

## به نام ایزد منان

تمرین اول درس مبانی هوش محاسباتی، «شبکه‌های عصبی»



استاد درس: دکتر عبادزاده

بهار 1403 - دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر



نکاتی در مورد این تمرین نیاز به توجه و دقت دوستان دارد.

- 1- هرگونه کپی کردن باعث عدم تعلق نمره به تمامی افراد مشارکت کننده در آن می‌شود.
- 2- آخرین مهلت ارسال تمرین، ساعت ۲۳:۵۵ دقیقه روز 10 فروردین 1403 می‌باشد. این زمان با توجه به جمع‌بندی‌های صورت گرفته، شرایط و با توجه به سایر تمرین‌ها در نظر گرفته شده است و هیچ عنوان قابل تمدید نمی‌باشد.
- 3- دوستان فایل ارسالی خود را به صورت فشرده و به صورت «شماره دانشجویی HW1\_» مانند HW1\_97310000 نام گذاری کنید.
- 4- در صورت هرگونه سوال یا مشکل می‌توانید با تدریس‌یاران درس از طریق discussion کانال [https://t.me/CI\\_aut](https://t.me/CI_aut) در ارتباط باشید.

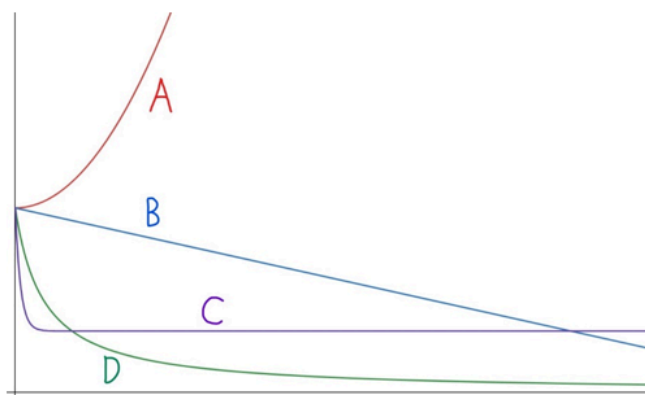
1) به سوال‌های زیر درباره‌ی توابع فعالیت پاسخ دهید.

- 1) در صورتی که از تابع فعالیت در شبکه عصبی استفاده نکنیم یا فقط از توابع خطی در یک شبکه عصبی چند لایه استفاده کنیم چه اتفاقی می افتد؟ با انجام محاسبه و مثال توضیح دهید.
- 2) تابع فعالیت Sigmoid چه مشکلاتی دارد؟
- 3) صفر محور نبودن یک تابع فعالیت چه مشکلی را ایجاد میکند؟
- 4) تابع Relu را تعریف کنید و مزایا و معایب آن را بنویسید.
- 5) تابع فعالیت Leaky ReLU چگونه تعریف میشود؟ کدام مشکل تابع فعالیت ReLU را حل میکند؟
- 6) توابع sigmoid و tanh را باهم مقایسه کنید و شباهت‌ها و تفاوت‌هایشان را بنویسید.
- 7) تابع Softmax را تعریف کنید و برای چه کاربردی در لایه آخر از آن استفاده میکنیم؟

2) به سوالات زیر پاسخ کوتاه بدهید:

- 1) چرا در شبکه‌های عصبی از تابع فعالیت پله استفاده نمی شود؟
- 2) میخواهیم یک شبکه عصبی طراحی کنیم که تصویر یک گربه یا سگ را دریافت کند و بگوید که این تصویر کدام حیوان است. به نظر شما در لایه آخر از Leaky ReLU استفاده کنیم یا Sigmoid؟ چرا؟
- 3) در صورتی که داده‌های کمی برای آموزش شبکه عصبی داشته باشیم دچار چه مشکلی خواهیم شد؟
- 4) فرمول Binary Cross Entropy چگونه تعریف می‌شود و در حل چه مسائلی از آن استفاده می شود؟
- 5) مراحل آموزش MLP را به طور مختصر شرح دهید.
- 6) چهار نرخ آموزش متفاوت در یک شبکه عصبی به کار رفته است و نتیجه آموزش آن‌ها در نمودار رو به رو نشان داده شده است. مقادیر این چهار نرخ آموزش را مقایسه کنید و به صورت مختصر توضیح دهید.

خطا



زمان

(3)

توضیح دهید که Cross Entropy , Sum of Squared Error (SSE) چگونه در ارزیابی عملکرد مدل‌های یادگیری ماشین مورد استفاده قرار می‌گیرند و هر یک در حل چه مسائلی مورد استفاده قرار می‌گیرند؟ مثال بزنید.

---

(4)

الف) تفاوت های Batch Gradient Descent و Stochastic Gradient Descent را بیان کنید و توضیح دهید که هر کدام در چه مواقعی استفاده می شوند.  
ب) نقش نرخ آموزش در الگوریتم backpropagation را بیان کنید و توضیح دهید که مقدار آن در طول الگوریتم چگونه تغییر میکند.

---

(5)

یک شبکه عصبی دو لایه‌ای ساده را در نظر بگیریم.  
این شبکه ورودی با اندازه 2، یک لایه مخفی با اندازه 3 و خروجی با اندازه 2 دارد.

ورودی:  $\begin{bmatrix} 6 \\ 10 \end{bmatrix}$

وزن و بایاس هم به صورت زیر تعریف میشود:

$$W_2 = \begin{bmatrix} 1.8 & 0.9 & -0.5 \\ 0.5 & -1.1 & 0.2 \end{bmatrix} \text{ و } W_1 = \begin{bmatrix} -0.3 & 0.2 \\ 0.6 & 0.6 \\ 0.4 & 0.8 \end{bmatrix}$$
$$b_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, b_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

1) مقدار  $z_1 = W_1x + b_1$  را محاسبه کنید.

2) فرض کنید برای فعال‌سازی لایه مخفی از تابع Sigmoid باشیم، فرمول تابع Sigmoid را بنویسید و

$a_1 = S(z_1)$  را محاسبه کنید.

3) مقدار  $z_2 = W_2a_1 + b_2$  را محاسبه کنید.

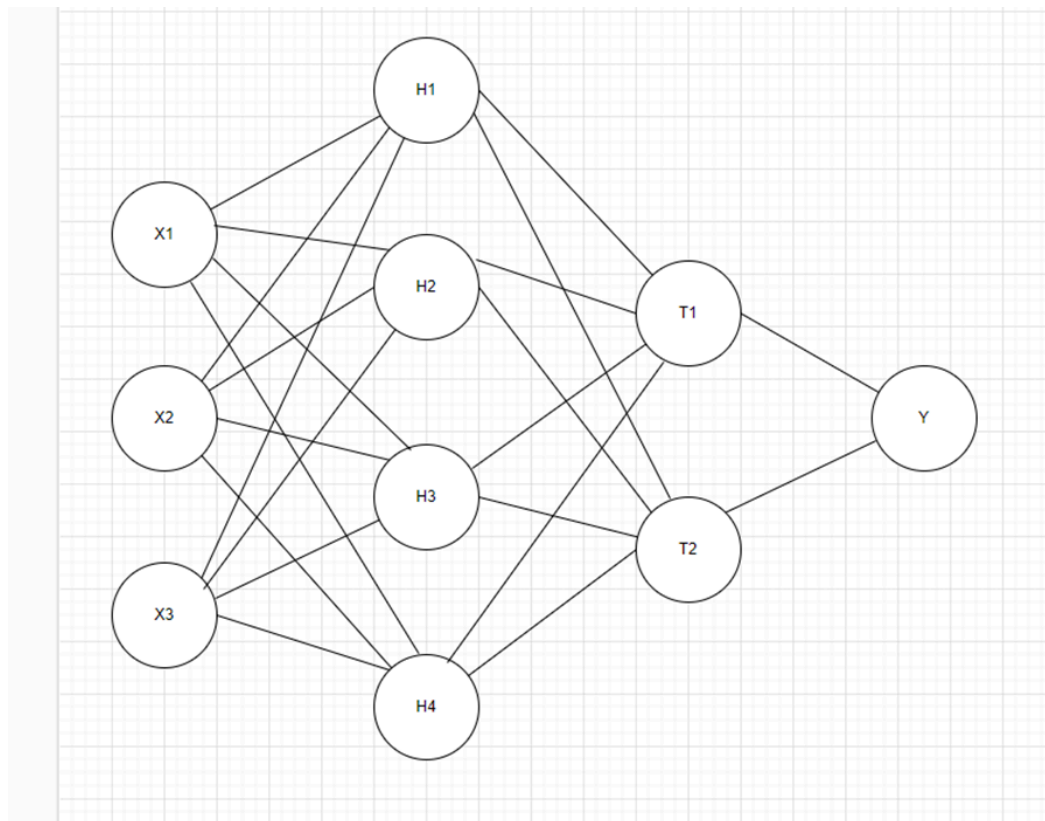
4) از تابع Softmax یا  $\sigma$  برای محاسبه  $\hat{y} = \sigma(z_2)$  استفاده کنید.

(6)

شبکه عصبی MLP زیر را در نظر بگیرید. این شبکه مطابق شکل دارای یک لایه ورودی، دو لایه پنهان و یک لایه خروجی است. لایه ورودی دارای سه پرسپترون، لایه پنهان اول دارای چهار پرسپترون، لایه پنهان دوم دارای 2 پرسپترون و لایه خروجی دارای یک پرسپترون می‌باشند. با توجه به توضیحات داده شده به سوالات زیر پاسخ دهید (تابع فعالیت لایه اول را  $\tanh$  و برای بقیه لایه‌ها را sigmoid در نظر بگیرید).

الف) با مقداردهی اولیه بایاس‌ها به صفر، وزن‌های بین لایه ورودی و پنهانی اول به 0.5، بقیه وزن‌ها به 1 و ورودی‌ها به شکل زیر، مقدار خروجی شبکه عصبی را محاسبه کنید.  
 $X_1 = 1, X_2 = 0.5, X_3 = 0.5$

ب) اگر خروجی مورد انتظار برای شبکه  $Y^* = 1$  باشد، با تابع هزینه کمترین مربعات خطا (MSE) و با نرخ یادگیری (Learning Rate) 0.2، وزن‌های خروجی H3 را بروزرسانی کنید (مراحل را به طور کامل بنویسید).



## سوال امتیازی:

شبکه عصبی زیر را در نظر بگیرید.

در این شبکه گره های تک دایره متغیر و گره های دو دایره توابع هستند.

$$h_1 = \frac{1}{1 + e^{-x_1 w_1 - x_2 w_2}} \quad \text{برای مثال}$$

اگر تابع هزینه و مشتق تابع سیگموئید به صورت زیر باشد،

$$L(y, \hat{y}) = \|y - \hat{y}\|^2 \quad \sigma'(x) = \sigma(x)(1 - \sigma(x))$$

مشتق زنجیره ای و عبارت جبری که برای محاسبه نیاز هست را بیابید. این مقدار چه کاربردی در آموزش شبکه عصبی دارد؟

