

تمرین اول تئوری درس مبانی هوش محاسباتی

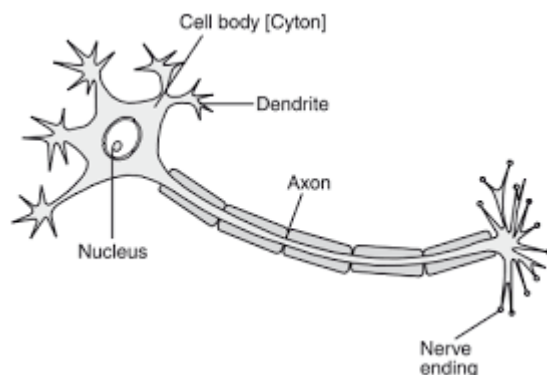
شبکه‌های عصبی

امیرحسین رجب‌پور 9731085

سوال 1:

(الف)

- **دندریت:** نقش ورودی را در نورون دارد که معمولاً ۳ یا ۴ تا می‌باشد. ورودی را به صورت ولتاژ دریافت می‌کند.
- **آکسون:** نقش خروجی را در نورون دارد. پیام به شکل پالس فرکانسی از آن خارج می‌شود.
- **هسته:** ارتباط بین ورودی‌ها (دندریت‌ها) و خروجی (آکسون) در نورون را برقرار می‌کند بدین صورت که مانند یک تابع پله عمل می‌کند یعنی اگر مجموع ولتاژ ورودی‌ها از یک حد آستانه بیشتر شود باعث می‌شود آکسون یک شلیک انجام دهد.



(ب) دندریت‌ها و آکسون‌ها از طریق یک اتصال شیمیایی (نه فیزیکی) به یکدیگر وصل می‌شوند که به این اتصال سیناپس گویند. در پاقع سیناپس سلول‌های عصبی را به

یکدیگر متصل می‌کند. از وظایف سیناپس می‌توان به تبدیل فرکانس به ولتاژ و همچنین تقویت و یا تضعیف خروجی آکسون اشاره کرد.

(پ)

- دندریت: معادل ورودی‌های مدل ما هستند.
- آکسون: معادل خروجی مدل ما است (y).
- سیناپس: معادل وزن‌های میان لایه‌های شبکه‌ی عصبی هستند (w ها).
- هسته: معادل محاسباتی است که انجام می‌شود تا ورودی‌ها به خروجی تبدیل شوند مانند تابع فعالیت (sigmoid, relu, etc).

(ت) چهار مورد از تفاوت‌های شبکه‌های عصبی بیولوژیکی و مصنوعی:

1. شبکه‌های عصبی بیولوژیکی دارای میلیون‌ها نورون و سیناپس می‌باشند و این تعداد بسیار بیشتر از شبکه‌های مصنوعی می‌باشد. شبکه‌های مصنوعی بسیار کوچکترند.

2. شبکه‌های عصبی بیولوژیکی دارای قدرت و سرعت بسیار بالاتر و بیشتر در پردازش اطلاعات نسبت به شبکه‌های مصنوعی هستند.

3. در شبکه‌های عصبی بیولوژیکی آکسون که در واقع همان خروجی ما هست به صورت پالس‌های فرکانسی است (یعنی مانند یک تابع پله) اما در شبکه‌های مصنوعی به این صورت نیست و ما برای انجام محاسبات نیاز به یک تابع مشتق پذیر به عنوان تابع فعالیت داریم.

4. در شبکه‌های مصنوعی برای یادگیری فرمول‌ها و روش‌هایی داریم اما این قضیه در شبکه‌های بیولوژیکی مشخص نیست که این شبکه‌ها دقیقا به چه صورت آموزش می‌بینند.

سوال 2:

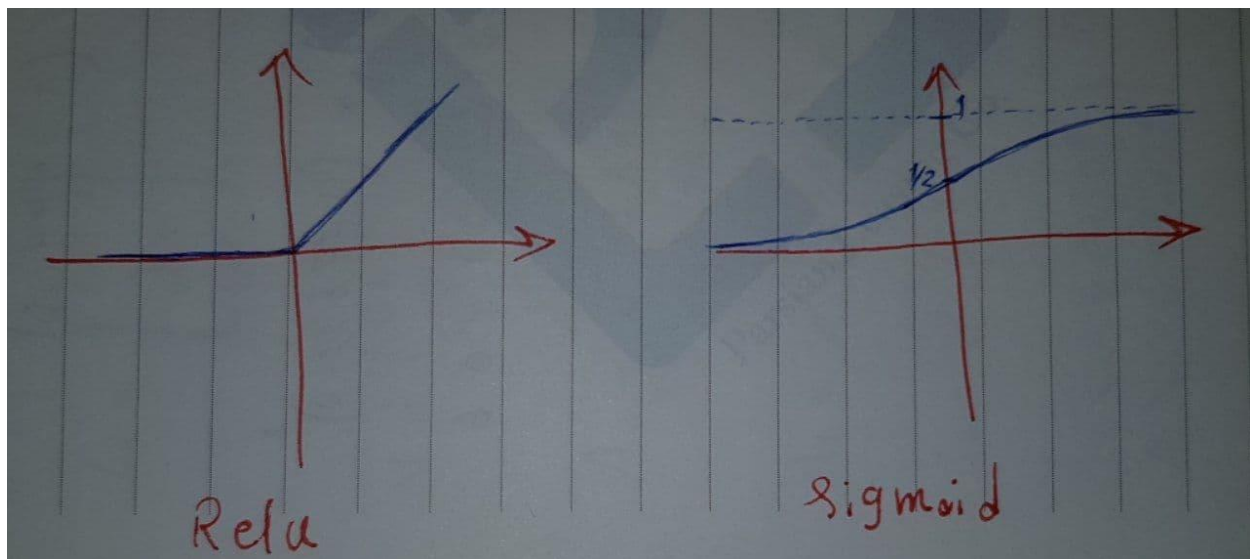
الف) افزودن بایاس این امکان را به تابع فعالسازی می‌دهد که بتواند علاوه بر مبدا از نقاط دیگر در نمودار نیز عبور کند و در واقع قدرت مانور این تابع را بیشتر می‌کند.

ب) تابع XOR را نمی‌توانیم مدل کنیم زیرا این تابع نیاز به دو خط جهت جدا کردن نقاطش دارد که بایاس در این زمینه کمکی به ما نمی‌کند و نیاز به دو خط داریم.

پ) خیر تنها با افزایش تعداد نوروها و لایه‌ها نمی‌توان این مشکل را حل کرد و به تابع فعالسازی نیز نیاز داریم. زیرا اگر تنها تعداد نوروها و لایه‌ها را زیاد کنیم جواب ما در نهایت باز خطی خواهد بود پس به تابع فعالسازی‌ای نیاز است تا از این خطی بودن خارج شویم.

ت) بله.

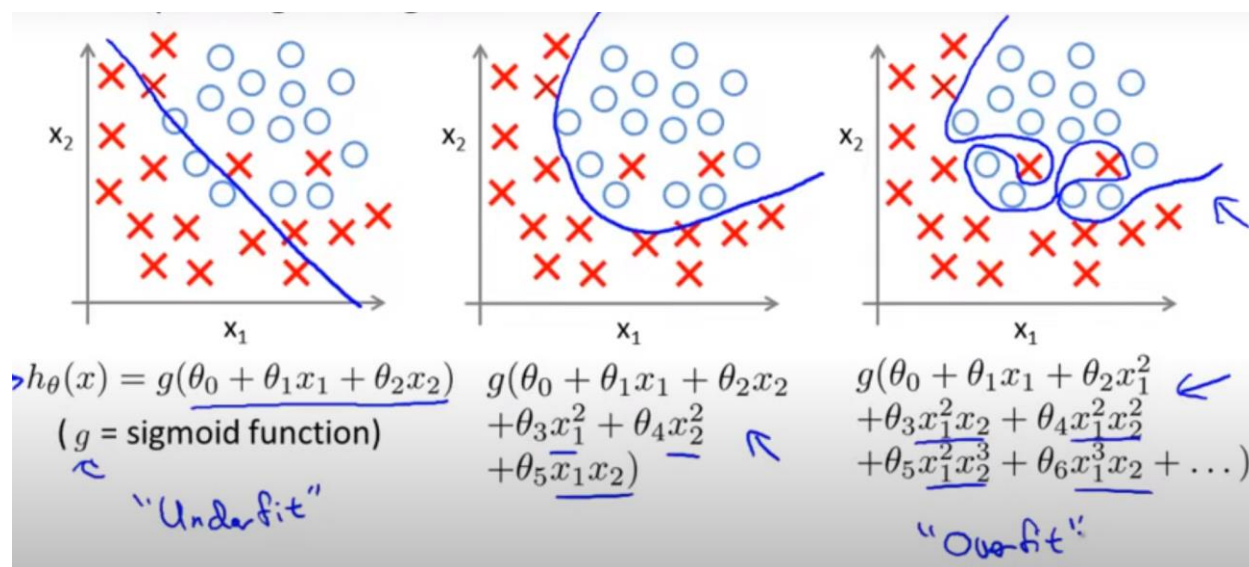
ث) این دو تابع هر دو مشتق پذیرند. همانطور که در شکل زیر مشخص است تابع سیگموید مقادیر بین ۰ و ۱ دارد ولی تابع رلو در اعداد منفی مقدار صفر دارد که این قضیه باعث می‌شود که اگر ورودی‌اش منفی باشد مقدار تابع صفر شده و این یک نقطه‌ی ضعف برای آن است زیرا در جاهایی که مقدار منفی می‌گیرد مدل به خوبی یاد نمی‌گیرد (برای جلوگیری و یا کم کردن اثر این مشکل از تابع leaky relu می‌توان استفاده کرد که در اعداد منفی صفر نیست و یک شیب ملایمی دارد). تابع سیگموید در مقادیر منفی خروجی نزدیک به صفر دارد و در این بازه به خوبی آموزش نمی‌بیند. هر دو تابع نیز صعودی هستند. تابع سیگموید برای binary classification مناسب‌تر است.



سوال 3:

الف) واریانس بالا در شبکه‌های عصبی به معنای حساسیت زیاد مدل به تغییرات کوچک روی داده‌های آموزشی می‌باشد یعنی اینکه مدل داده‌های آموزشی را بیش از حد یاد گرفته است و اگر داده‌ای به مدل نشان دهیم که تا به حال آن را ندیده باشد عملکرد ضعیفی از خود نشان خواهد داد در واقع مدل عملکرد پایینی دارد هنگامی که داده‌های واقعی و یا داده‌های اعتبارسنجی را می‌بیند.

اما از طرف دیگر **بایاس بالا** به این معنا است که مدل حتی روی داده‌های آموزشی نیز خوب یاد نگرفته است و عملکرد پایینی خواهد داشت در مواجهه با داده‌های آموزشی یا حتی واقعی و اعتبارسنجی نیز ضعیف خواهد بود.



همانطور که در این شکل مشخص است در نمودار سمت چپ بایاس بالا داریم و مدل ما نتوانسته است که خط مناسبی که نقاط را از یکدیگر جدا کند بکشد و در مدل سمت راست واریانس بالا داریم به این معنا که مدل ما خط مناسبی برای داده‌های آموزشی کشیده و خط تمام اشکال را از یکدیگر جدا کرده است اما اگر داده‌های جدیدی به مدل بدهیم که تا بحال آن‌ها را ندیده است عملکرد خوبی نخواهد داشت زیرا درگیر جزئیات داده‌های آموزشی شده است. اما در نمودار وسط مدل یادگیری مناسبی داشته.

ب) برای تشخیص واریانس بالا: اگر ارور روی داده‌های آموزشی پایین باشد اما روی داده‌هایی که مدل ما قبلاً ندیده مانند داده‌های اعتبارسنجی و یا تست زیاد باشد یعنی واریانس بالا داریم.

برای تشخیص بایاس بالا: اگر ارور روی داده‌های آموزشی ما بالا باشد (فاصله‌ی زیادی با خطای انسانی داشته باشد) در این صورت بایاس بالا داریم.

پ) می‌توان مدل را بیشتر آموزش داد (**train** کرد) و همچنین می‌توان از شبکه‌ی بزرگتری استفاده کرد (تعداد لایه‌ها یا نودهای لایه‌ها را زیاد کرد). و یا اینکه از

الگوریتم‌های بهینه‌سازی بهتری مانند Adam استفاده کرد. همچنین می‌توان مقدار فاکتور regularization را نیز کم کرد.

ت) در روش **drop out** به صورت رندوم یکسری از نودها را در هر لایه حذف می‌کنیم (به غیر از لایه‌ی آخر) بدین صورت در هر بار پردازش شبکه‌ی کوچکتری خواهیم داشت و جزییات کمتری در نظر گرفته می‌شوند که همین امر از وقوع واریانس بالا جلوگیری می‌کند و همچنین سرعت پردازش بالا می‌رود. این امر باعث می‌شود تا مدل ما زیاد از حد درگیر جزییات مجموعه‌ی آموزشی نشود و در کل آموزش بهتری داشته باشد تا عملکرد بهتری روی داده‌هایی که تا بحال ندیده مانند داده‌های اعتبارسنجی داشته باشد و مدل ما overfit نشود.

در روش **regularization** به تابع هزینه یک عبارت دیگر نیز اضافه می‌کنیم که به صورت زیر است:

$$\lambda \|w\|_2^2$$

که به این حالت L2 norm گویند که در آن لاندا همان متغیر regularization می‌باشد.

سوال 4:

(الف)

4- الف)

$$\frac{\partial \text{Cost}}{\partial W_1} = \frac{\partial \text{Cost}}{\partial O_1} \times \frac{\partial O_1}{\partial B_0} \times \frac{\partial B_0}{\partial A_0} \times \frac{\partial A_0}{\partial W_1}$$

$$= 2(O_1 - 1) \left(\frac{S(BT + b_3) + 1}{(B_0 t_0)} \right) B_0 \left(\frac{-S(AV + b_2) + 1}{(A_0 V_0 + A_1 V_1)} \right) V_0 \left(\frac{-S(WX + b_0) + 1}{(W_0 x_0 + W_1 x_1 + W_2 x_2)} \right) X_0$$

(ب)

4- ب)

$$W = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.2 \end{bmatrix}, \quad u = \begin{bmatrix} 0.3 \\ 0.4 \\ 0.5 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.6 \\ 0.7 \end{bmatrix}, \quad t = [0.9], \quad x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$A_0 = \sigma(Wx + b_0) = \sigma(0.9 + 0.5) = \sigma(1.4) = 0.8218$$

$$A_1 = \sigma(Vx + b_1) = \sigma(1.5) = 0.81757$$

$$B_0 = \sigma(A_0V_0 + A_1V_1 + b_2) =$$

$$\left. \begin{array}{l} A_0V_0 = 0.49308 \\ A_1V_1 = 0.572299 \end{array} \right\} B_0 = \sigma(1.565379) = 0.82712$$

$$O_1 = \sigma (BT + b_3) = \sigma (1.244408)$$

$$\Rightarrow O_1 = 0.77633$$

$$\text{Cost func} = (O_1 - Y_1)^2 = (0.77633 - 0)^2$$

$$\Rightarrow \text{Cost func} = 0.6026882689 \approx 0.6$$

سوال 5: (امتیازی)

سوال قبل را عدد تباری کنیم $(\alpha = 0.1)$

5-

$$W_0 = W_0 - \alpha \frac{\partial \text{Cost}}{\partial W_0} = 0.4 - (0.1 \times 0.003) = 0.3997$$

$$W_1 = W_1 - \alpha \frac{\partial \text{Cost}}{\partial W_1} = 0.3 - (0.1 \times 0.003) = 0.2997$$

$$W_2 = W_2 - \alpha \frac{\partial \text{Cost}}{\partial W_2} = 0.2 - (0.1 \times 0.003) = 0.1997$$