



**دانشگاه صنعتی امیرکبیر**  
( پلی تکنیک تهران )

تمرین اول درس هوش محاسباتی

استاد عبادزاده

عطیه براتی نیا 9631010

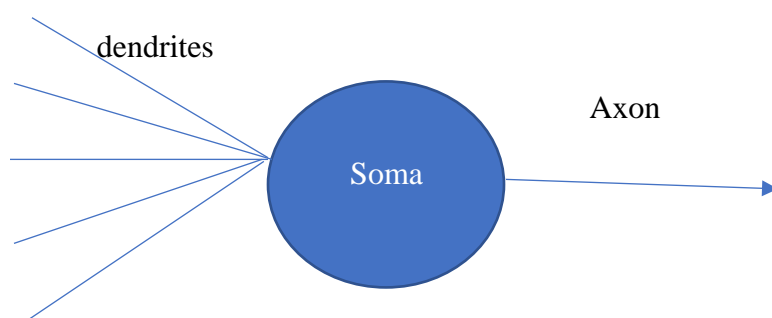
## سوال ۱

**Dendrites:** ورودی نوروں است و پالس های الکتریکی که شامل داده میشود را از سیناپس های نوروں های دیگر میگیرد و به soma میفرستد.

**Soma:** قسمت اصلی نوروں soma است که پالس های الکتریکی و یا در واقع داده ها با ترکیب کردن با هم پردازش میشوند.

**Axon:** آکسون داده ای که پردازش نوروں بوده را از soma میگیرد و به سیناپس میفرستد.

**Synapse:** قسمتی است که داده ی نوروں را به هزاران نوروں دیگر که با آنها ارتباط دارد میفرستد. سیناپس ها به عنوان تضعیف کننده یا تقویت کننده عمل میکنند، انگار خروجی آکسون را در یک ضربی ضرب میکنند.



دندریت ها به عنوان ورودی یک نوروں مصنوعی و آکسون ها به عنوان خروجی نوروں های مصنوعی مدل سازی میشوند، تعداد آکسون ها میتواند خیلی باشد ولی برای سادگی فقط یکی نشان داده شده است. سیناپس ها به عنوان وزن اتصالات گره ها عمل میکنند. soma نیز به صورت ترکیبات خطی که روی ورودی ها انجام میشود مدل سازی میشود.

سیگنال های بین نوروں ها الکتریکی است. این سیگنال ها از طریق آکسون یک نوروں پس از تقویت یا تضعیف شدن از طریق سیناپس ها وارد دندریت نوروں دیگر میشود و در سوما ی نوروں پردازش شده و به آکسون فرستاده میشود و به همین ترتیب ادامه پیدا میکند.

## سوال ۲

نرخ یادگیری میزان قدم های ما را برای رسیدن به مینیمم خطای مطلوب نشان میدهد. نرخ یادگیری در منفی گرادیان ضرب میشود و با وزن مورد نظر جمع میشود تا وزن بهتر با خطای کمتر به دست آید. نرخ یادگیری اگر

خیلی کوچک مقداردهی شود باعث میشود تا خیلی زمان ببرد تا به مینیمم برسیم یا ممکن است در مینیمم محلی گیر کنیم و اگر نرخ یادگیری را خیلی بزرگ مقداردهی کنیم باعث میشود که الگوریتم نوسانی و ناپایدار شود. مقداردهی این پارامتر باید متناسب با شناختی باشد که از نمودار داریم و عدد خیلی بزرگ یا خیلی کوچک انتخاب نکنیم. معمولاً نرخ یادگیری بین 0.1 تا 0.01 انتخاب میشود. در حالت کلی به نظرم بهتر است وقتی که شیب گرادیان خیلی زیاد است، یعنی از مینیمم خیلی فاصله داریم پس نرخ یادگیری را زیاد بگیریم و وقتی شیب گرادیان خیلی کم است یعنی داریم به گرادیان صفر و مینیمم خطا نزدیک میشویم پس نرخ یادگیری را کم بگیریم و کلاً نرخ یادگیری را متناسب با گرادیان اتخاذ کنیم.

افزایش عمق شبکه باعث دقت بیشتر شبکه میشود و باعث میشود تا بتوان دسته بندی را با خطوط شکسته بیشتری انجام داد. اما باید توجه کرد که افزایش بیش از حد عمق باعث overfit شدن داده های آموزشی میشود این بدان معناست که دقت آن برای داده های آموزشی بسیار خوب میشود اما باعث میشود که شبکه خیلی خاص شود و از عمومیت شبکه عصبی میکاهد و در نتیجه میزان خطا برای داده های جدید را به شدت زیاد میکند.

نرون ها نشان دهنده ی متغیرها هستند و افزایش تعداد نرون ها باعث افزایش تعداد متغیرها میشود. تا حد امکان باید تلاش کرد تعداد نرون ها یا به عبارت بهتر تعداد متغیرها را کم نگه داشت تا از پیچیدگی شبکه جلوگیری کنیم ولی گاه به دلیل پیچیدگی شبکه به تعداد نرون بیشتری نیاز هست.

برای اینکه متوجه بشیم که مدل دچار بیش برآزش شده است یا نه باید داده های تست را روی آن به مرور آزمایش کنیم در اوایل ایپاک ها، شبکه عصبی داده های آموزشی را خوب مدل میکند و خطا بر روی داده های تست کاهش میابد ولی زمانی که دیدیم که داده های آموزشی بهتر مدل میشوند ولی خطا بر روی داده های تست به جای کاهش شروع به افزایش کرده اند یعنی مدل دچار بیش برآزش شده است. وقتی به چنین موقعیتی رسیدیم باید آموزش را متوقف کنیم چون بهترین آموزشی که شبکه عصبی موردنظر میتواند داشته باشد را به دست آورده ایم و ادامه دادن فقط باعث بیش برآزش میشود.

### سوال ۳

مدل پرسپترون NOR:  $Y = X_1(-0.5) + X_2(-0.5) + 0.3$

اگر  $Y > 0$  بود خروجی 1 خواهد بود اگر  $Y < 0$  بود خروجی صفر خواهد بود.

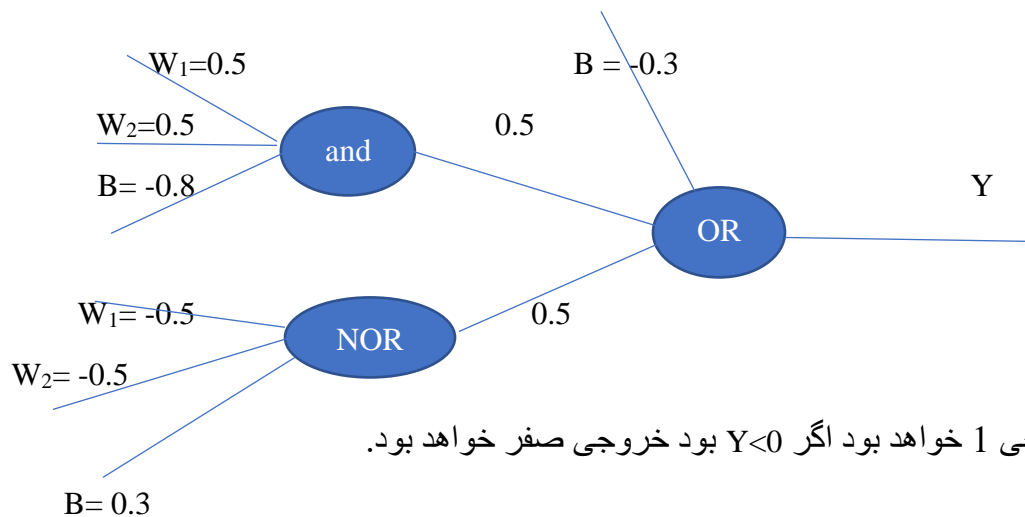
مدل پرسپترون NAND:  $Y = X_1(-0.5) + X_2(-0.5) + 0.8$

اگر  $Y > 0$  بود خروجی 1 خواهد بود اگر  $Y < 0$  بود خروجی صفر خواهد بود.

مدل پرسپترون چند لایه است:  $X_1X_2 + X_1'X_2'$

این معادله به این صورت خواهد بود:->

$$Y = (X_1(-0.5) + X_2(-0.5) + 0.3)(0.5) + (X_1(0.5) + X_2(0.5) - 0.8)(0.5) - 0.3$$



اگر  $Y > 0$  بود خروجی 1 خواهد بود اگر  $Y < 0$  بود خروجی صفر خواهد بود.

#### سوال ۴

##### یادگیری با نظارت

یادگیری با نظارت، یک روش عمومی در یادگیری ماشین است که در آن به یک سیستم، مجموعه‌ای از جفت‌های ورودی - خروجی ارائه شده و سیستم تلاش می‌کند تا تابعی از ورودی به خروجی را فرا گیرد. یادگیری تحت نظارت نیازمند تعدادی داده ورودی به منظور آموزش سیستم است. یادگیری تحت نظارت خود به دو دسته تقسیم می‌شود: رگرسیون و طبقه‌بندی رگرسیون آن دسته از مسائل هستند که خروجی یک عدد پیوسته یا یک سری اعداد پیوسته هستند مانند پیش‌بینی قیمت خانه بر اساس اطلاعاتی مانند مساحت، تعداد اتاق خوابها، و غیره و دسته طبقه‌بندی به آن دسته از مسائل گفته می‌شود که خروجی یک عضو از یک مجموعه باشد مانند پیش‌بینی اینکه یک ایمیل هرزنامه هست یا خیر یا پیش‌بینی نوع بیماری یک فرد از میان ۱۰ بیماری.

##### یادگیری بدون نظارت

یادگیری بی نظارت یا یادگیری بدون نظارت (Unsupervised machine learning)، یکی از انواع یادگیری در یادگیری ماشینی است. اگر یادگیری بر روی داده‌های بدون برچسب و برای یافتن الگوهای پنهان در این داده‌ها

انجام شود، یادگیری، بدون نظارت خواهد بود. از انواع یادگیری بدون نظارت می‌توان به الگوریتمهای خوشه‌بندی (Clustering)، تخصیص پنهان دیریکله (LDA) و جاسازی لغات (Word Embedding) اشاره کرد. از یادگیری نظارت نشده در دنیای امروز می‌توان مثال‌های متعددی زد. یکی از پرکاربردترین آن‌ها پیشنهادهایی است که به کاربران در شبکه‌های اجتماعی داده می‌شود. بعنوان مثال در اینستاگرام داده‌های فوق‌العاده زیادی از هر فرد در مورد علایق، کسانی که دنبال می‌کند، دنبال‌کنندگان او و ... وجود دارد. اینستاگرام براساس داده‌هایی که از کاربر بدست می‌آورد، ویژگی‌های او را تعیین کرده و با خوشه‌بندی می‌کند. در نهایت با توجه به خوشه‌ای که کاربر درون آن قرار گرفته است، پیشنهادهای متعددی به وی در جهت درگیر کردن بیشتر او با این شبکه اجتماعی می‌دهد.

### یادگیری تقویتی

یادگیری تقویتی (Reinforcement learning | RL) نوع مهمی از یادگیری ماشین است که در آن یک عامل می‌آموزد چگونه در محیط با انجام اقدامات و دیدن نتایج آن‌ها رفتار کند. هدف یادگیری تقویتی این است که چگونه عامل‌های نرم افزاری، باید یک عمل را مناسب محیط انتخاب کنند تا پاداش بهینه بیشینه شود. این رشته به دلیل کلی بودن، در بسیاری از رشته‌های دیگر از جمله نظریه بازی، تئوری کنترل، تحقیق در عملیات، تئوری اطلاعات، بهینه سازی مبتنی بر شبیه سازی، سیستم‌های چند عامل، هوشمند جمعی، آمار و الگوریتم‌های ژنتیکی مورد مطالعه قرار می‌گیرد.

شبکه عصبی در یادگیری با نظارت: یک مجموعه داده به عنوان مجموعه آموزش (training set) با پاسخ صحیح برای آموزش شبکه استفاده می‌شود. قانون یادگیری (learning rule) بر اساس اصلاح و تنظیم وزن‌ها و بایاس (bias) شبکه برای نزدیکی به مقادیر پاسخ تولید شده و از این قانون برای پیش‌بینی داده‌های جدید استفاده می‌شود. روش یادگیری با نظارت برای مدل سازی، طبقه بندی، رگرسیون و درون یابی به کار می‌رود.

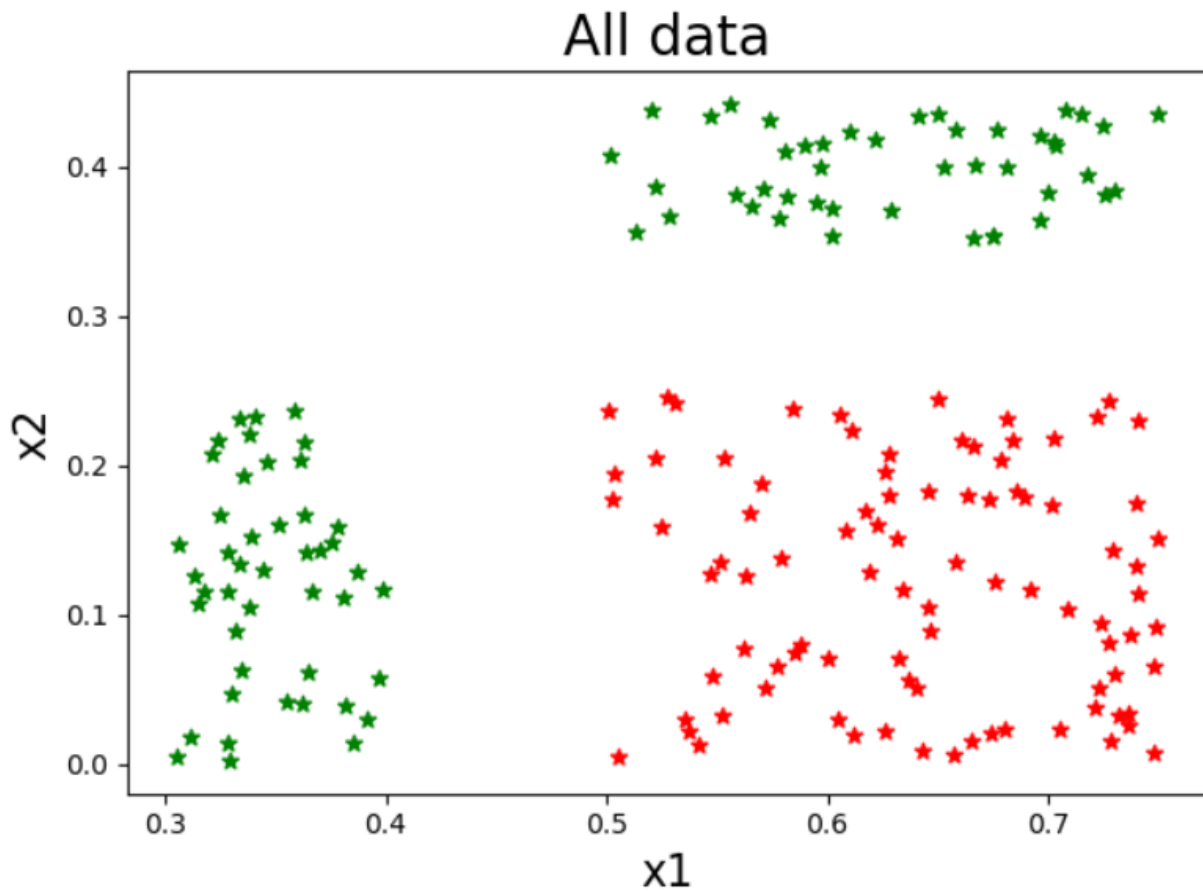
شبکه عصبی در یادگیری بدون نظارت: در روش یادگیری بدون نظارت داده‌های اولیه یا ورودی‌های شبکه بدون پاسخ یا هدف (Target) هستند و وزن‌ها و بایاس شبکه فقط بر اساس ورودی‌ها اصلاح می‌گردند. بیشتر روش‌های یادگیری بدون نظارت برای خوشه بندی (Clustering) استفاده می‌شوند که در آن یافتن ساختار و رابطه درون مجموعه داده‌ها بررسی می‌شود.

شبکه عصبی در یادگیری تقویتی: در این شبکه‌ها آموزش از طریق مشاهده و سعی و خطا صورت می‌گیرد. در واقع با توجه به تعاملی که با محیط پیرامون خود دارد عملی انجام داده و نتیجه عمل را به صورت پاداش و تنبیه برای نزدیکی و دوری به هدف مشخص می‌شود. شبکه کارهایی انجام می‌دهد که منجر به کسب پاداش بیشتری می‌شود.

## بخش دوم - مسائل برنامه نویسی

سوال یک)

در ابتدا ۱۸۰ داده را میتوان در شکل زیر دید، آنهایی که برچسب صفر دارند با قرمز و آنهایی که برچسب یک دارند با سبز مشخص شده اند.



سپس چهل داده اول به عنوان تست و ۱۴۰ داده بعدی به عنوان داده‌ی آموزشی انتخاب شد.

سوال دو)

$$\frac{d\text{cost}}{dW} = \frac{d\text{cost}}{dy} * \frac{dy}{dW} = \frac{d\text{cost}}{dy} * \frac{dS(WX + b)}{d(WX + b)} * \frac{d(WX + b)}{dW}$$

$$= \left( - \sum_{\forall x} \left[ y_T * \frac{1}{y \ln(10)} + (1 - y_T) \frac{-1}{(1 - y) \ln(10)} \right] \right) * S(WX + b) (1 - S(WX + b)) * X$$

$$\frac{dcost}{db} = \frac{dcost}{dy} * \frac{dy}{db} = \frac{dcost}{dy} * \frac{dS(WX + b)}{d(WX + b)} * \frac{d(WX + b)}{db}$$

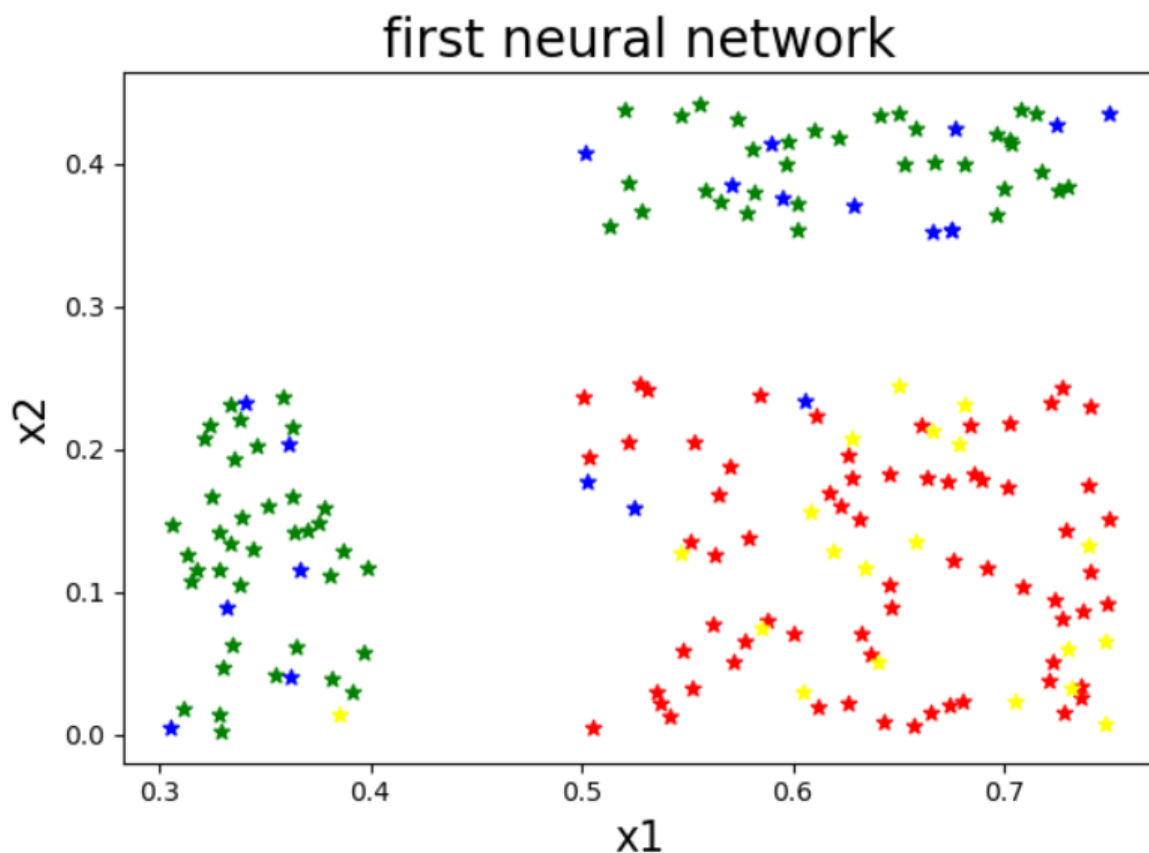
$$= \left( - \sum_{\forall x} \left[ y_T * \frac{1}{y \ln(10)} + (1 - y_T) \frac{-1}{(1 - y) \ln(10)} \right] \right) * S(WX + b) (1 - S(WX + b))$$

سوال ۳

کد پایتون در کنار این فایل قابل دسترسی است، متد train1 مربوط به آموزش این شبکه عصبی است. طبق بررسی های مختلف lr=0.1 و n\_epoch=2000 بهترین نتیجه ممکن را داد.

سوال ۴

دقت دسته بندی برای ۴۰ داده اول که به عنوان تست انتخاب شده بود 0.9 به دست آمد.



ستاره های سبز داده های آموزشی با برچسب یک هستند و ستاره های قرمز داده های آموزشی با برچسب صفر هستند. ستاره های آبی داده های آزمایشی هستند که برچسب آنها توسط شبکه عصبی یک دسته بندی شد و ستاره های زرد داده های آزمایشی هستند که برچسب آنها توسط شبکه عصبی صفر دسته بندی شد.

## سوال ۵

$$\begin{aligned} \diamond \frac{dcost}{dW} &= \frac{dcost}{dy} * \frac{dy}{dz_0} * \frac{dz_0}{dW} = 2(y - y_T) * \frac{dy}{d(ZU+b_2)} * \frac{d(ZU+b_2)}{dz_0} * \frac{dz_0}{d(XW+b_0)} * \frac{d(XW+b_0)}{dW} = \\ &= 2(y - y_T) * S(ZU + b_2)(1 - S(ZU + b_2)) * u_0 * S(XW + b_0)(1 - S(XW + b_0)) * X \end{aligned}$$

$$\begin{aligned} \diamond \frac{dcost}{dV} &= \frac{dcost}{dy} * \frac{dy}{dz_1} * \frac{dz_1}{dV} = 2(y - y_T) * \frac{dy}{d(ZU+b_2)} * \frac{d(ZU+b_2)}{dz_1} * \frac{dz_1}{d(XV+b_1)} * \frac{d(XV+b_1)}{dV} = \\ &= 2(y - y_T) * S(ZU + b_2)(1 - S(ZU + b_2)) * u_1 * S(XV + b_1)(1 - S(XV + b_1)) * X \end{aligned}$$

$$\diamond \frac{dcost}{dU} = \frac{dcost}{dy} * \frac{dy}{d(ZU+b_2)} * \frac{d(ZU+b_2)}{dU} = 2(y - y_T) * S(ZU + b_2)(1 - S(ZU + b_2)) * Z$$

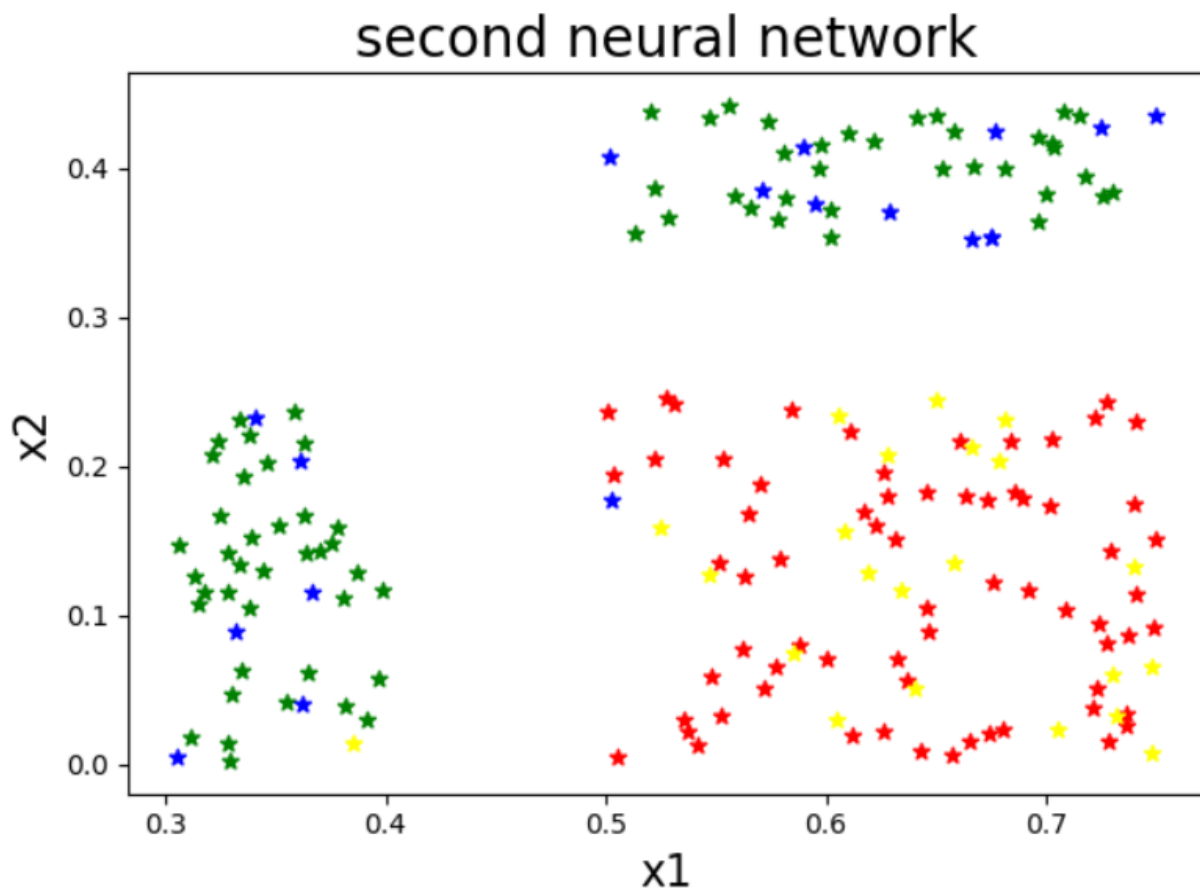
$$\begin{aligned} \diamond \frac{dcost}{db_0} &= \frac{dcost}{dy} * \frac{dy}{dz_0} * \frac{dz_0}{db_0} = 2(y - y_T) * \frac{dy}{d(ZU+b_2)} * \frac{d(ZU+b_2)}{dz_0} * \frac{dz_0}{d(XW+b_0)} * \frac{d(XW+b_0)}{db_0} = \\ &= 2(y - y_T) * S(ZU + b_2)(1 - S(ZU + b_2)) * u_0 * S(XW + b_0)(1 - S(XW + b_0)) \end{aligned}$$

$$\begin{aligned} \diamond \frac{dcost}{db_1} &= \frac{dcost}{dy} * \frac{dy}{dz_1} * \frac{dz_1}{db_1} = 2(y - y_T) * \frac{dy}{d(ZU+b_2)} * \frac{d(ZU+b_2)}{dz_1} * \frac{dz_1}{d(XV+b_1)} * \frac{d(XV+b_1)}{db_1} = \\ &= 2(y - y_T) * S(ZU + b_2)(1 - S(ZU + b_2)) * u_1 * S(XV + b_1)(1 - S(XV + b_1)) \\ \diamond \frac{dcost}{db_2} &= \frac{dcost}{dy} * \frac{dy}{d(ZU+b_2)} * \frac{d(ZU+b_2)}{db_2} = 2(y - y_T) * S(ZU + b_2)(1 - S(ZU + b_2)) \end{aligned}$$



کد پایتون در کنار این فایل قابل دسترسی است، متد train2 مربوط به آموزش این شبکه عصبی است. طبق بررسی های مختلف  $lr=0.05$  و  $n\_epoch=3000$  بهترین نتیجه ممکن را داد. دقت این شبکه عصبی برای ۴۰ داده ای تست 0.95 به دست آمد.

ستاره های سبز داده های آموزشی با برجسب یک هستند و ستاره های قرمز داده های آموزشی با برجسب صفر هستند. ستاره های آبی داده های آزمایشی هستند که برجسب آنها توسط شبکه عصبی یک دسته بندی شد و ستاره های زرد داده های آزمایشی هستند که برجسب آنها توسط شبکه عصبی صفر دسته بندی شد.



## سوال ۶

این شبکه چون لایه مخفی دارد میتواند دقیق تر از شبکه قبل عمل کند، به عبارت دیگر شبکه قبلی فقط یک خط صاف است در حالی که این شبکه میتواند یک خط شکسته باشد بنابراین میتواند دقت بهتری در یادگیری داشته باشد و دسته بندی بهتری ارائه دهد.

## سوال ۷

دو لایه مخفی، در لایه اول چهار نرون و در لایه دوم دو نرون استفاده کنیم. لایه‌ی مخفی اول باعث میشود تا دو خط شکسته ایجاد شود و لایه مخفی دوم باعث میشود تا این دو خط شکسته بهم متصل شوند و در نتیجه بتوان دسته بندی درستی انجام داد.