \frac{۱}{۱ + e^{-x}})$$

$$\Rightarrow \frac{d\sigma(x)}{dx} = \sigma(x)(۱ - \sigma(x)) \tag{۱}$$

$$\xrightarrow{(۱)} \frac{\partial o}{\partial w_۶} = o(۱ - o) = ۰/۹۰۵(۱ - ۰/۹۰۵) \Rightarrow \boxed{\frac{\partial o}{\partial w_۶} = ۰/۰۸۶}$$

$$\xrightarrow{(۱)} \frac{\partial o}{\partial w_۵} = g(o(۱ - o)) = ۳/۵(۰/۹۰۵(۱ - ۰/۹۰۵)) \Rightarrow \boxed{\frac{\partial o}{\partial w_۵} = ۰/۳۰۱}$$

برای محاسبه‌ی باقی مشتقات جزئی، نیاز است یک لایه به عقب‌تر برگردیم:

$$\frac{\partial o}{\partial w_۴} = \frac{\partial o}{\partial g}\frac{\partial g}{\partial w_۴} \xrightarrow{(۱)} \frac{\partial o}{\partial w_۴} = w_۵(o(۱ - o)) \times w_۳zf = \frac{۱}{۲}۰/۰۸۶ \times \frac{۱}{۲}۷ \times ۲ \Rightarrow \boxed{\frac{\partial o}{\partial w_۴} = ۰/۳۰۱}$$

$$\frac{\partial o}{\partial w_۳} = \frac{\partial o}{\partial g}\frac{\partial g}{\partial w_۳} \xrightarrow{(۱)} \frac{\partial o}{\partial w_۳} = w_۵(o(۱ - o)) \times w_۴zf = \frac{۱}{۲}۰/۰۸۶ \times \frac{۱}{۲}۷ \times ۲ \Rightarrow \boxed{\frac{\partial o}{\partial w_۳} = ۰/۳۰۱}$$

مجددا برای محاسبه باقی مشتقات جزئی، نیاز است یک لایه به عقب‌تر برگردیم:

$$\frac{\partial o}{\partial w_۲} = \frac{\partial o}{\partial g}\frac{\partial g}{\partial f}\frac{\partial f}{\partial w_۲} \xrightarrow{(۱)} \frac{\partial o}{\partial w_۲} = w_۵(o(۱ - o)) \times w_۴w_۳zy = \frac{۱}{۲}۰/۰۸۶ \times \frac{۱}{۲}\frac{۱}{۲}۷ \times ۳ \Rightarrow \boxed{\frac{\partial o}{\partial w_۲} = ۰/۲۲۵}$$

$$\frac{\partial o}{\partial w_۱} = \frac{\partial o}{\partial g}\frac{\partial g}{\partial f}\frac{\partial f}{\partial w_۱} \xrightarrow{(۱)} \frac{\partial o}{\partial w_۱} = w_۵(o(۱ - o)) \times w_۴w_۳zx = \frac{۱}{۲}۰/۰۸۶ \times \frac{۱}{۲}\frac{۱}{۲}۷ \times ۱ \Rightarrow \boxed{\frac{\partial o}{\partial w_۱} = ۰/۰۷۵}$$

اکنون این عملیات نیز به پایان می‌رسد و با توجه به مقدار هدف خروجی و η یعنی learning rate، وزن یال‌های شبکه را با استفاده از gradient descent می‌توان آپدیت کرد.