

به نام خدا

مهدی فیروزبخت

۴۰۰۱۳۱۰۲۷

تمرین سری اول درس شبکه های عصبی

(۱)

برای پیاده سازی ۲ مجموعه داده جداپذیر خطی با توزیع گوسی ، از پارامتر های زیر استفاده کرده ایم :

برای داده های اول:

$\text{Mean} = [ -50 , 0 ]$

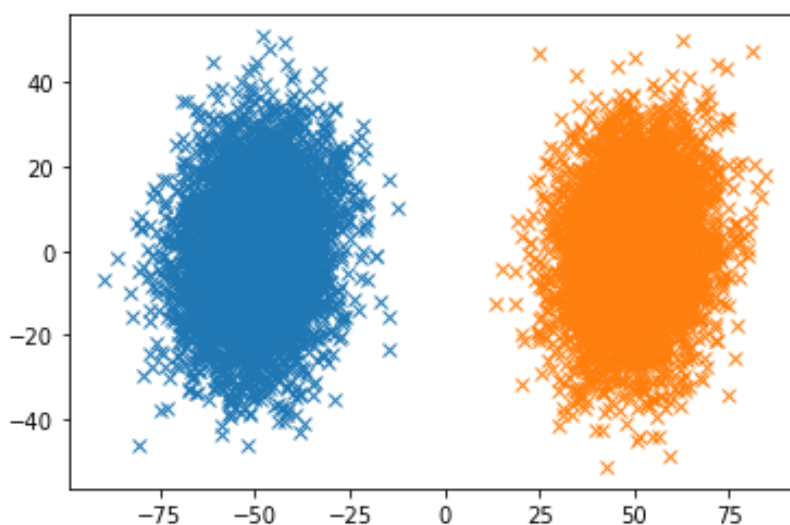
$\text{Cov} = [ 100 , 20 ] , [ 20 , 200 ]$

برای داده های دوم:

$\text{Mean} = [ 50 , 0 ]$

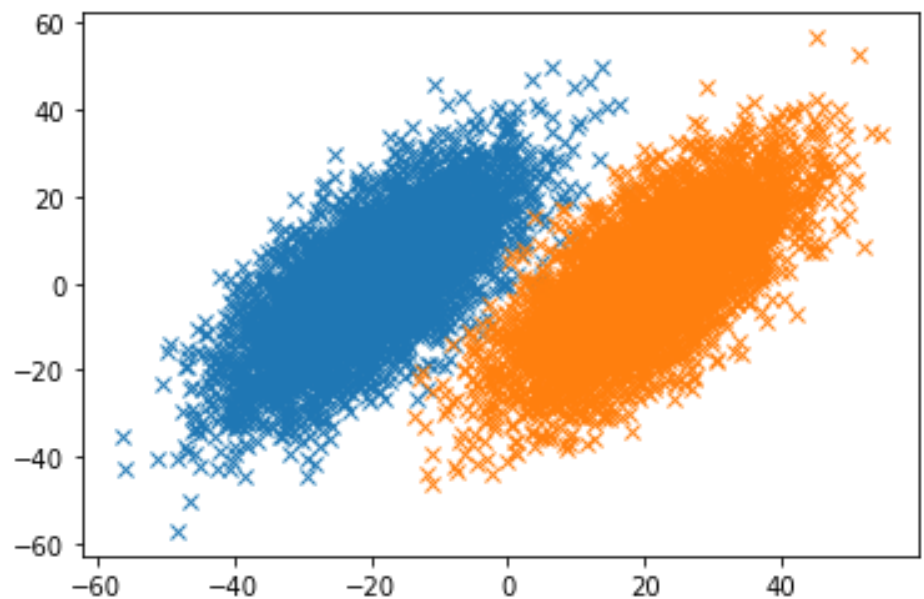
$\text{Cov} = [ 100 , 20 ] , [ 20 , 200 ]$

در این صورت شکل داده ها به صورت زیر خواهد بود :



برای اینکه داده ها حتما جدا پذیر خطی باشد ، سعی کرده ایم مقدار میانگین آن را از یکدیگر تا حدی دورتر قرار دهیم . همچنین در ماتریس کواریانس سعی کرده ایم پراکندگی را به گونه ای قرار دهیم که از فاصله میانگین ها بیشتر نشود تا داده ها درون یکدیگر قرار نگیرند . اگر برای مثال میانگین داده ها را به یکدیگر نزدیک کنیم و کواریانس آنها را نیز زیاد کنیم داده ها درون یکدیگر قرار میگیرند .

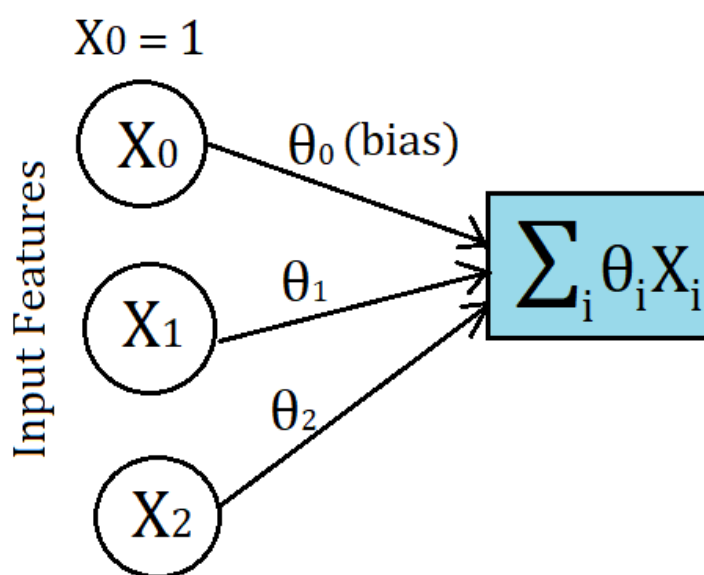
داده های زیر میانگین  $[-20, 0]$  و  $[20, 0]$  و ماتریس کواریانس  $[100, 200]$  ،  $[100, 100]$  را دارا هستند که به این صورت درون یکدیگر قرار گرفته اند :



برای ساخت شبکه عصبی پرسپترونی ، در کلاس Perceptron یک شبکه پرسپترونی ایجاد کرده ایم.

در این شبکه به عنوان ورودی نرخ یادگیری ، تعداد تکرار ها و نوع تابع فعالیت را دریافت کرده ایم. در بخش init بسته به تابع فعالیت ورودی ، اگر نوع آن sig باشد ، تابع فعالیت سیگمویدی ، اگر tanh باشد تابع فعالیت tanh و اگر unit باشد تابع فعالیت unit step خواهد شد.

ساختار ورودی داده ها به این صورت است که ابتدا داده های  $X$  که به صورت ۲ بعدی بوده اند وارد تابع fit شده و در ادامه در هنگام استفاده از قانون perceptron برای bias داده  $x$  برابر ۱ در نظر گرفته میشود. به طور کلی مانند شکل زیر خواهد بود:



برای مقدار اولیه وزن ها ، تمام مقادیر اولیه آن را برابر ۱ در نظر گرفته ایم.

توابع فعالیت ، ۳ نوع تابع فعالیت برای این شبکه عصبی ایجاد کرده ایم. تابع سیگموید ، تابعی است که از قانون سیگموید و فرمول آن استفاده کرده است که به این صورت است :

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

برای تابع tanh نیز از فرمول آن استفاده کرده ایم که اگر از تانژانت مقادیر استفاده میکند.

برای تابع unit step ، اگر مقدار تابع از ۰ کمتر بود ، مقدار ۰ را قرار می‌دهیم اما اگر مقدار آن بیشتر از ۰ باشد ، به جای آن مقدار ۱ را قرار می‌دهیم.

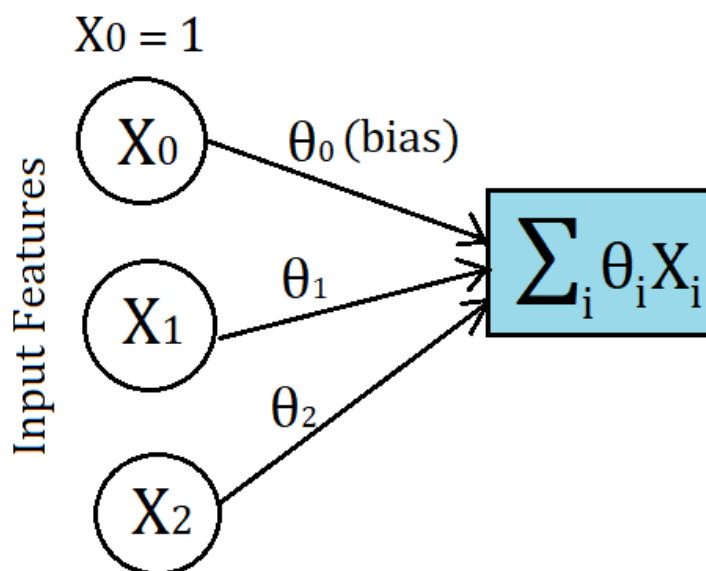
برای محاسبه مقدار وزن ها نیز از فرمول آموزش پرسپترون استفاده کرده ایم که به صورت زیر است :

$$\theta_j := \theta_j + \alpha \left( y^{(i)} - h_{\theta}(x^{(i)}) \right) x_j^{(i)}.$$

برای ساخت شبکه عصبی ادالاین ، ، در کلاس Adaline یک شبکه ادالاین ایجاد کرده ایم.

در این شبکه به عنوان ورودی نرخ یادگیری ، تعداد تکرار ها و نوع تابع فعالیت را دریافت کرده ایم. در بخش init بسته به تابع فعالیت ورودی ، اگر نوع آن sig باشد ، تابع فعالیت سیگموئیدی ، اگر tanh باشد تابع فعالیت tanh و اگر unit باشد تابع فعالیت unit step فعال خواهد شد.

ساختار ورودی داده ها به این صورت است که ابتدا داده های  $X$  که به صورت ۲ بعدی بوده اند وارد تابع fit شده و در ادامه در هنگام استفاده از قانون adaline برای bias داده  $x$  برابر ۱ در نظر گرفته میشود. به طور کلی مانند شکل زیر خواهد بود:



برای مقدار اولیه وزن ها ، تمام مقادیر اولیه آن را برابر ۱ در نظر گرفته ایم. توابع فعالیت ، ۳ نوع تابع فعالیت برای این شبکه عصبی ایجاد کرده ایم. تابع سیگموئید ، تابعی است که از قانون سیگموئید و فرمول آن استفاده کرده است که به این صورت است :

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

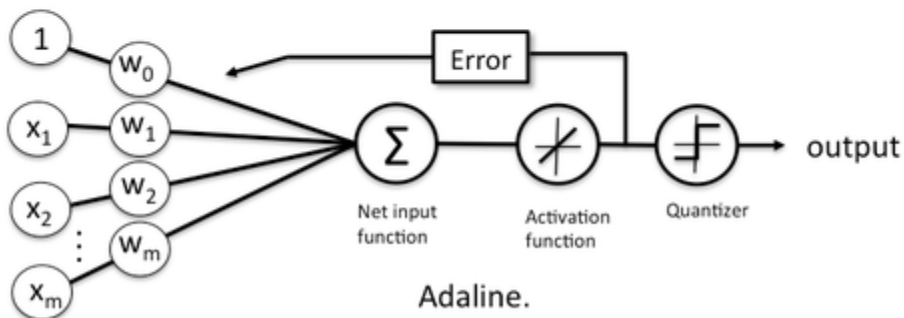
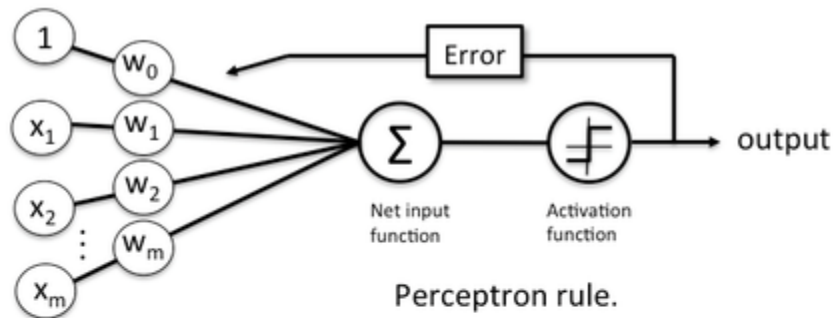
برای تابع tanh نیز از فرمول آن استفاده کرده ایم که اگر از تانژانت مقادیر استفاده میکند.

برای تابع unit step ، اگر مقدار تابع از ۰ کمتر بود ، مقدار ۰ را قرار میدهیم اما اگر مقدار آن بیشتر از ۰ باشد ، به جای آن مقدار ۱ را قرار میدهیم.

برای محاسبه مقدار وزن ها در مرحله اول از مقادیر ورودی استفاده میشود و در مرحله های بعدی از مقدار محاسبه شده بدون رد شدن از تابع فعالیت استفاده میشود.

تفاوت اصلی این دو شبکه در این است که شبکه پرسپترونی ، در مرحله اول از داده های ورودی استفاده میکند و در مراحل بعدی از پاسخ شبکه ، که از درون تابع فعالیت رد شده است استفاده میکند در حالتی که در شبکه ادالاین ، در مرحله اول از داده های ورودی و در مراحل بعدی از پاسخ شبکه بدون رد شدن از درون تابع فعالیت استفاده میشود.

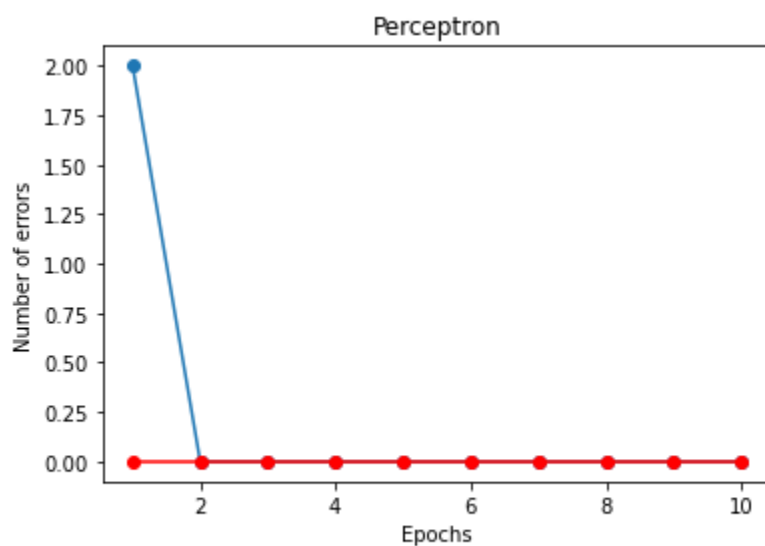
در کل میتواند برای تفاوت این دو شبکه به شکل زیر اشاره کرد :



(۳)

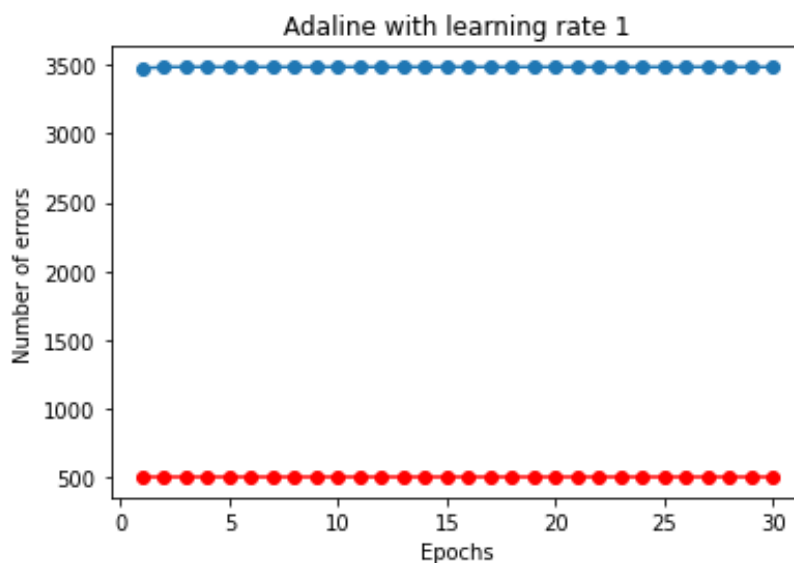
تابع فعالیت unit :

نرخ یادگیری : ۱



Accuracy is : 1.0

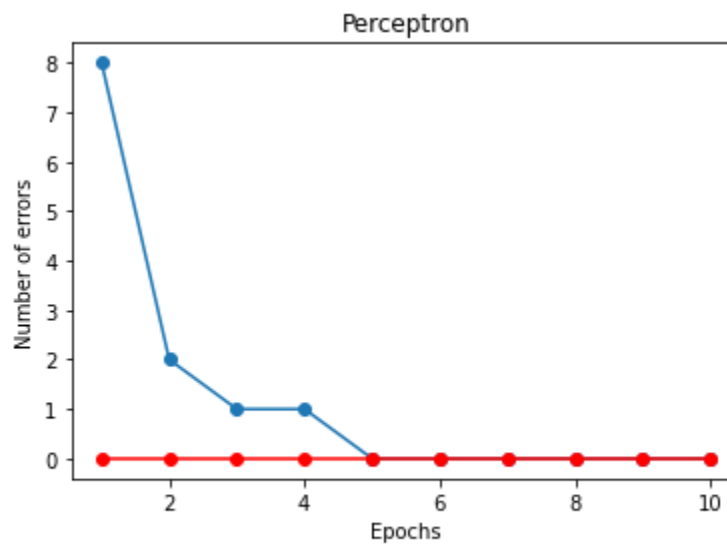
سرعت همگرایی برابر ۲ است زیرا در تکرار دوم همگرا شده اند.



Accuracy is : 0.4945

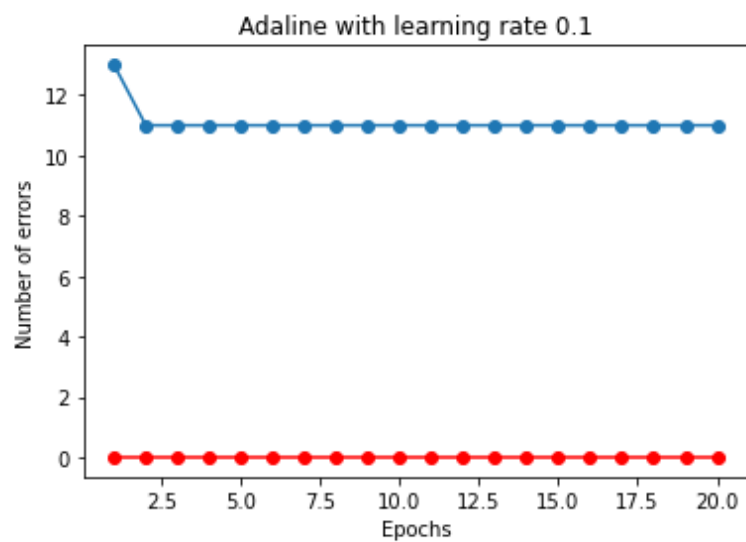
سرعت همگرایی برابر ۲ میباشد.

نرخ یادگیری : ۰/۱



Accuracy is : 1.0

سرعت یادگیری برابر ۵ است زیرا در مرحله ۵ همگرا شده اند.

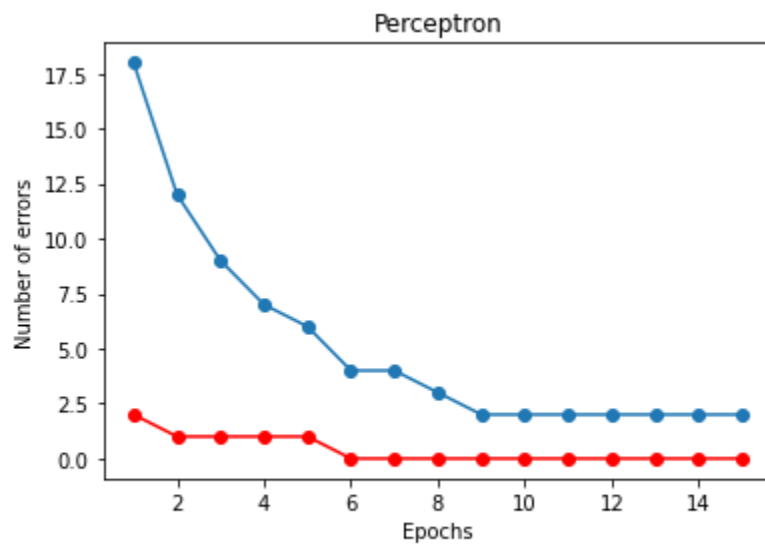


Accuracy is : 1.0

سرعت همگرایی برابر ۲ است.

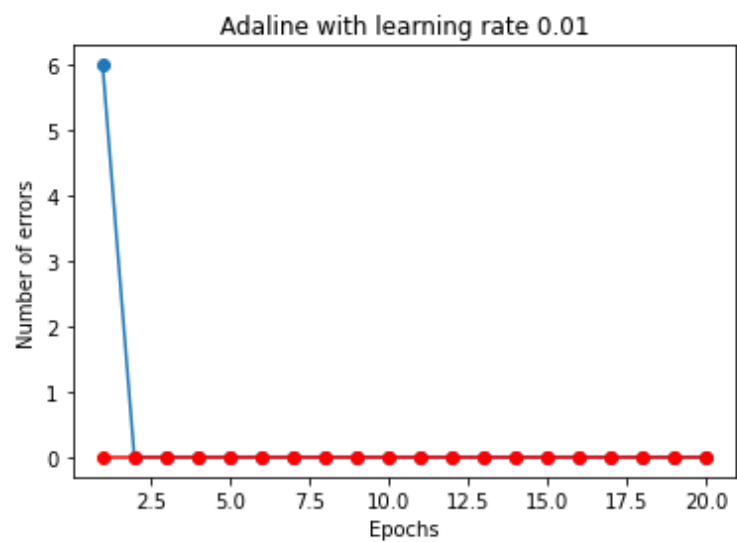


نرخ یادگیری ۰/۰۱



Accuracy is : 1.0

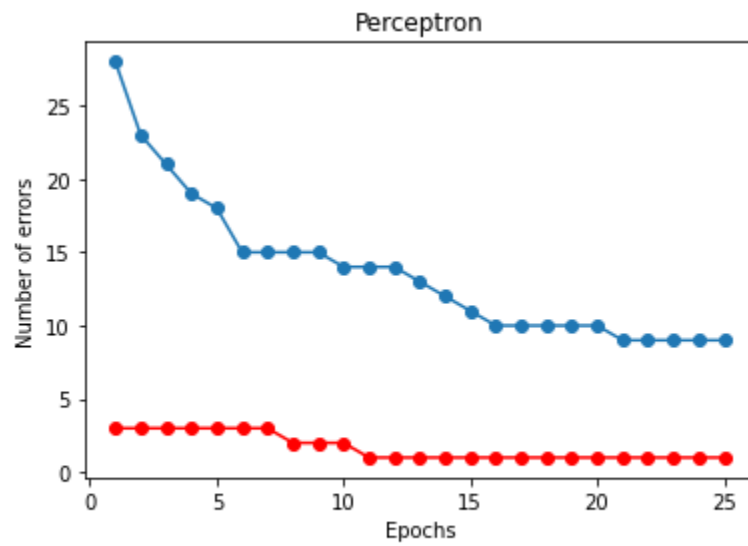
سرعت یادگیری برابر ۹ است.



Accuracy is : 1.0

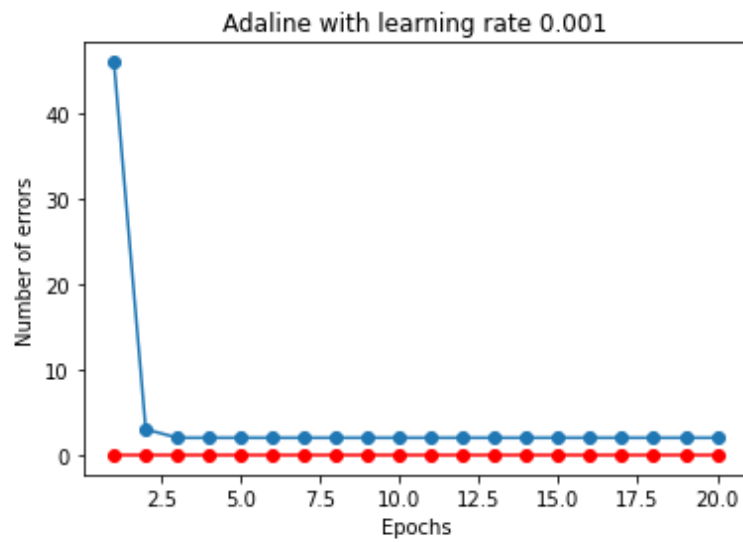
سرعت همگرایی برابر ۲ است.

نرخ یادگیری ۰/۰۰۱



Accuracy is : 0.9995

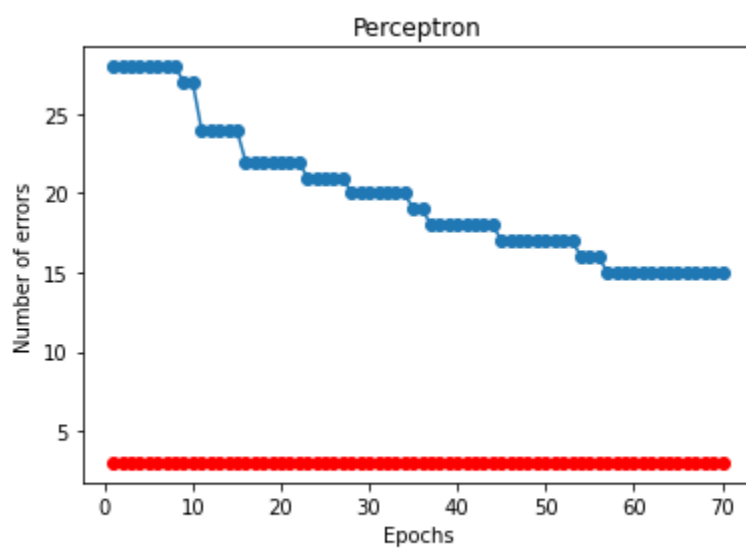
سرعت یادگیری برابر ۲۱ است.



Accuracy is : 1.0

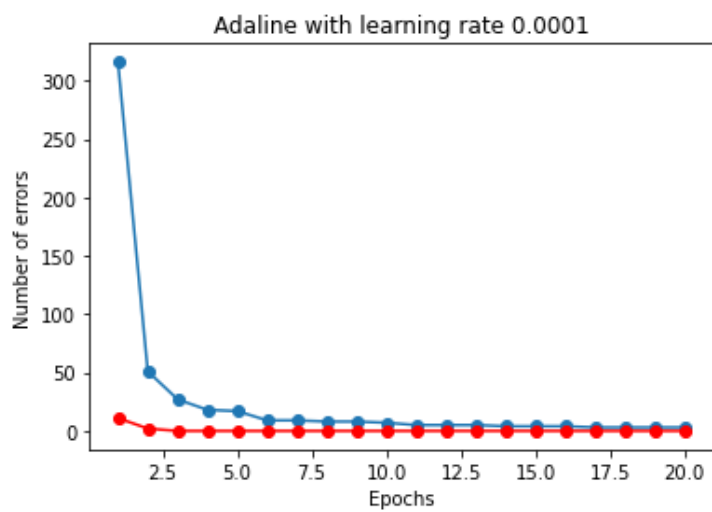
سرعت همگرایی برابر 3 است.

نرخ یادگیری ۰/۰۰۰۱



Accuracy is : 0.9985

سرعت همگرایی برابر ۵۷ است.

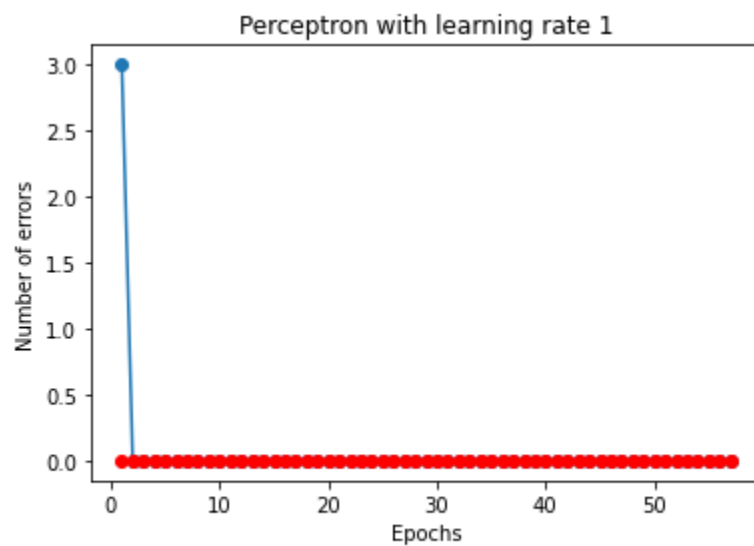


Accuracy is : 1.0

سرعت همگرایی برابر ۲ است.

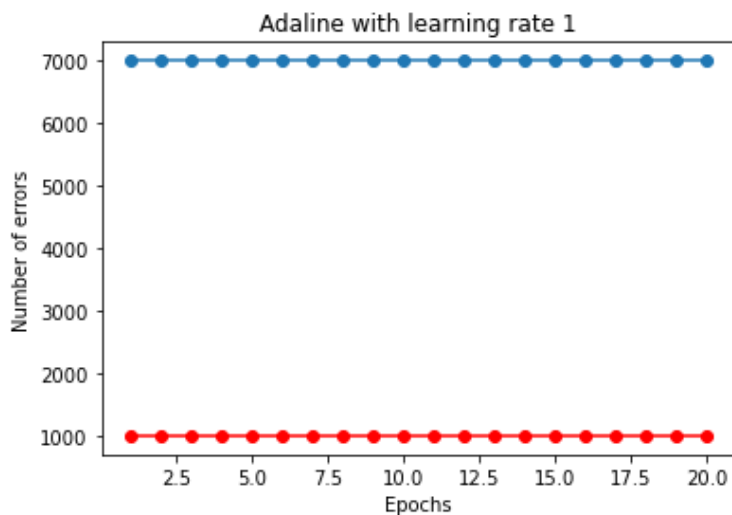
تابع فعالیت sigmoid

نرخ یادگیری ۱



Accuracy is : 1.0

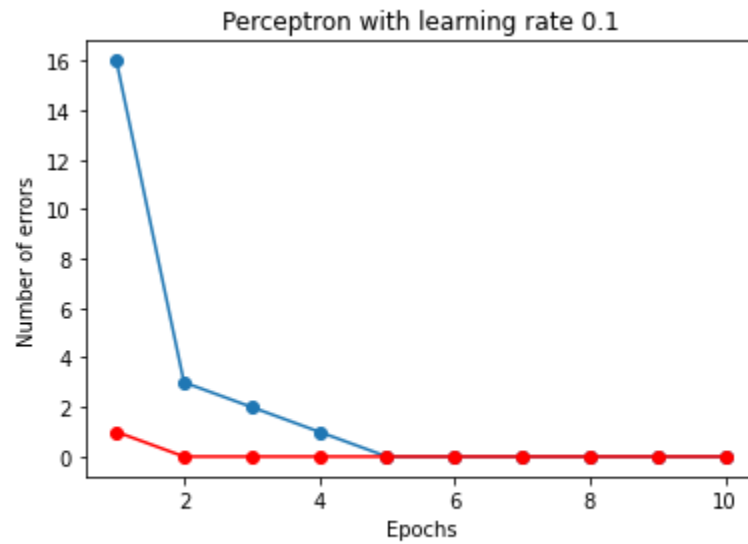
سرعت همگرایی برابر ۲ است.



Accuracy is : 0.0

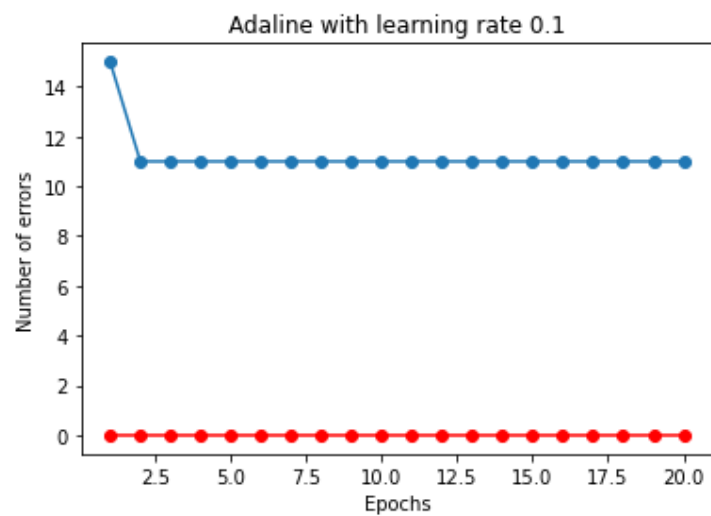
همگرایی ندارد.

نرخ یادگیری ۰/۱



Accuracy is : 1.0

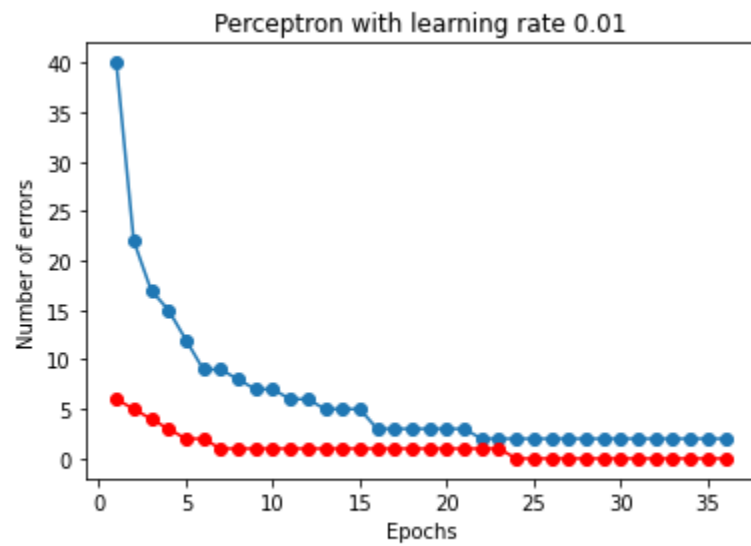
سرعت همگرایی ۵ است.



Accuracy is : 1.0

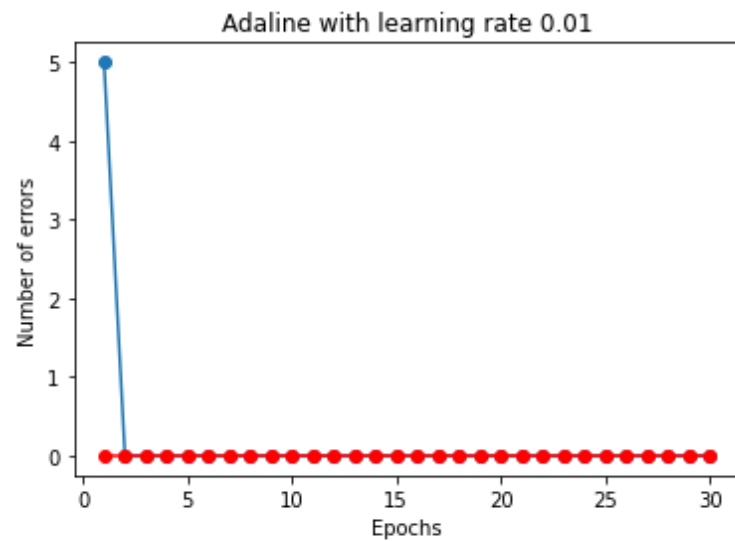
سرعت همگرایی ۲ است.

نرخ یادگیری ۰/۰۱



Accuracy is : 1.0

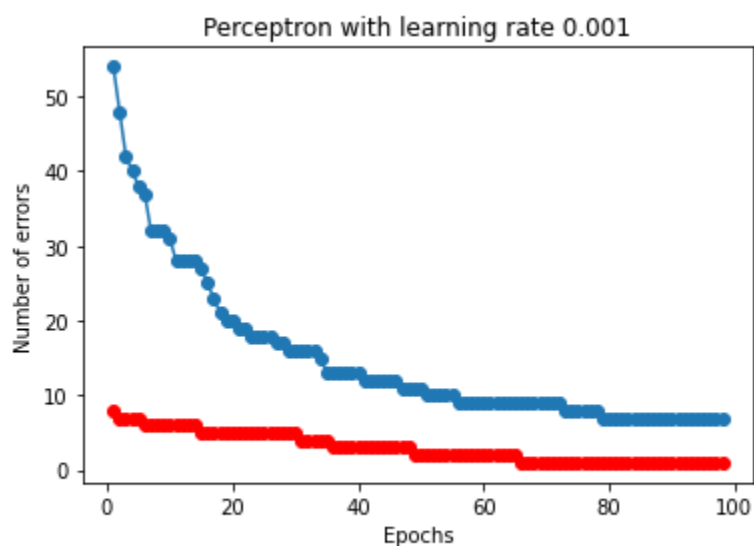
سرعت همگرایی برابر ۲۴ است.



Accuracy is : 1.0

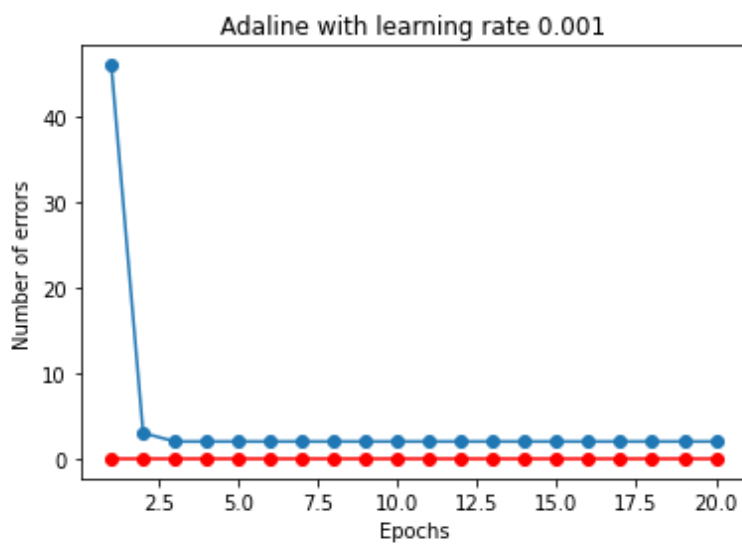
سرعت همگرایی برابر ۲ است.

نرخ یادگیری ۰/۰۰۱



Accuracy is : 0.9995

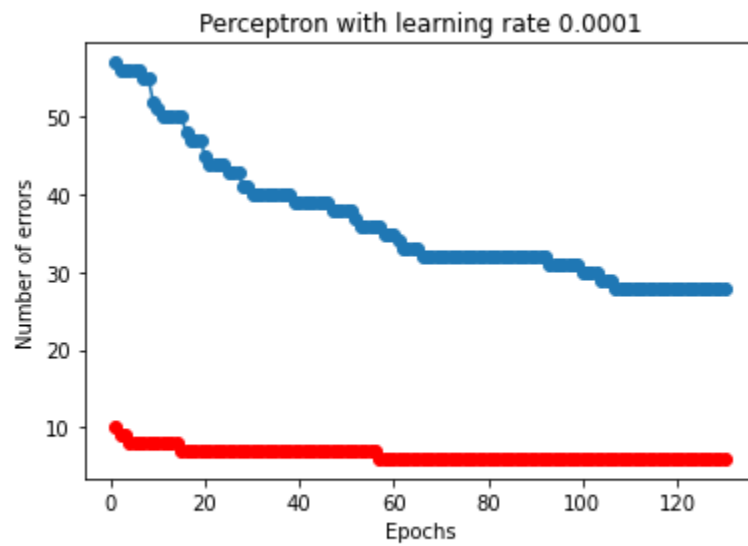
سرعت همگرایی برابر ۷۸ است.



Accuracy is : 1.0

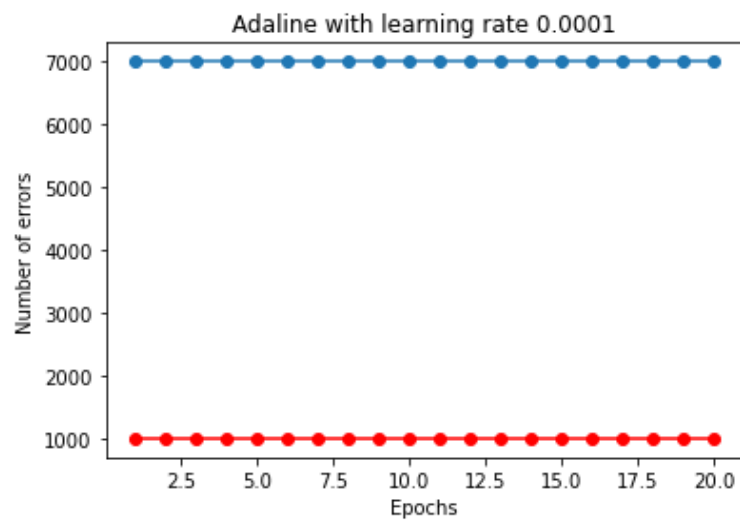
سرعت همگرایی برابر ۳ است.

نرخ یادگیری ۰/۰۰۰۱



Accuracy is : 0.9975

سرعت همگرایی برابر ۱۱۰ است.



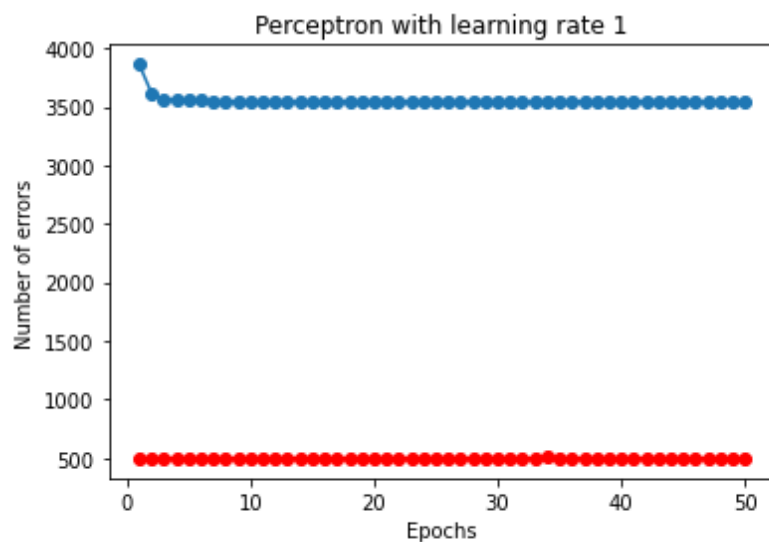
Accuracy is : 0.0

همگرایی ندارد.



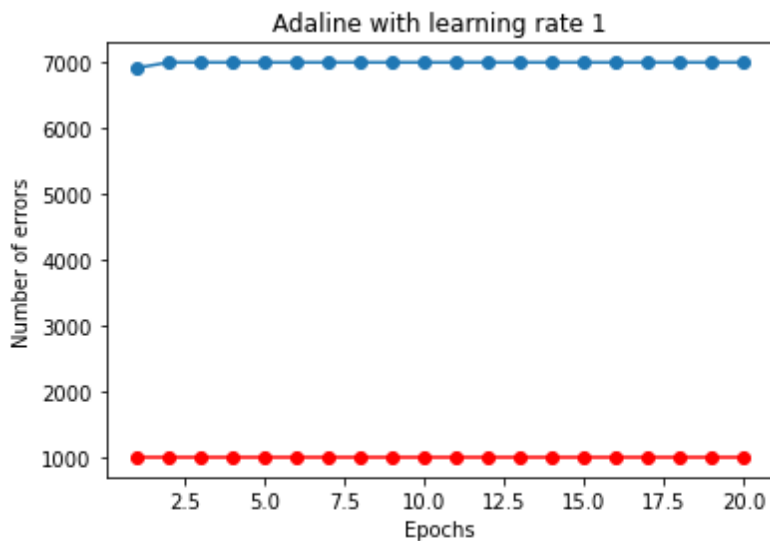
تابع فعالیت tanh :

نرخ یادگیری ۱ :



Accuracy is : 0.5055

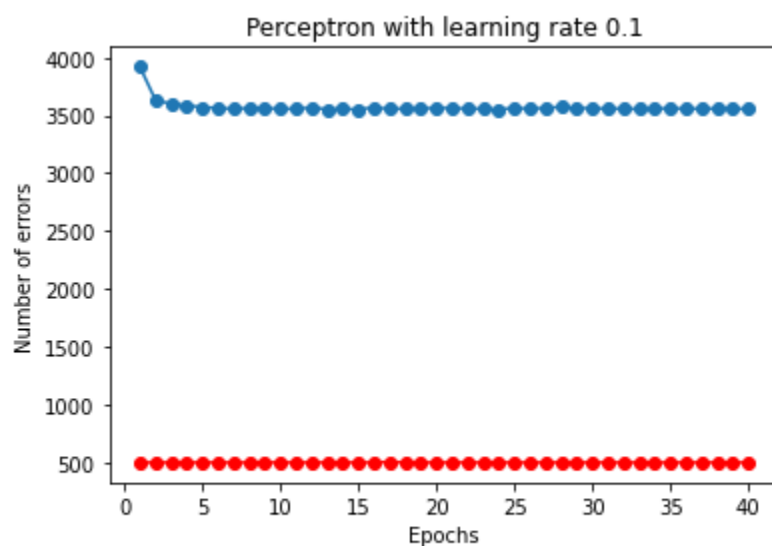
سرعت همگرایی برابر ۲ است.



Accuracy is : 0.0

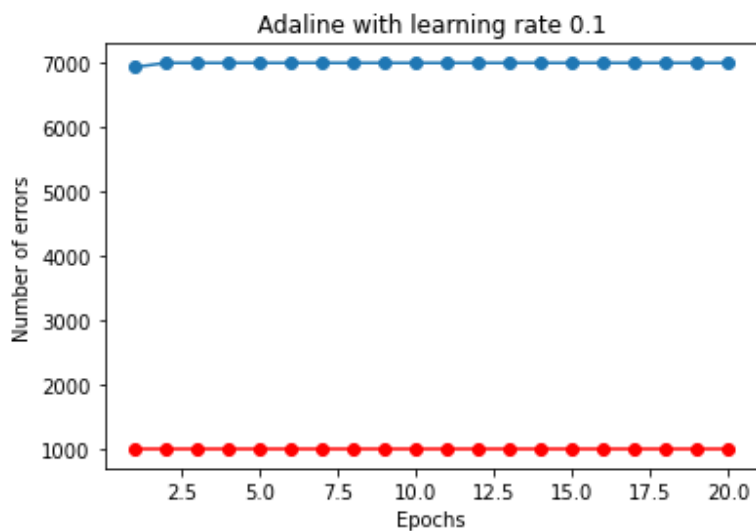
همگرایی ندارد.

نرخ یادگیری ۱٪



Accuracy is : 0.5055

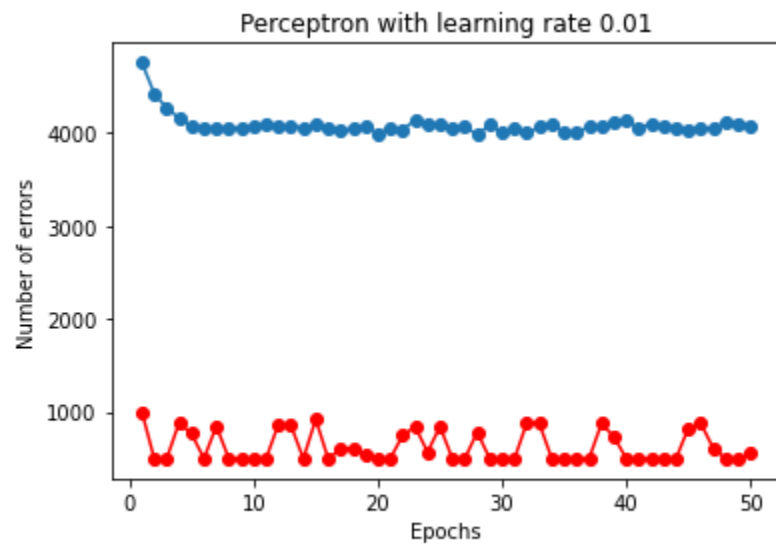
سرعت همگرایی برابر ۲۸ است.



Accuracy is : 0.0

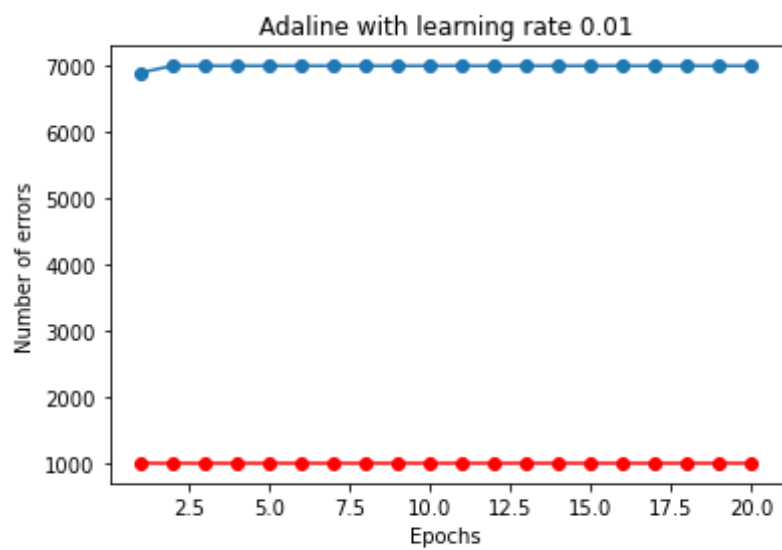
همگرایی ندارد.

نرخ یادگیری ۰/۰۱



Accuracy is : 0.4325

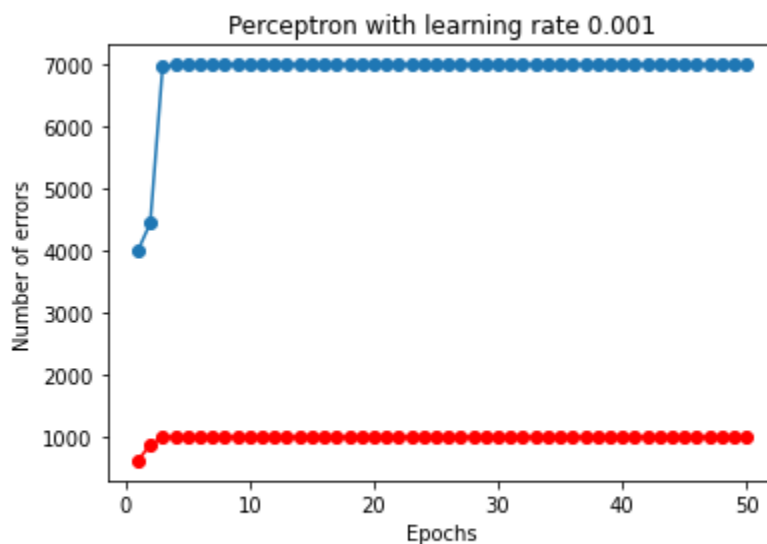
همگرایی ندارد.



Accuracy is : 0.0

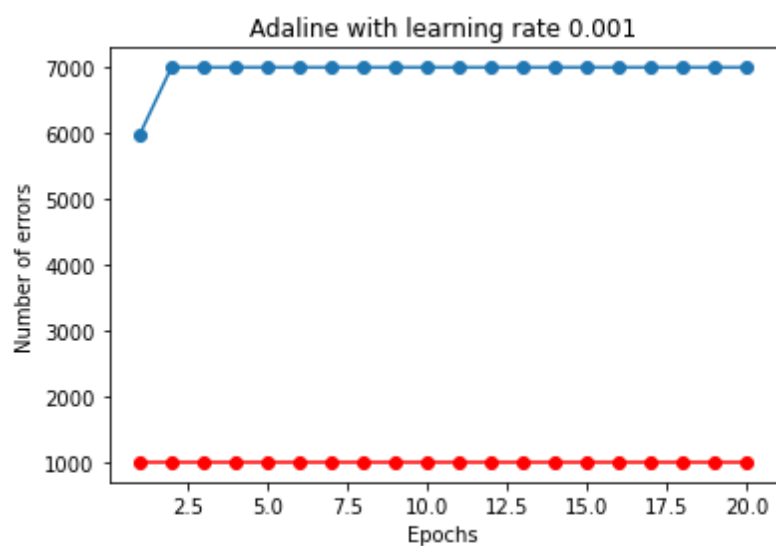
همگرایی ندارد.

نرخ یادگیری ۰/۰۰۱



Accuracy is : 0.0

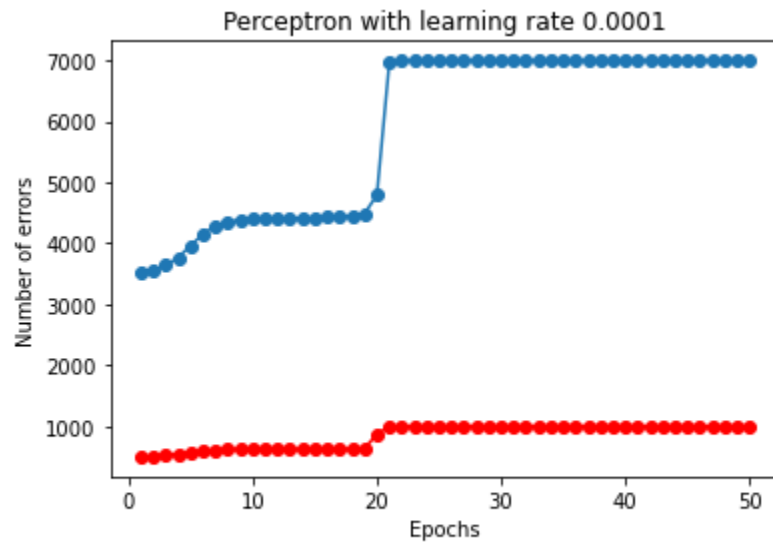
سرعت همگرایی برابر ۳ است.



Accuracy is : 0.0

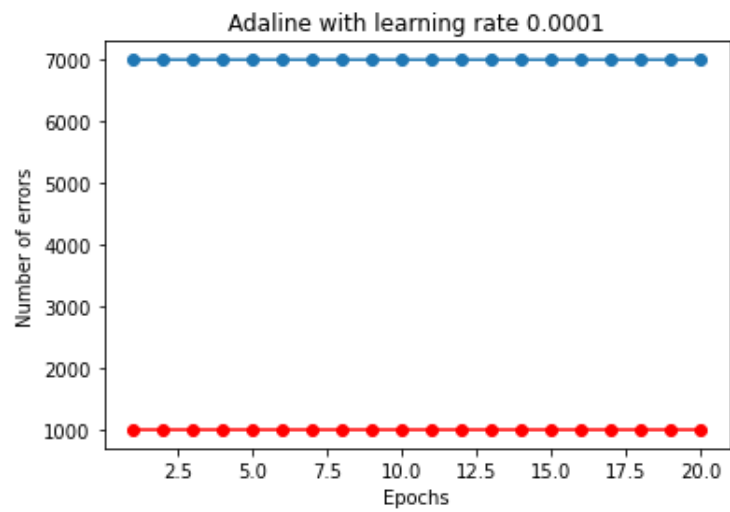
همگرایی ندارد.

نرخ یادگیری ۰/۰۰۰۱



Accuracy is : 0.0

سرعت همگرایی برابر ۲۲ است.



Accuracy is : 0.0

همگرایی برابر ۰ است.

هر دو شبکه میتوانند به راحتی مسئله را حل کنند. در هر دو شبکه تابع فعالیت unit بهترین حالت مسئله را حل میکند. زیر در هر دو شبکه ، در تعداد حالات بیشتری جواب های مناسبی برای مسئله ایجاد کرده است. پس از آن تابع sigmoid است که این تابع نیز پاسخ های مناسبی ایجاد میکند اما تعداد آن کمتر است.

در unit برای نورون پرسپترون بهترین نرخ برابر ۱ است و با کاهش آن ، شبکه بدتر میشود . در sigmoid نیز بهترین حالت برای نرخ ۱ است و با کاهش آن شبکه بدتر میشود. در تابع tanh بهترین حالت آن برای حالت ۱ است و با کاهش آن شبکه بدتر میشود هر چند این تابع برای حل این مسئله برای این نورون اصلا مناسب نیست.

در مورد شبکه ادالاین برای هر دو تابع sigmoid , unit بهترین حالت برابر ۰/۰۱ است و با کاهش و افزایش آن شبکه بدتر میشود. در مورد تابع tanh در هیچ حالت مقدار آن مناسب نیست. در کل تابع tanh به هیچ عنوان برای ادالاین مناسب نیست.

در کل برای حل این مسئله ، تابع unit و شبکه پرسپترونی بهتر است . زیرا در این نورون و تابع بهترین پاسخ ها ایجاد شده است.

(۴)

برای ساخت دیتاست جداناپذیر از ۴ توزیع گوسی استفاده کرده ایم که از ۴ میانگین متفاوت و ماتریس کواریانس متفاوت ساخته شده اند. از این ۴ توزیع برای ساخت یک دیتاست به صورت xor استفاده شده است. مقادیر استفاده شده در این مثال به صورت زیر هستند :

داده های اول:

$$\text{Mean} = [ -50 , 150 ]$$

$$\text{Cov} = [ [100 , 20 ] , [ 20 , 200 ] ]$$

داده های دوم :

$$\text{Mean} = [ 50 , 150 ]$$

$$\text{Cov} = [ [100 , 20 ] , [ 20 , 200 ] ]$$

داده های سوم:

$$\text{Mean} = [ 50 , 0 ]$$

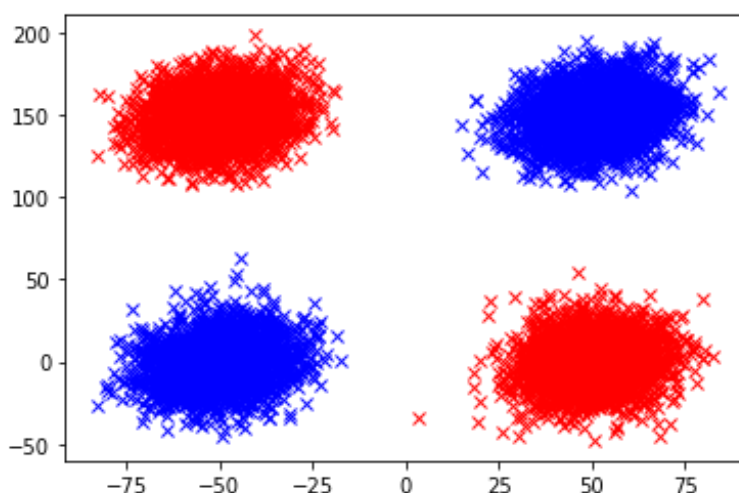
$$\text{Cov} = [ [100 , 20 ] , [ 20 , 200 ] ]$$

داده های چهارم:

$$\text{Mean} = [ -50 , 0 ]$$

$$\text{Cov} = [ [100 , 20 ] , [ 20 , 200 ] ]$$

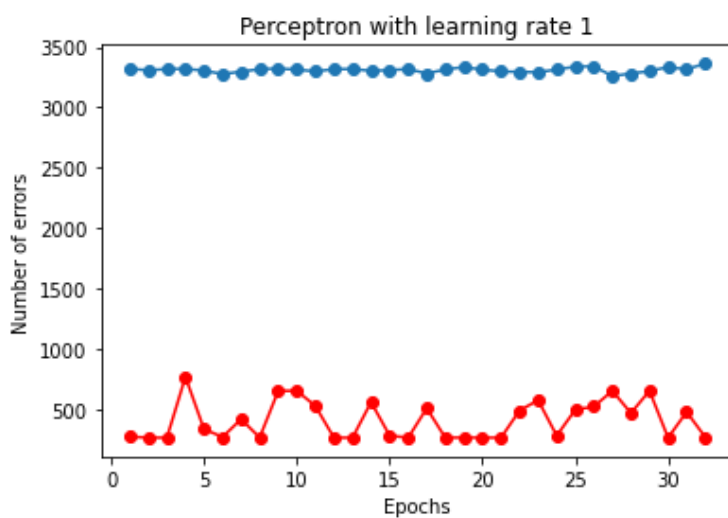
نمودار داده های ایجاد شده به صورت زیر است:



(۵)

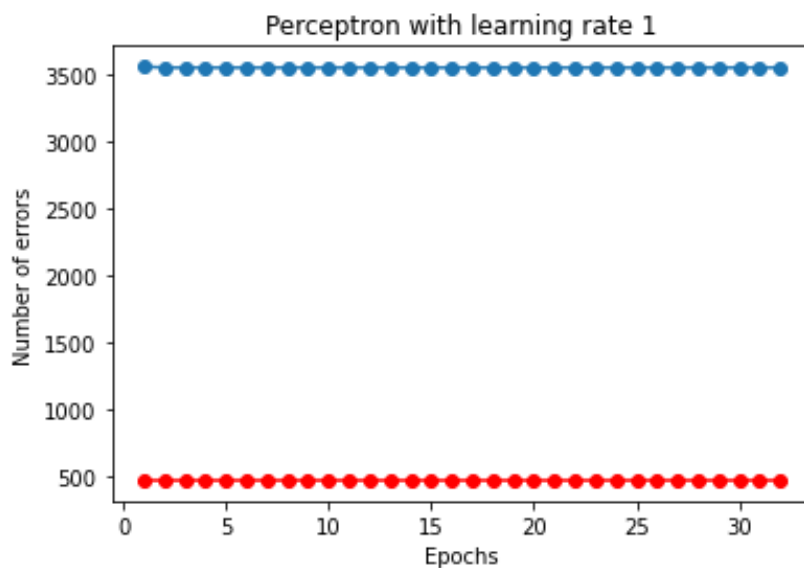
تابع فعالیت unit :

نرخ یادگیری : ۱



Accuracy is : 0.731

همگرایی ندارد.

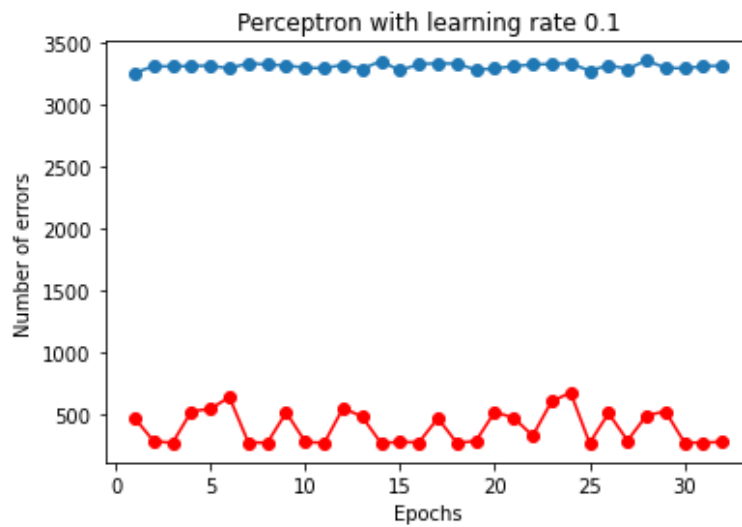


Accuracy is : 0.509

همگرایی ندارد .

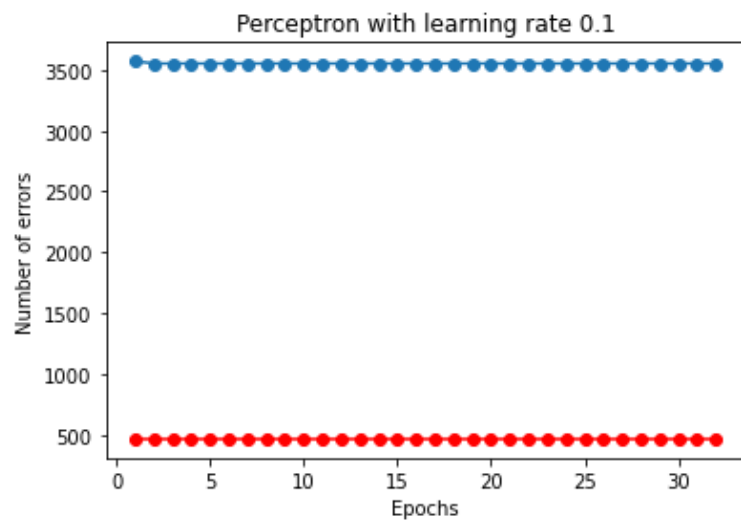


نرخ یادگیری : ۰/۱



Accuracy is : 0.73

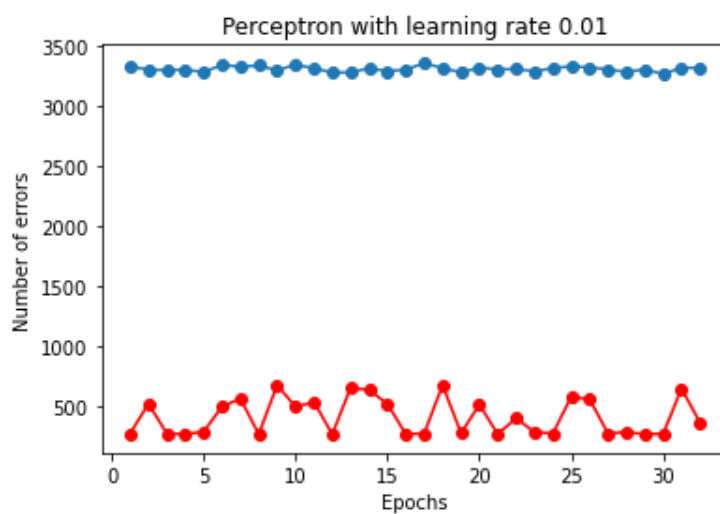
همگرایی ندارد.



Accuracy is : 0.509

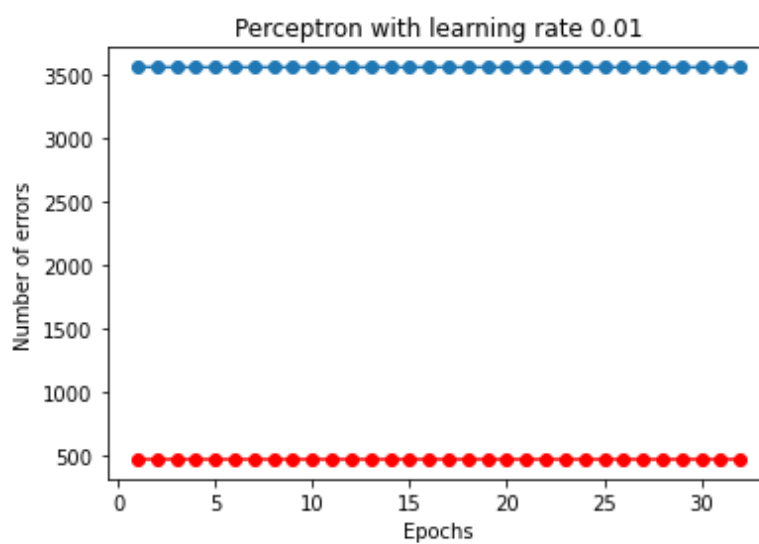
همگرایی ندارد .

نرخ یادگیری ۰/۰۱



Accuracy is : 0.6205

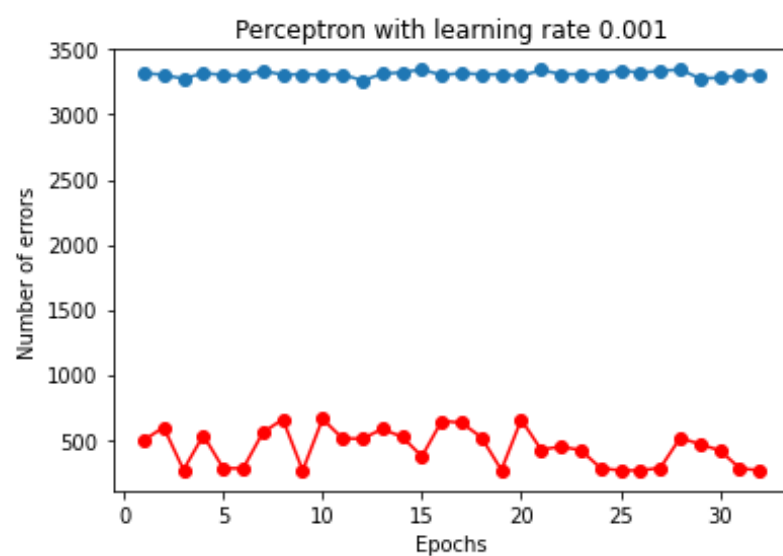
همگرایی ندارد .



Accuracy is : 0.509

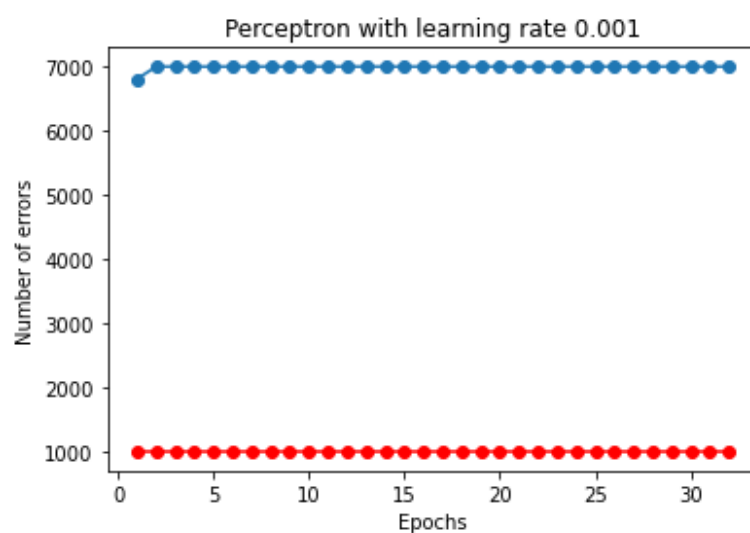
همگرایی ندارد .

نرخ یادگیری ۰/۰۰۱



Accuracy is : 0.74

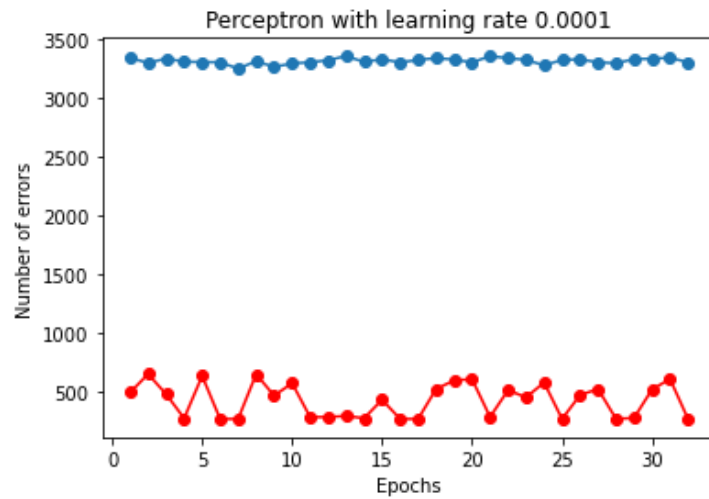
همگرایی ندارد.



Accuracy is : 0.0

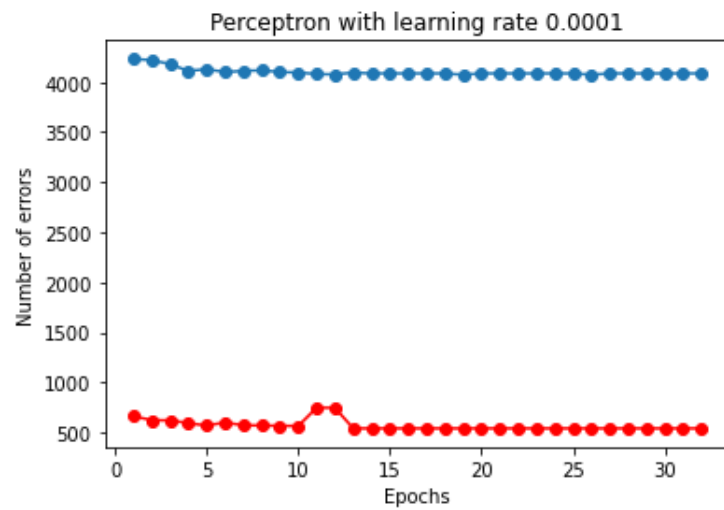
همگرایی ندارد .

نرخ یادگیری ۰/۰۰۰۱



Accuracy is : 0.7395

همگرایی ندارد.

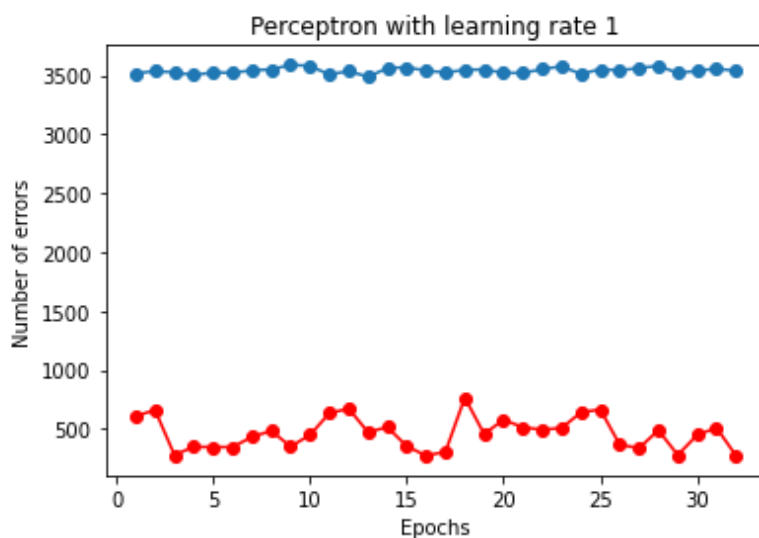


Accuracy is : 0.485

همگرایی ندارد .

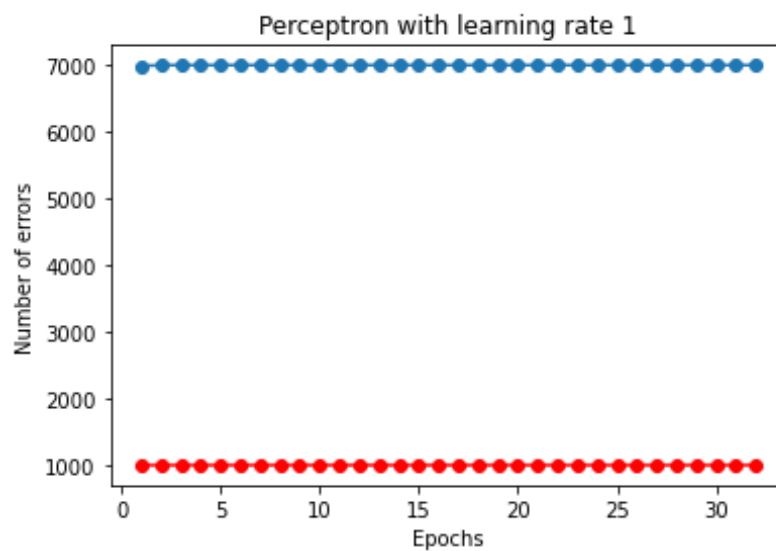
تابع فعالیت sigmoid

نرخ یادگیری ۱



Accuracy is : 0.7265

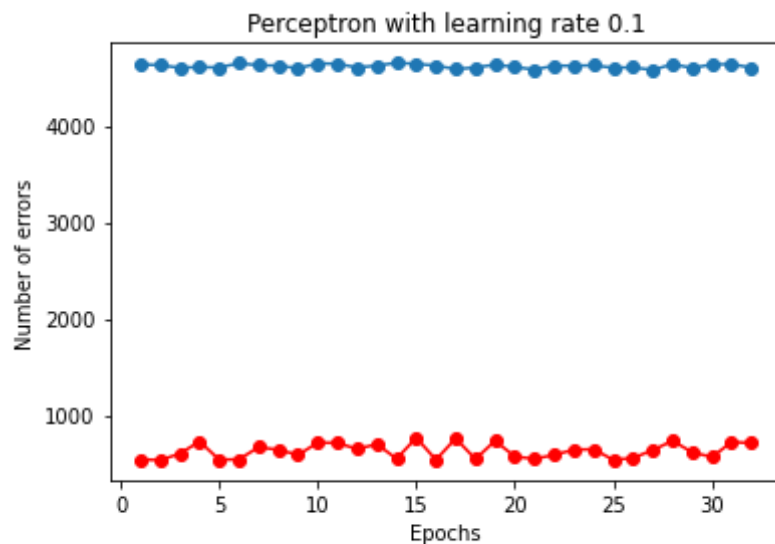
همگرایی ندارد.



Accuracy is : 0.0

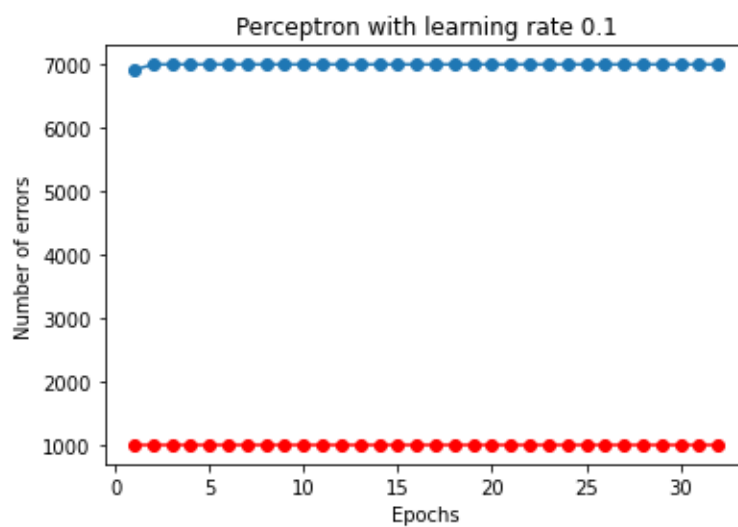
همگرایی ندارد .

نرخ یادگیری ۱٪



Accuracy is : 0.2935

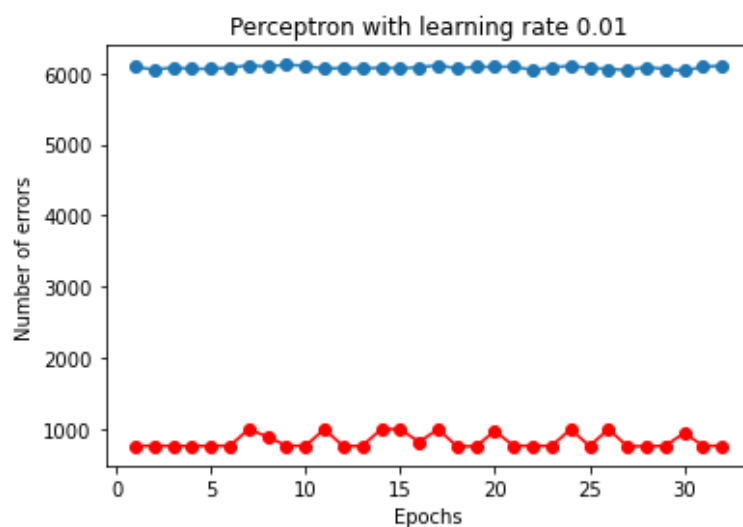
همگرایی ندارد .



Accuracy is : 0.0

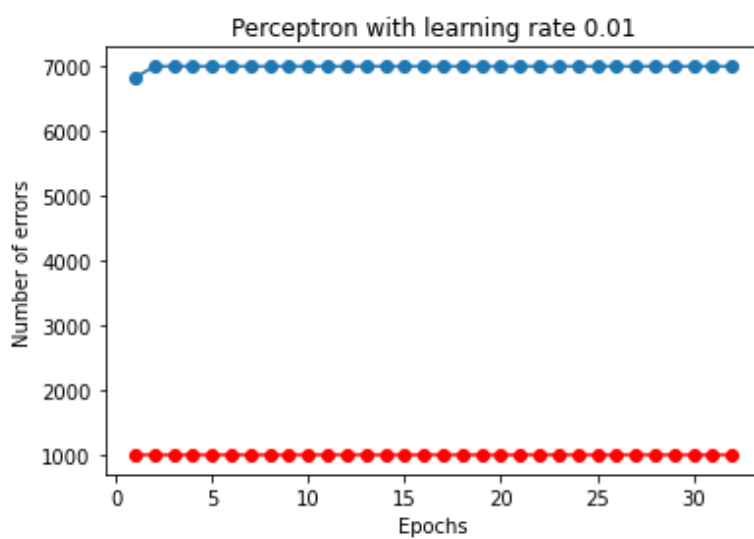
همگرایی ندارد .

نرخ یادگیری ۰/۰۱



Accuracy is : 0.2355

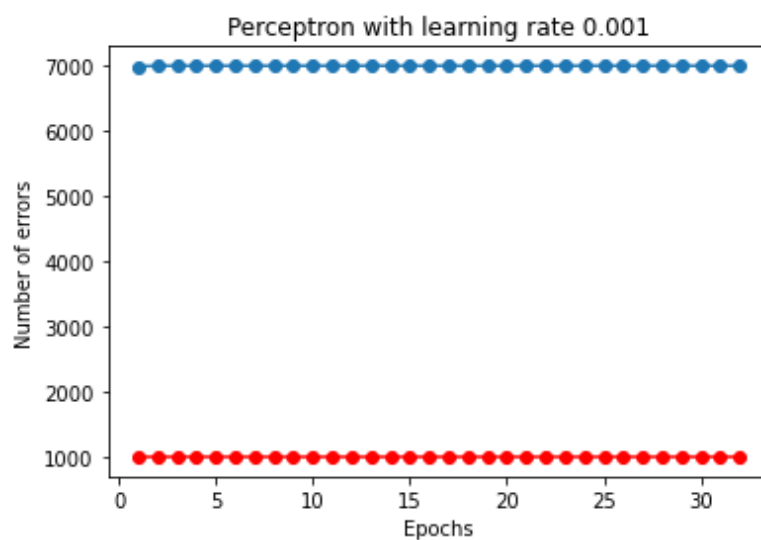
همگرایی ندارد.



Accuracy is : 0.0

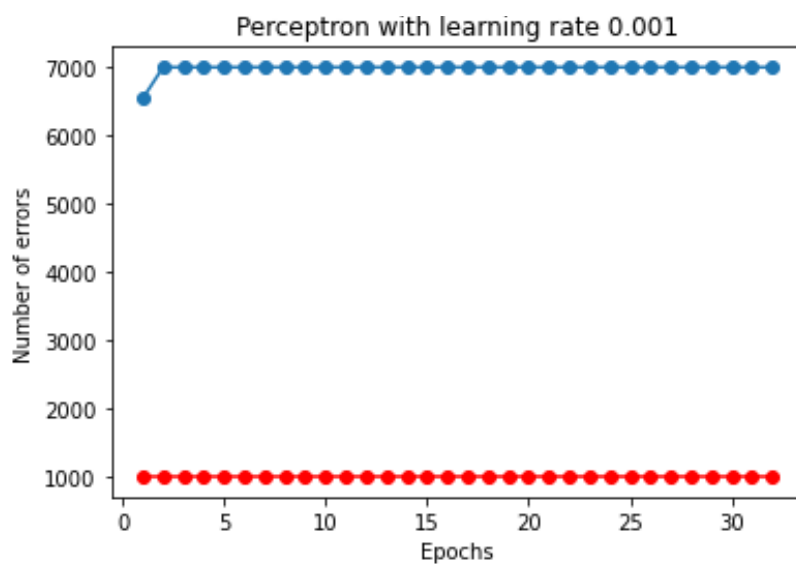
همگرایی ندارد .

نرخ یادگیری ۰/۰۰۱



Accuracy is : 0.0

همگرایی ندارد.

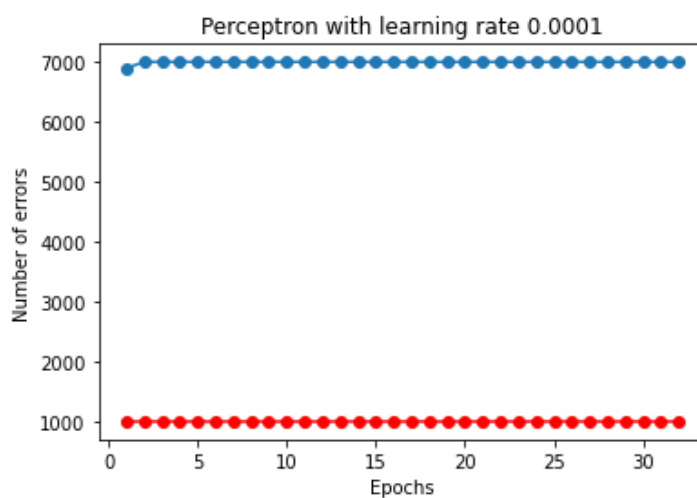


Accuracy is : 0.0

همگرایی ندارد .

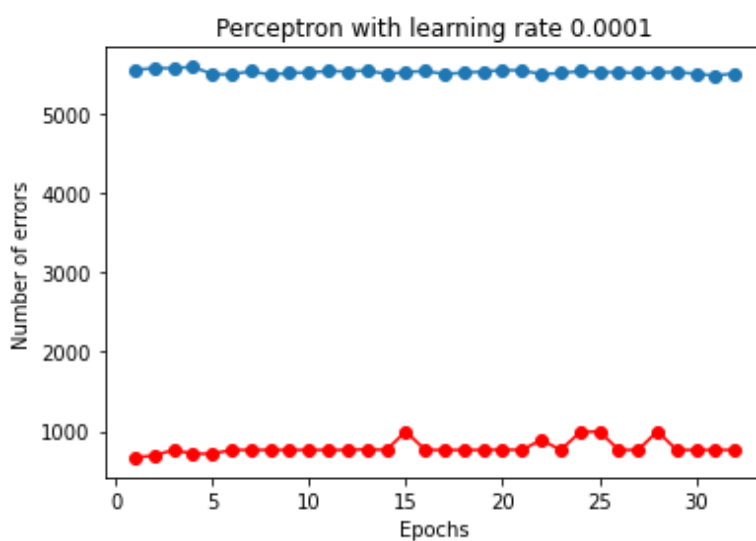


نرخ یادگیری ۰/۰۰۰۱



Accuracy is : 0.0

همگرایی ندارد .

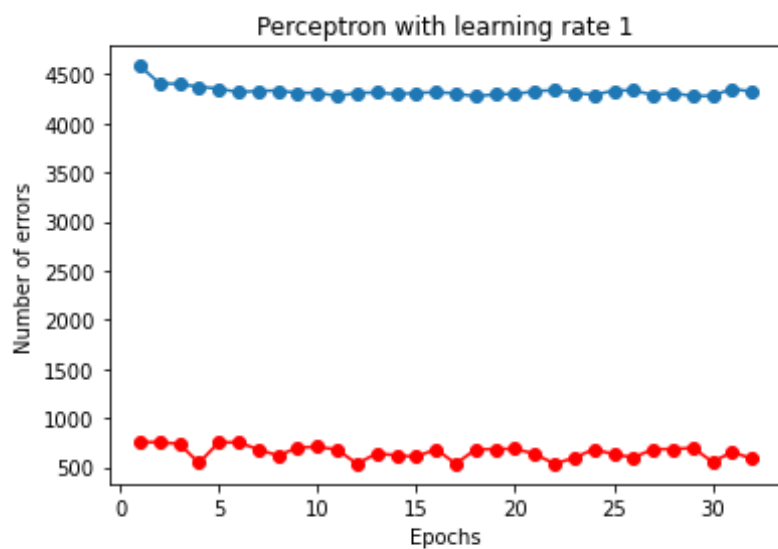


Accuracy is : 0.235

همگرایی ندارد .

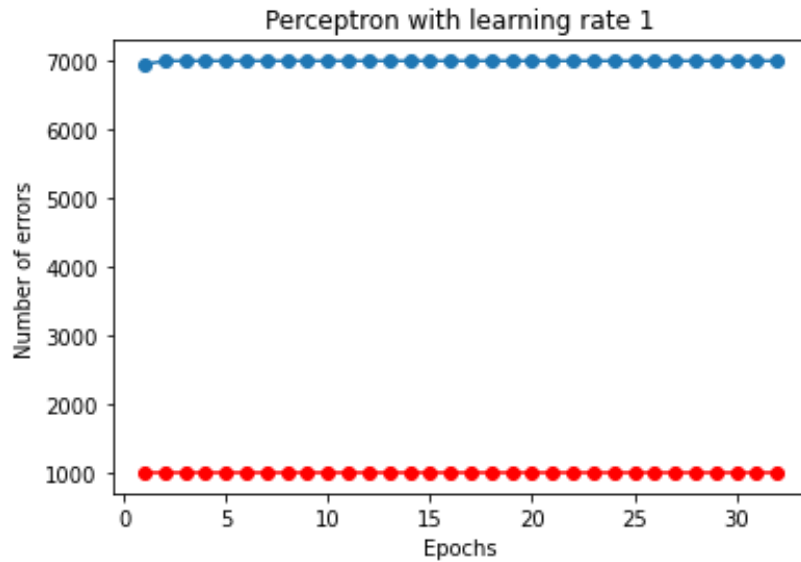
تابع فعالیت tanh :

نرخ یادگیری ۱ :



Accuracy is : 0.429

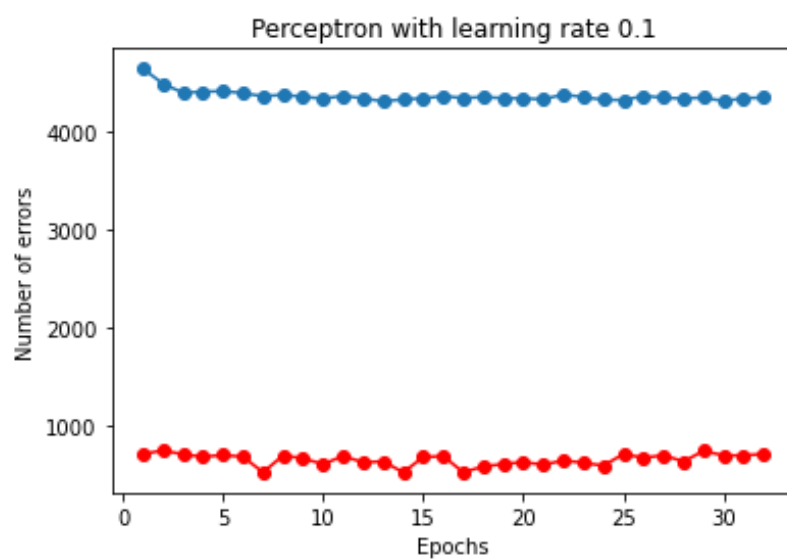
همگرایی ندارد .



Accuracy is : 0.0

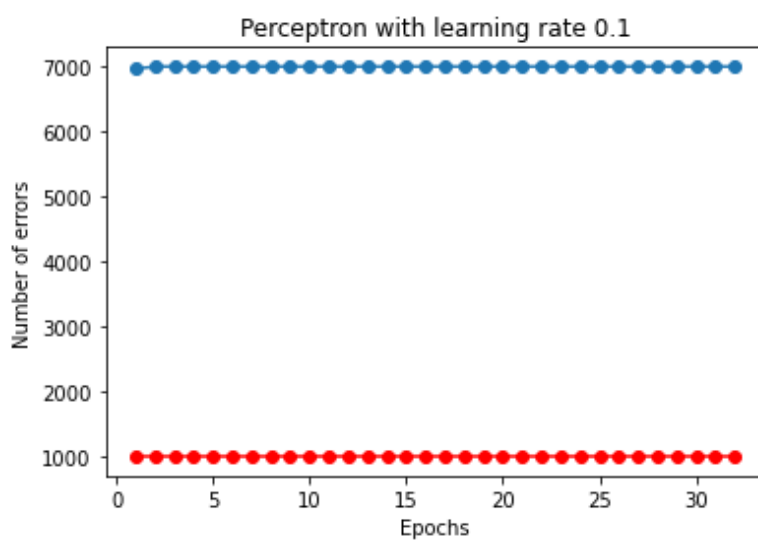
همگرایی ندارد.

نرخ یادگیری ۱٪



Accuracy is : 0.2895

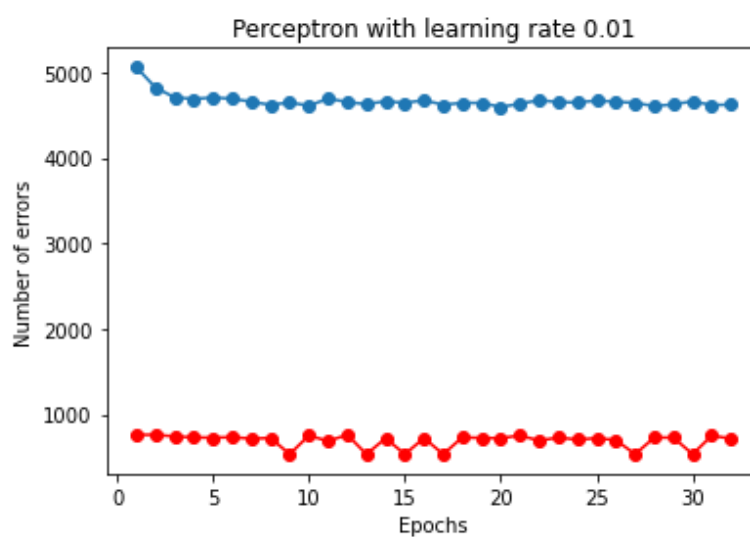
همگرایی ندارد .



Accuracy is : 0.0

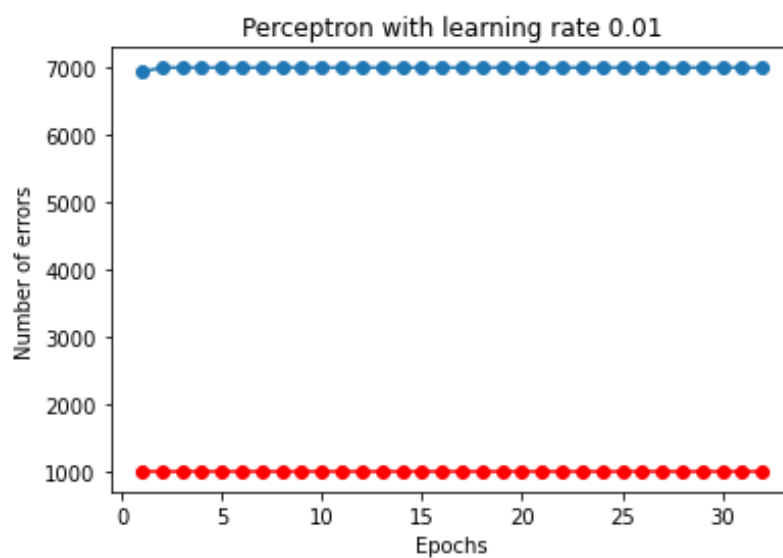
همگرایی ندارد.

نرخ یادگیری ۰/۰۱



Accuracy is : 0.296

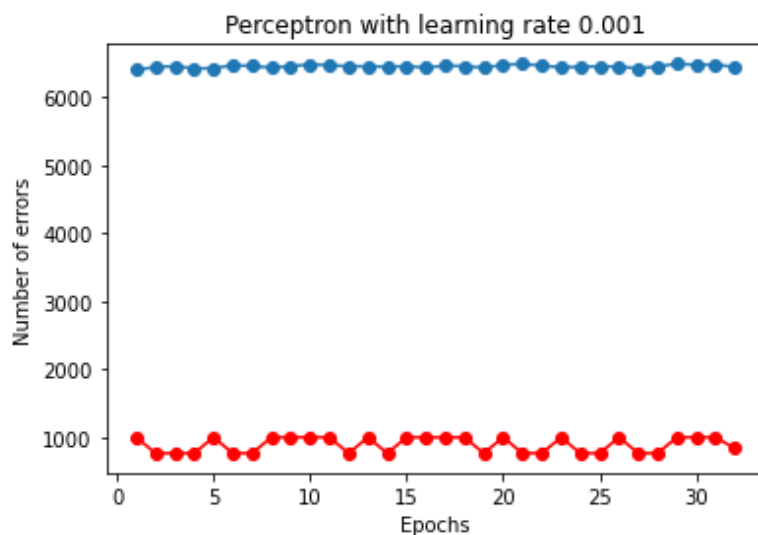
همگرایی ندارد.



Accuracy is : 0.0

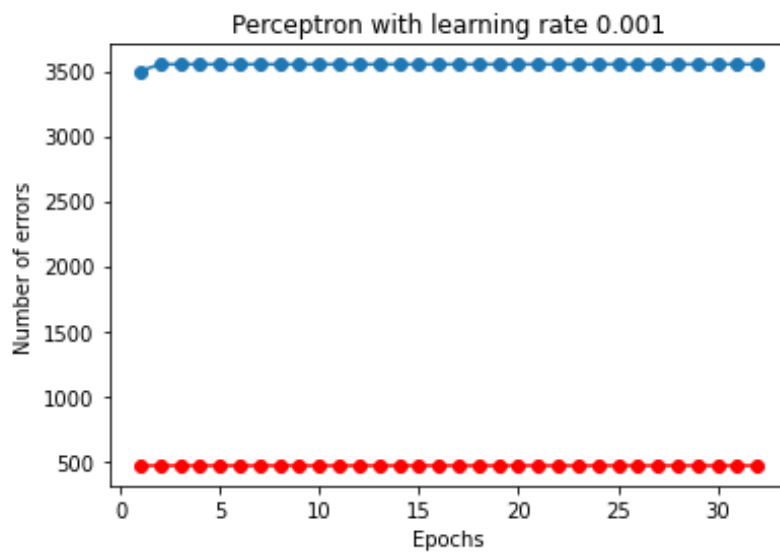
همگرایی ندارد.

نرخ یادگیری ۰/۰۰۱



Accuracy is : 0.1455

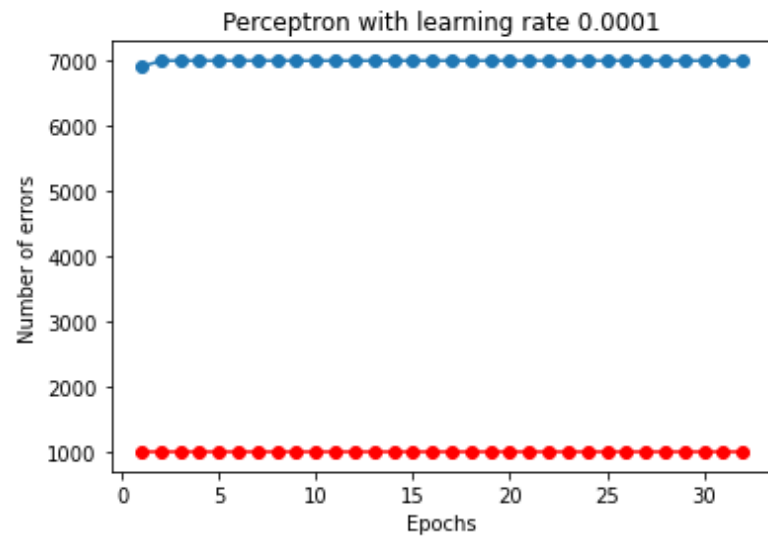
همگرایی ندارد .



Accuracy is : 0.509

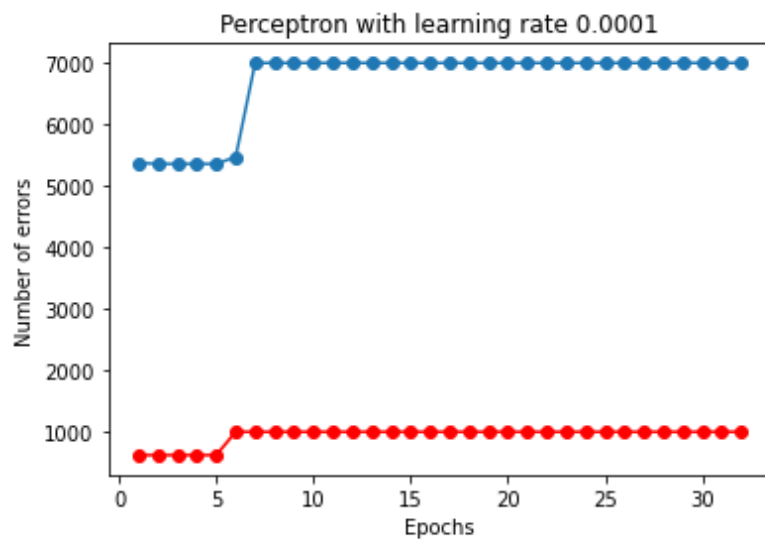
همگرایی ندارد.

نرخ یادگیری ۰/۰۰۰۱



Accuracy is : 0.0

همگرایی ندارد .



Accuracy is : 0.0

همگرایی ندارد .

همانطوری که انتظار میرود ، سیستم های خطی به خصوص ادالاین نمیتوانند داده های XOR را به طور درست کلاس بندی کنند. در شبکه ادالاین در بیشتر حالت ها با تابع فعالیت  $\tanh$  مقدار دقت برابر ۰ است و نمیتواند به درستی فعالیت داشته باشد که نشان از این است که تابع فعالیت تانژانت برای ادالاین به هیچ عنوان مناسب نیست. در تابع فعالیت های  $\text{unit}$  ,  $\text{sigmoid}$  در حالت با نرخ ۱ بهترین خروجی را دارد. دقت آن بیشترین حالت است هرچند شبکه به همگرایی نمیرسد. با کاهش نرخ مقدار دقت کاهش میابد.

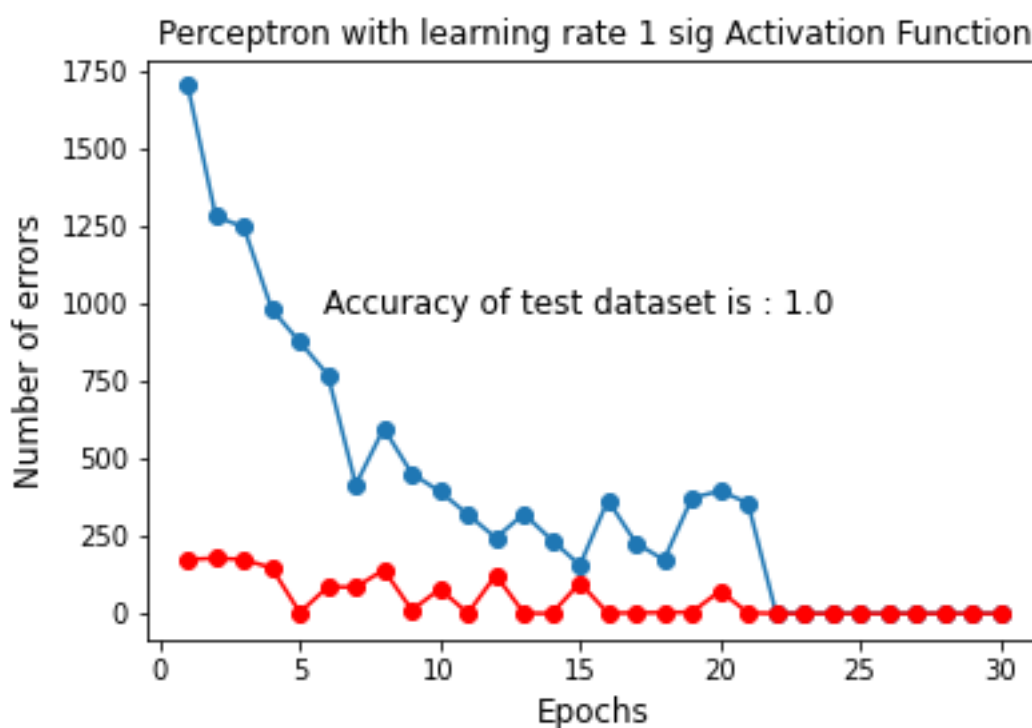
نورون پرسپترون با درصد های بهتر و در حالتی که تابع فعالیت آن  $\text{unit}$  است بهتر از شبکه ادالاین کار میکند هر چند هر دو شبکه نمیتواند به درستی کلاس بندی را انجام دهد.

از نظر بررسی تابع های فعالیت ، تابع فعالیت  $\text{unit}$  در هر دو شبکه به خوبی میتواند مسئله را حل کند و بهتر است. پس از آن تابع  $\text{sigmoid}$  است که در بعضی حالت ها پاسخ های خوبی برای مسئله ایجاد میکند. و در انتها تابع  $\tanh$  قرار دارد که در شبکه ادالاین به طور قطعی میتوان گفت برای این شبکه مناسب نیست. در مورد شبکه پرسپترونی نیز نسب به ۲ تابع دیگر دقت کمتری ایجاد میکند.

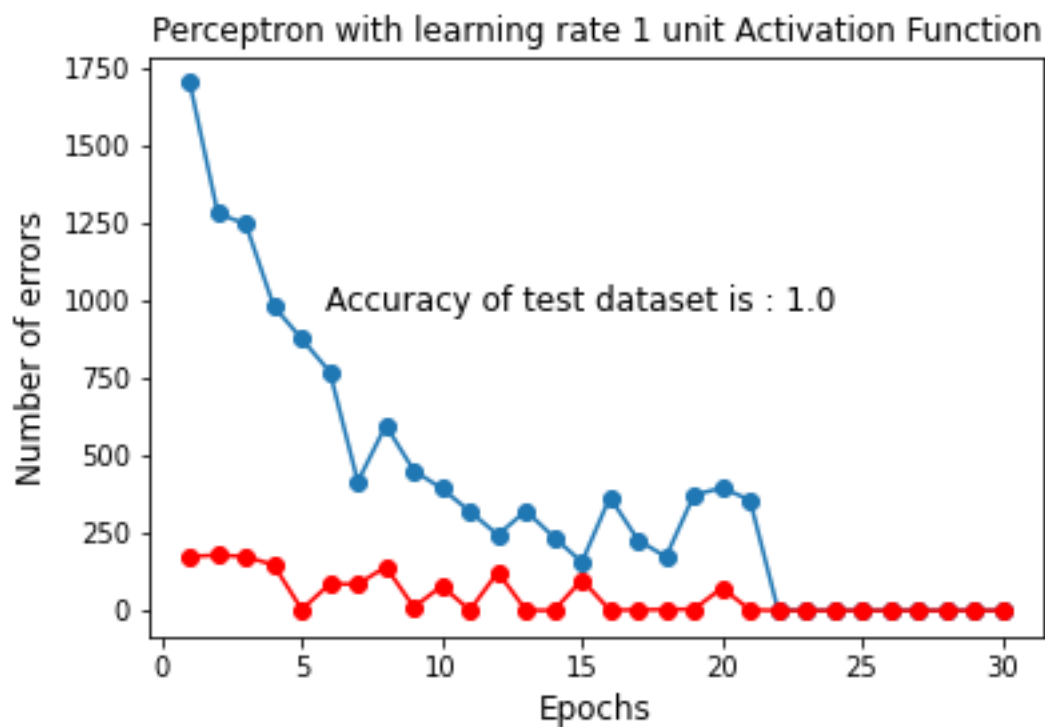
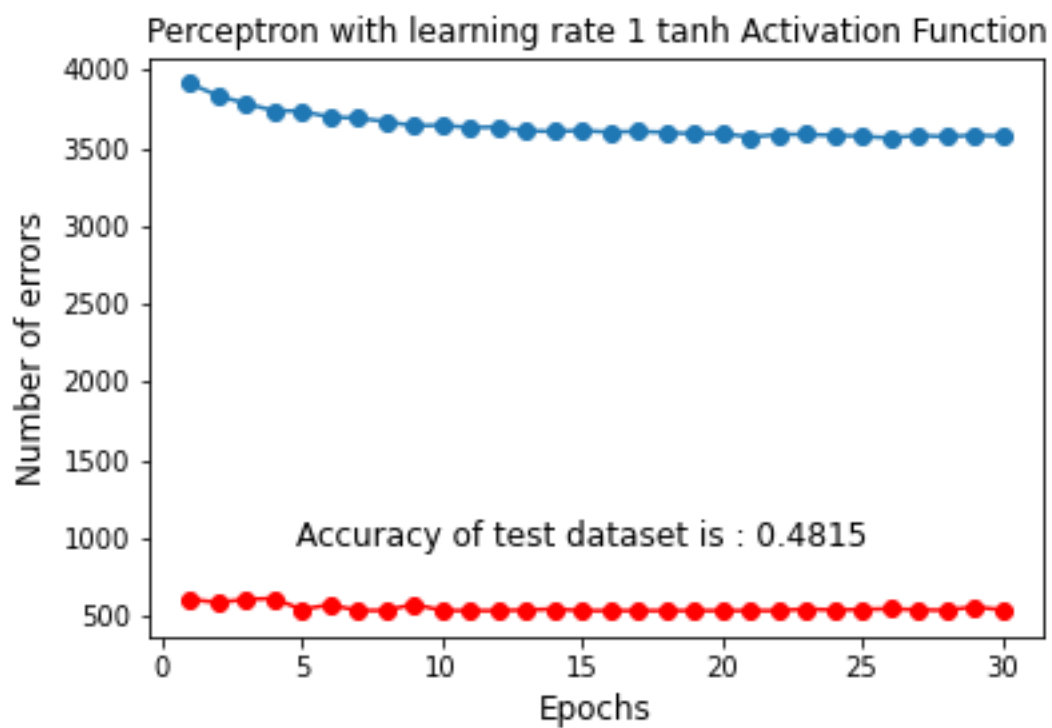
برای این کار ورودی های خود را به ورودی های غیر خطی تبدیل کرده ایم. در این هنگام ورودی ها به صورت  $X$  ,  $Y$  ,  $X^2$  ,  $Y^2$  ,  $XY$  پرسپترون و ادالاینی خود داده ایم. به این صورت سیستم درجه ۲ ایجاد کرده ایم. کد کلاس پرسپترون و ادالاین به گونه دینامیک نوشته شده است که بر اساس تعداد ورودی ها ، تعداد وزن های مورد نیاز برای آن نیز ایجاد شود.

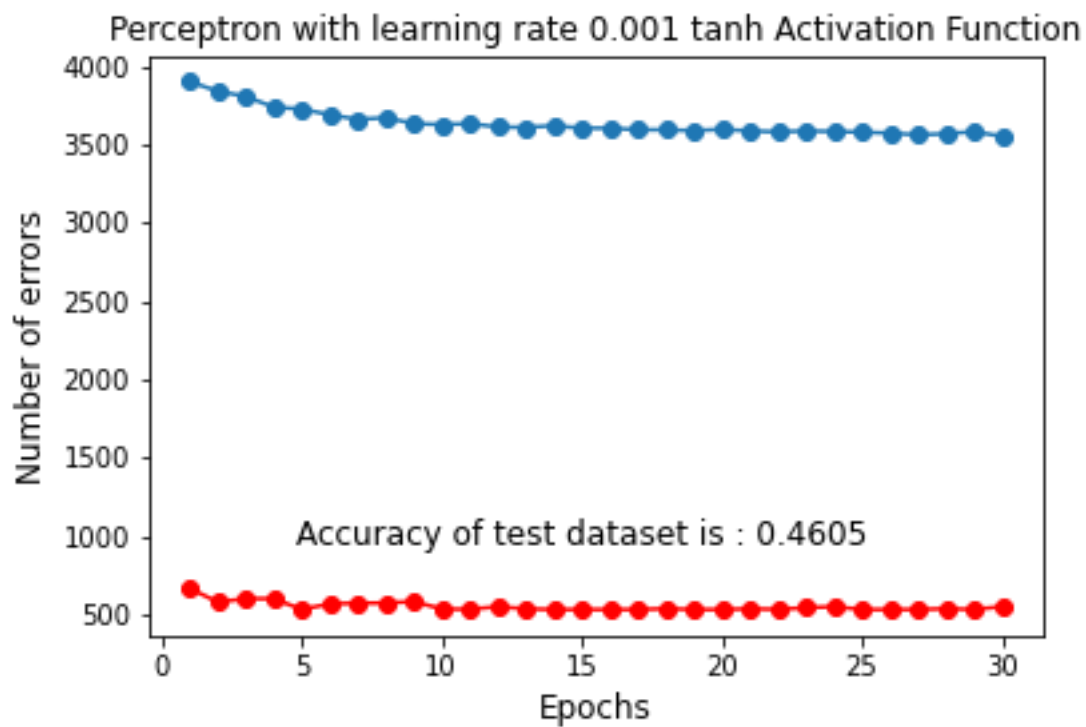
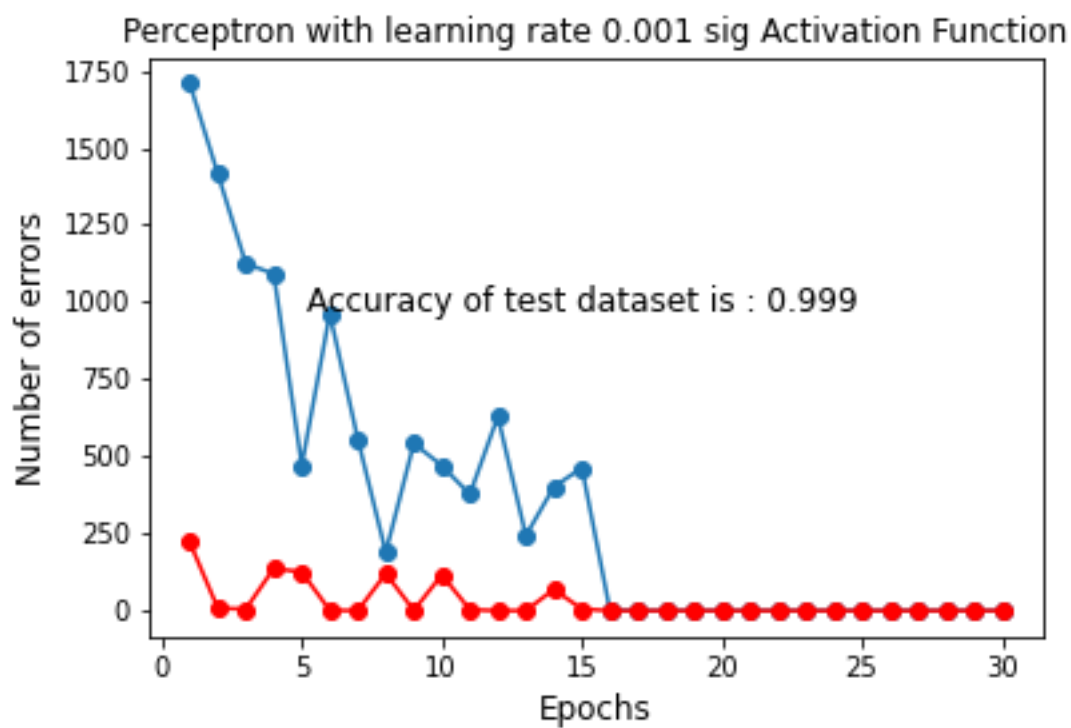
نتایج آزمایشات به صورت زیر است:

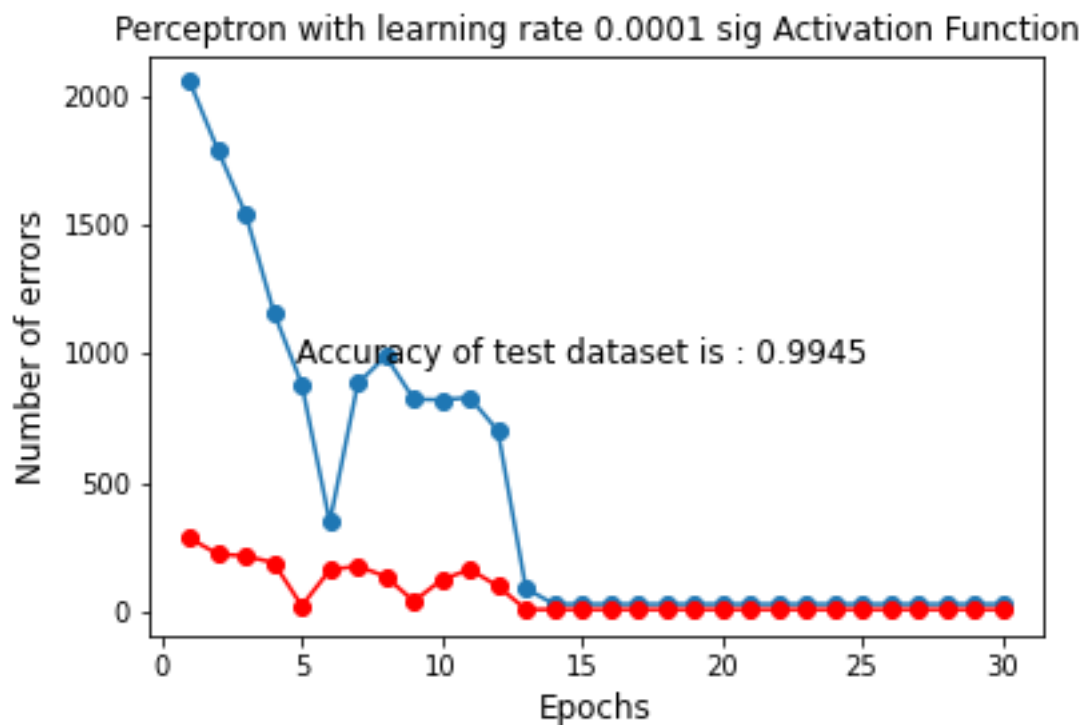
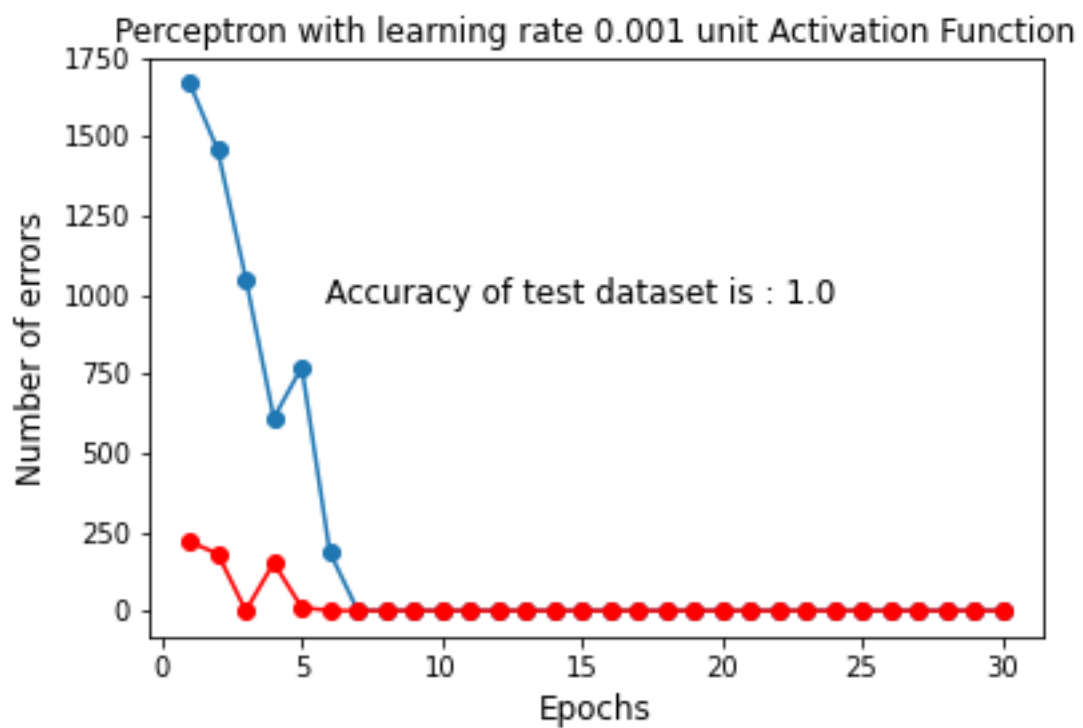
سیستم پرسپترونی :

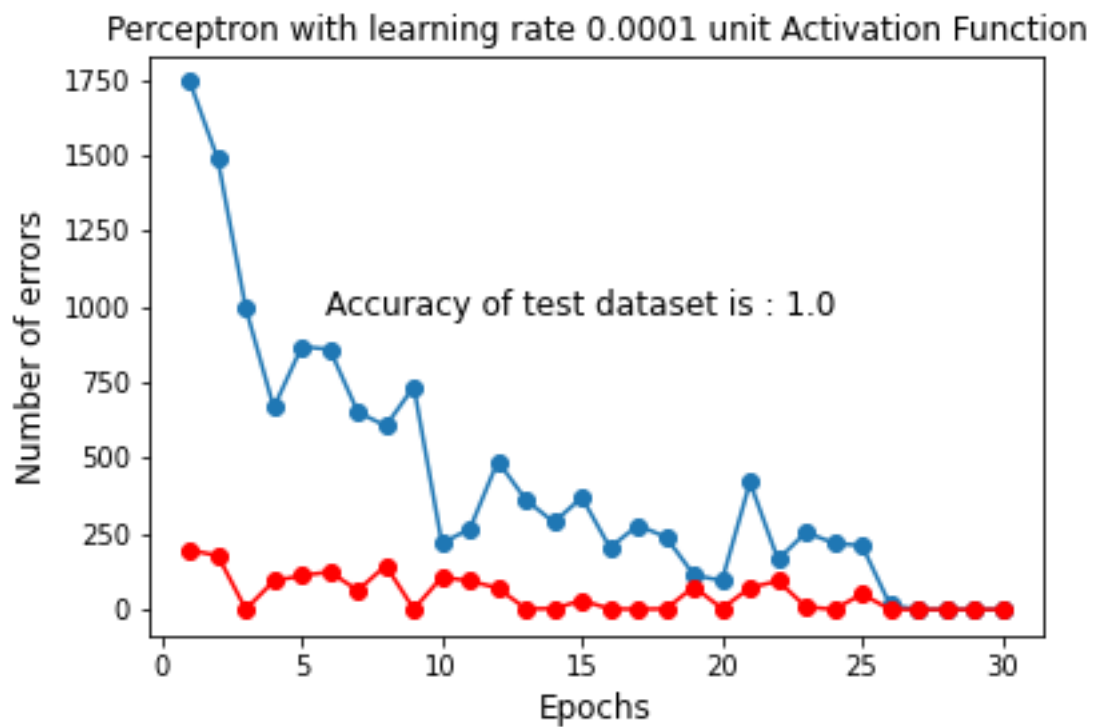
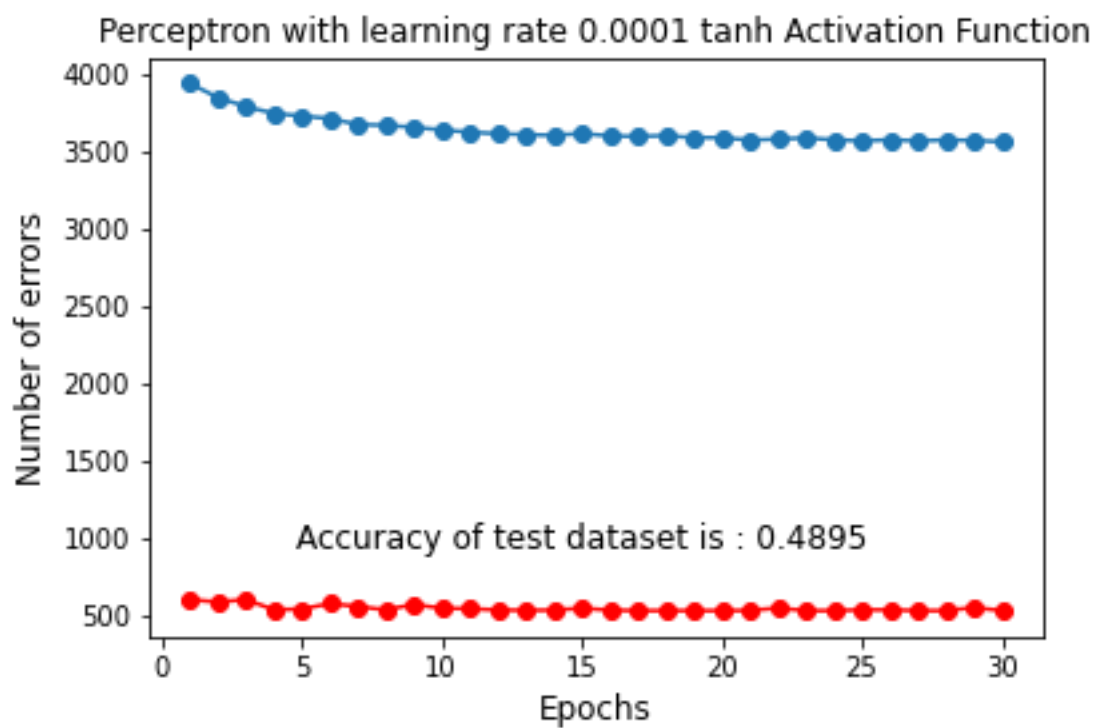




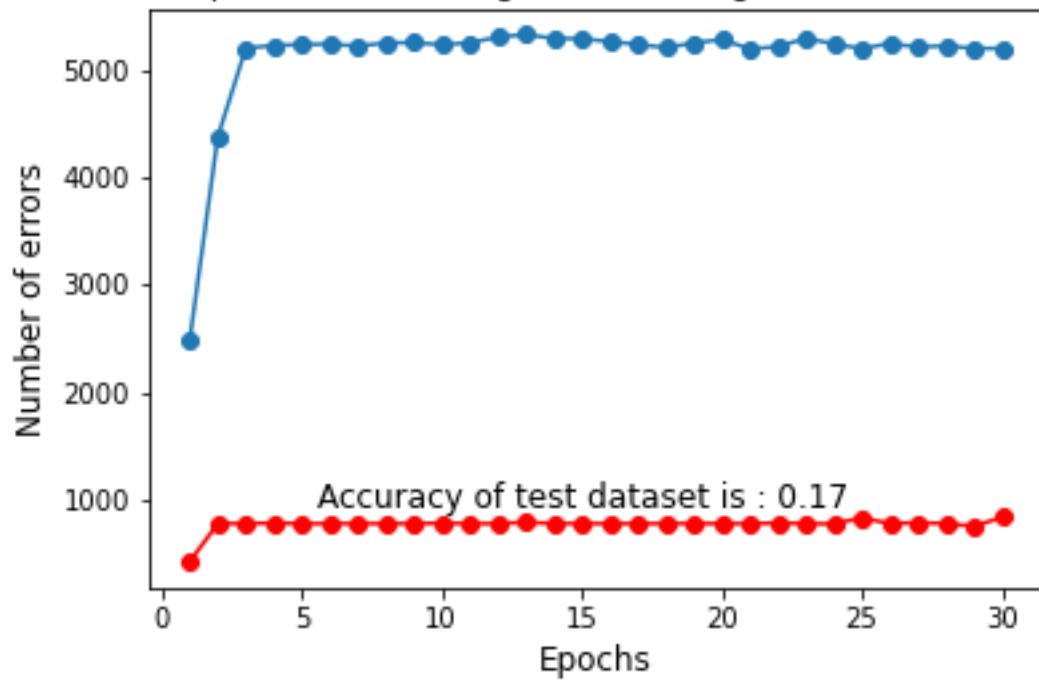




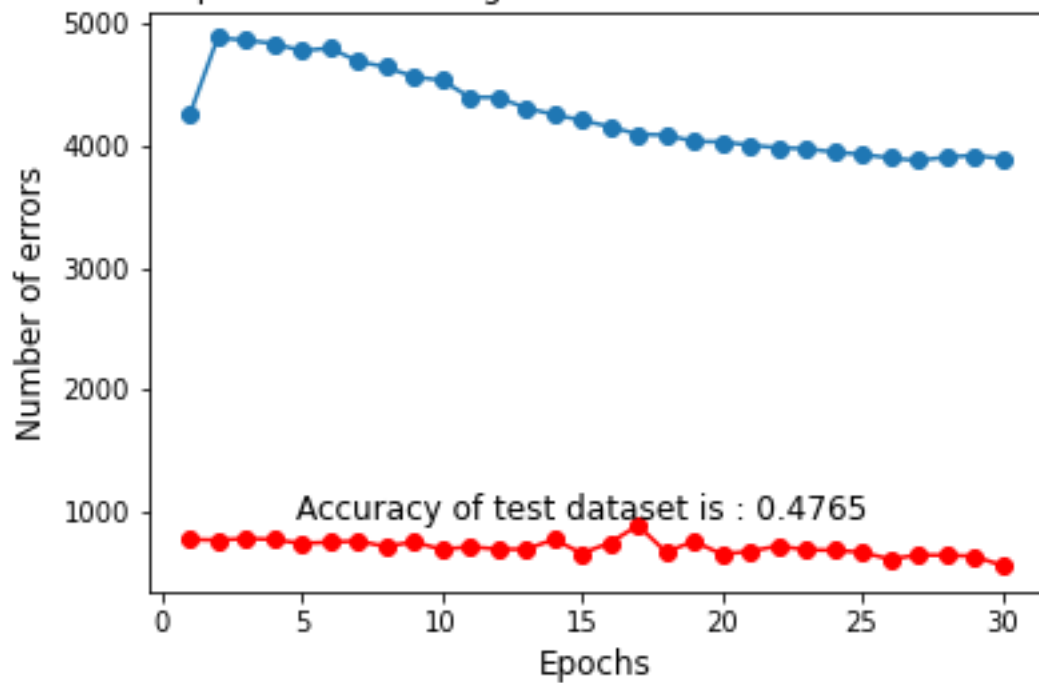




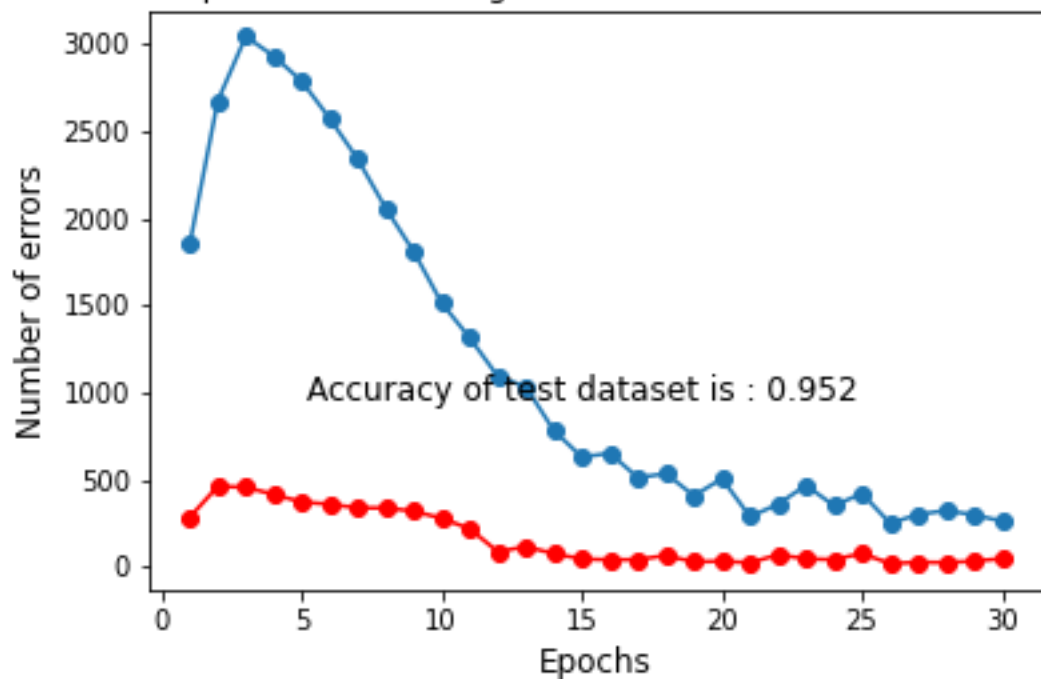
Perceptron with learning rate 1e-06 sig Activation Function



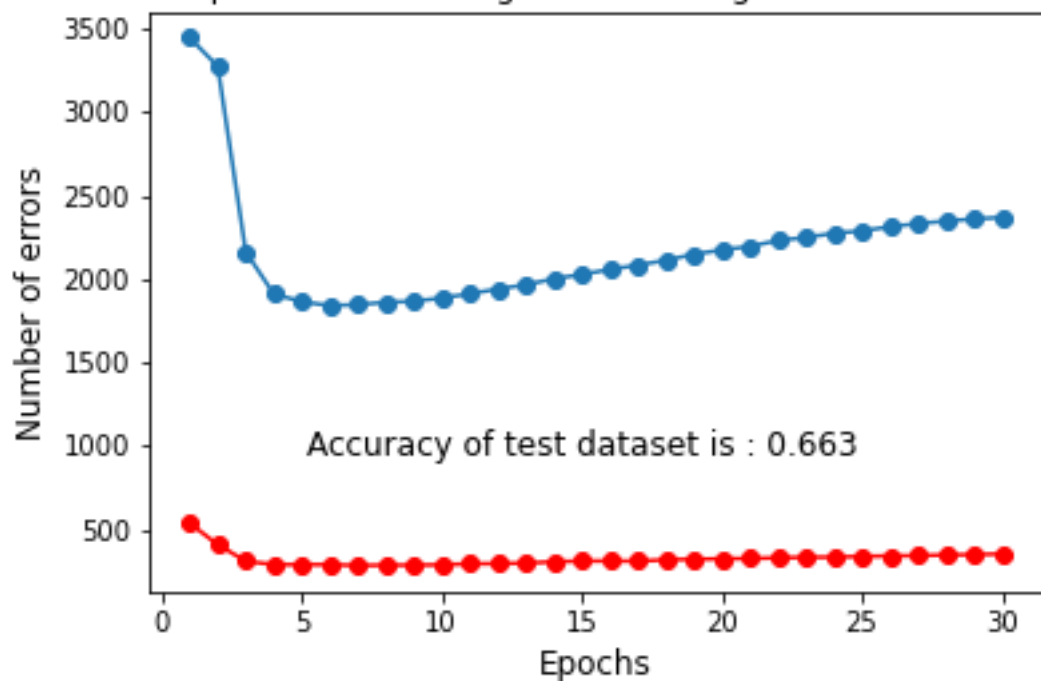
Perceptron with learning rate 1e-06 tanh Activation Function



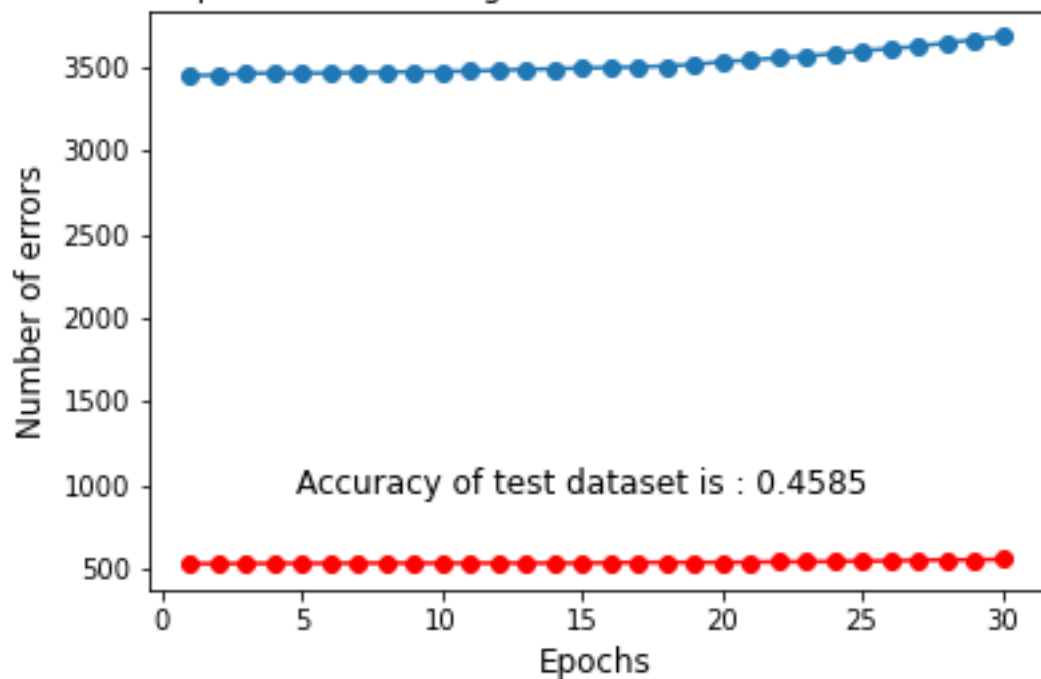
Perceptron with learning rate 1e-06 unit Activation Function



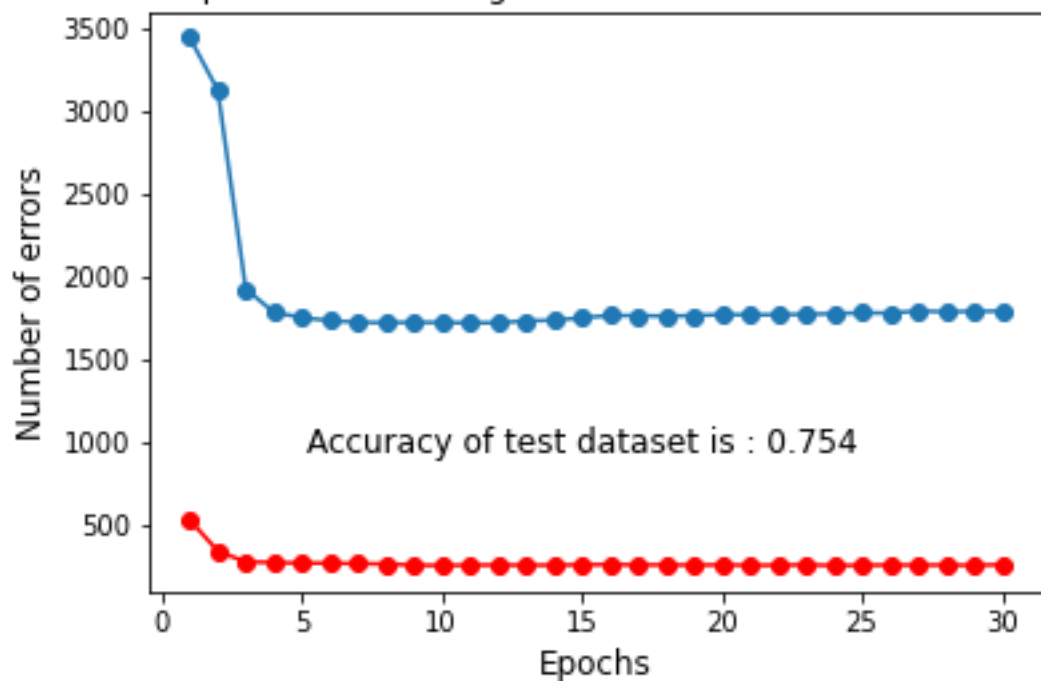
Perceptron with learning rate 1e-08 sig Activation Function



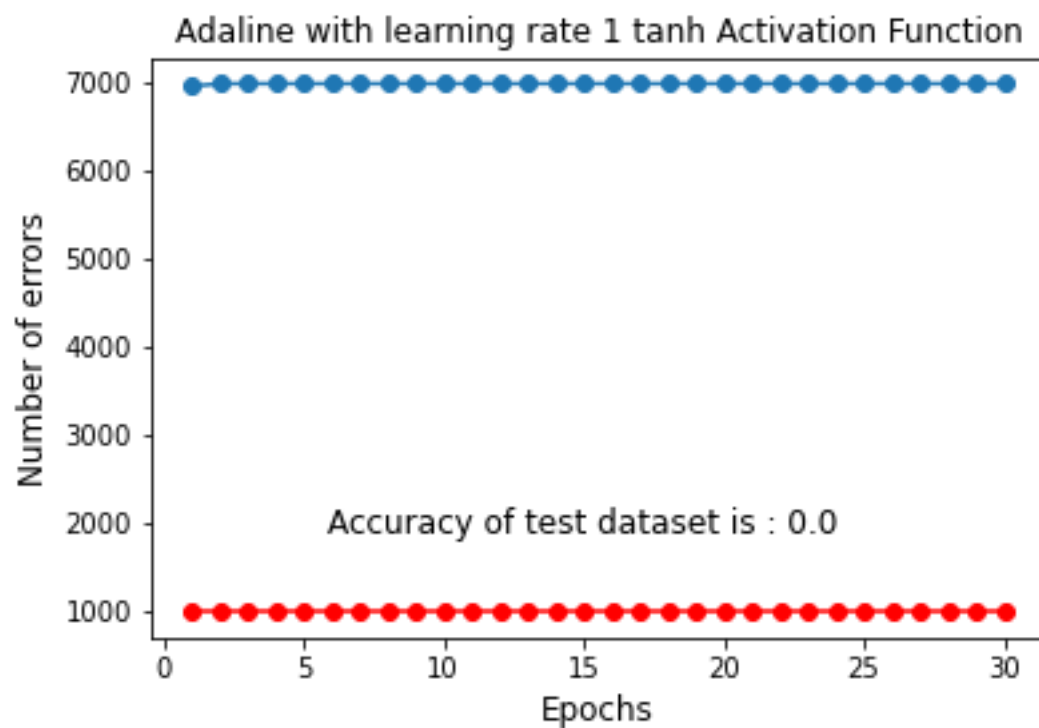
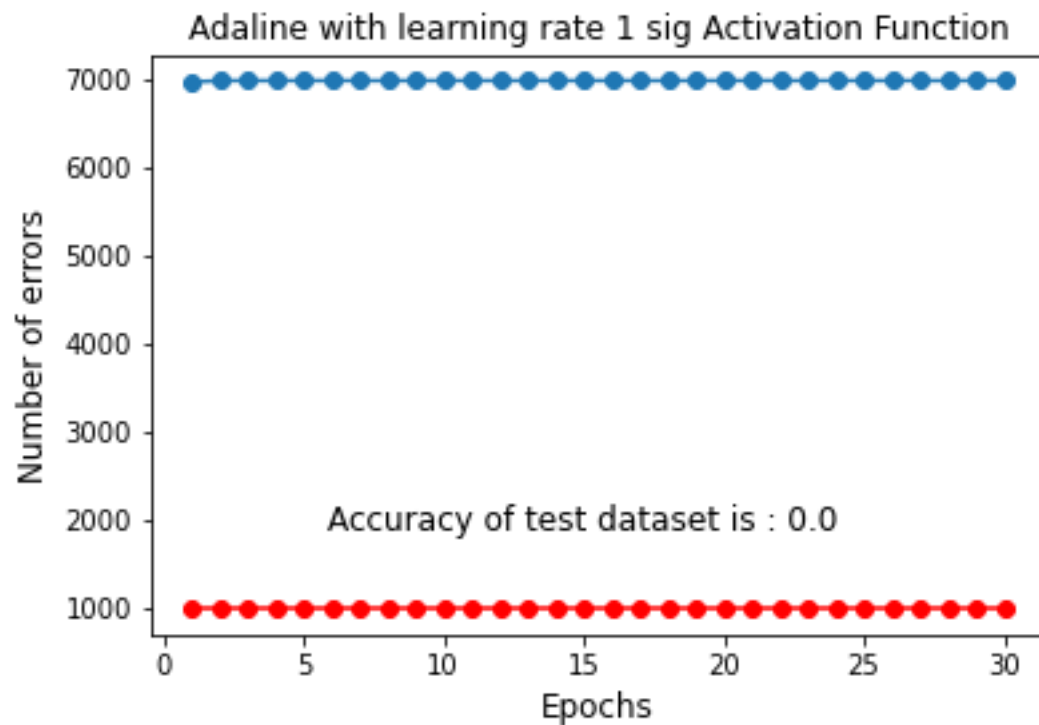
Perceptron with learning rate 1e-08 tanh Activation Function



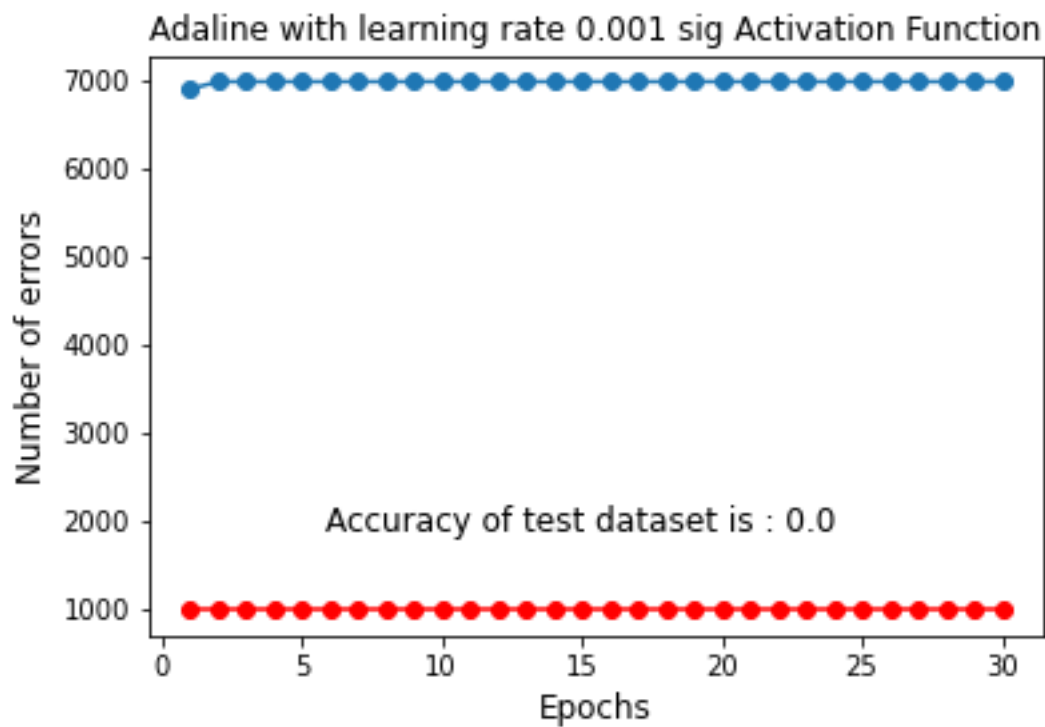
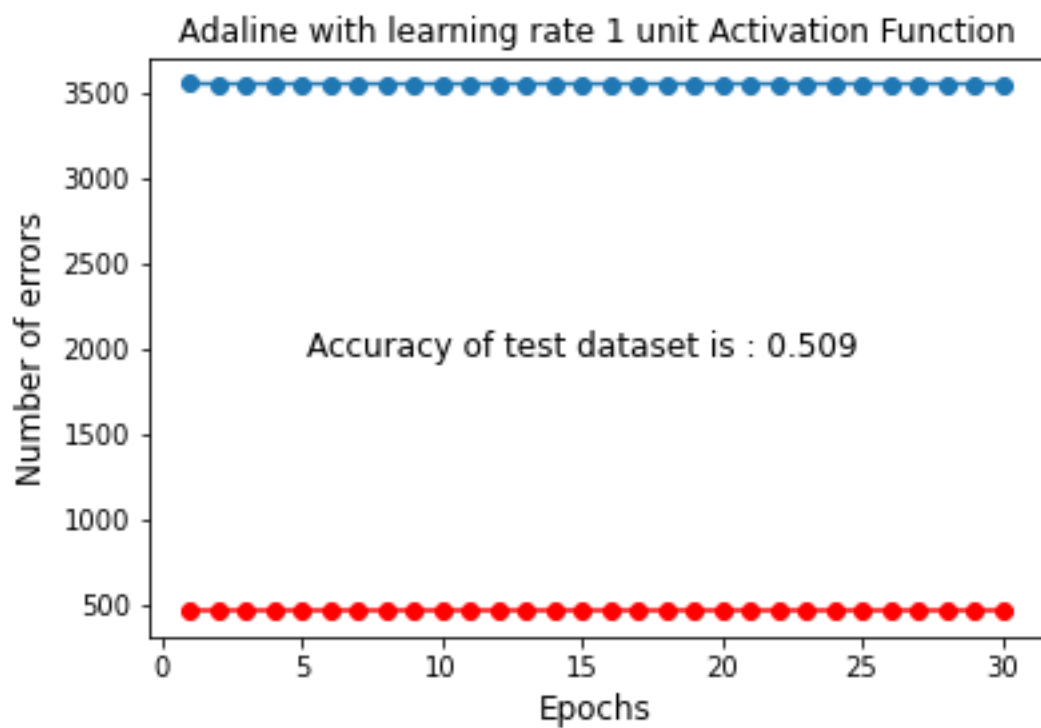
Perceptron with learning rate 1e-08 unit Activation Function

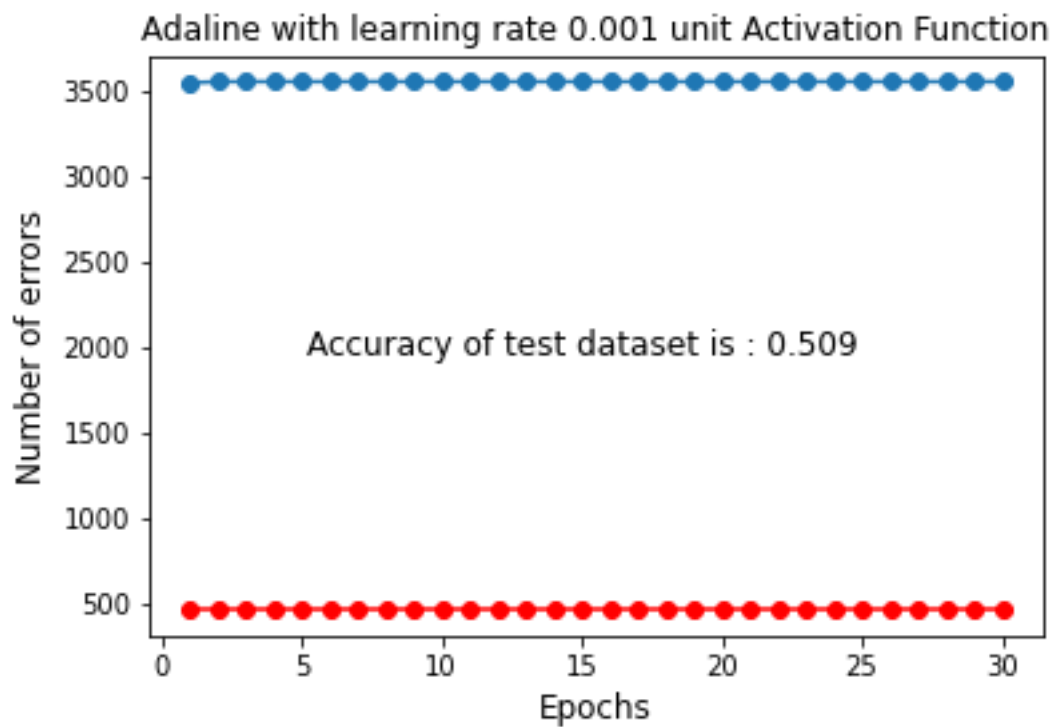
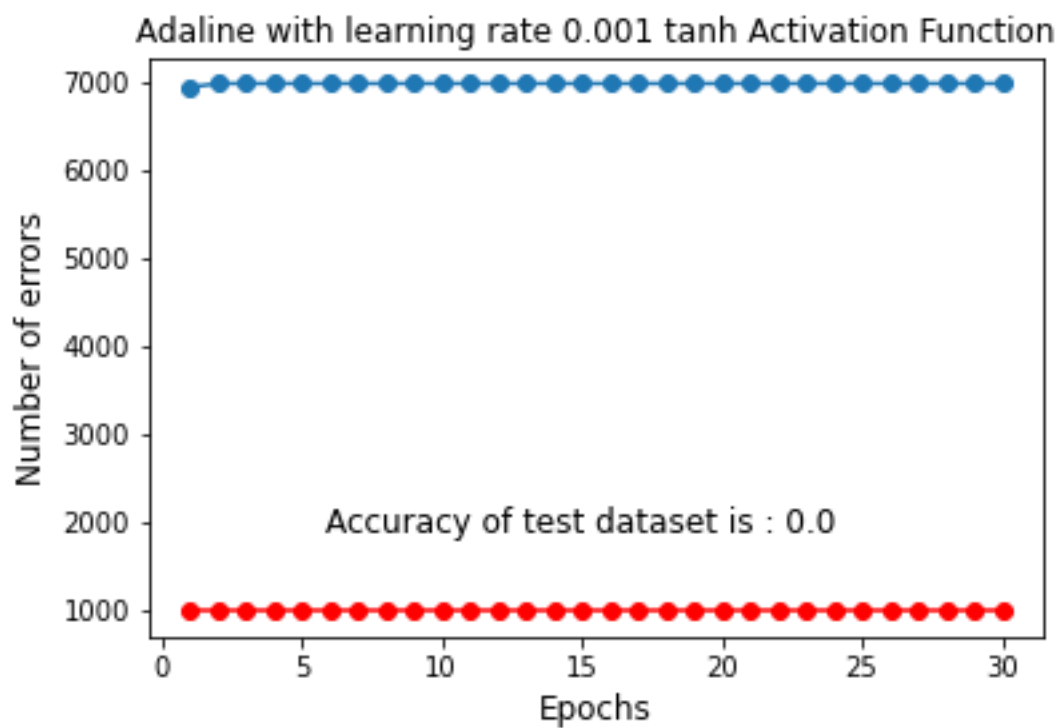


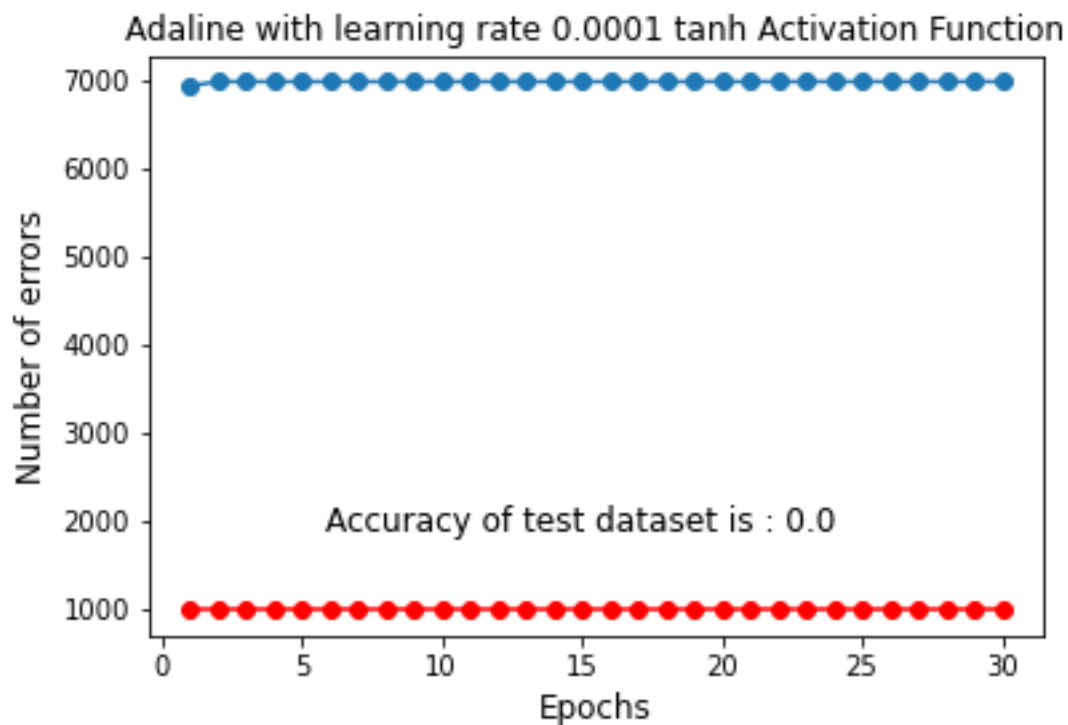
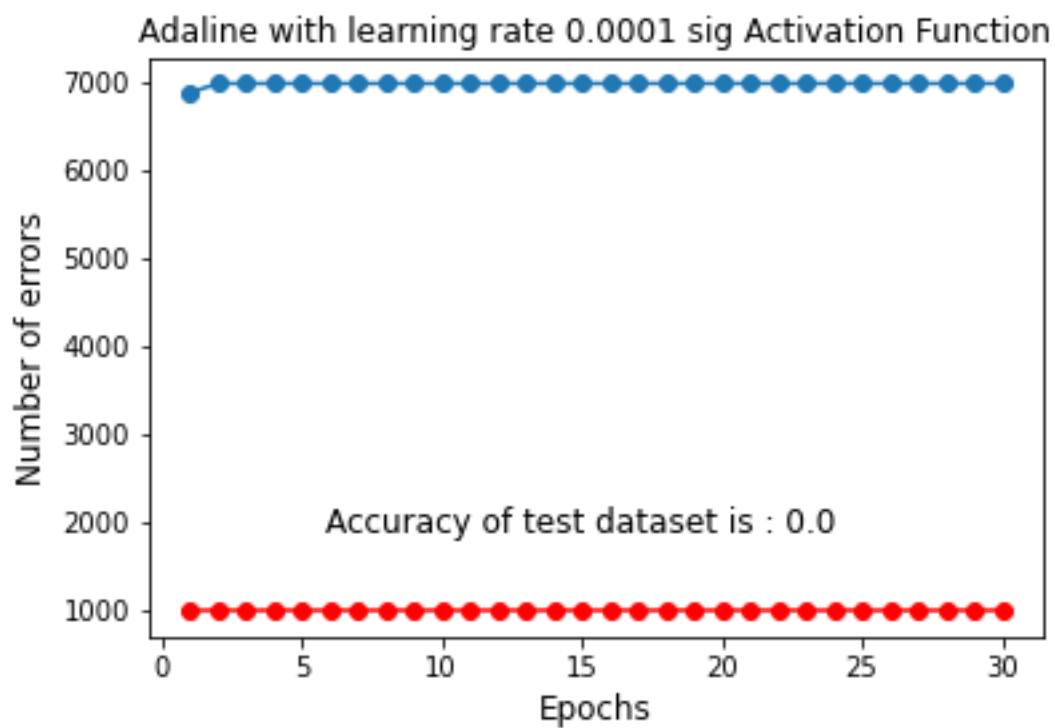
سیستم ادالاین :

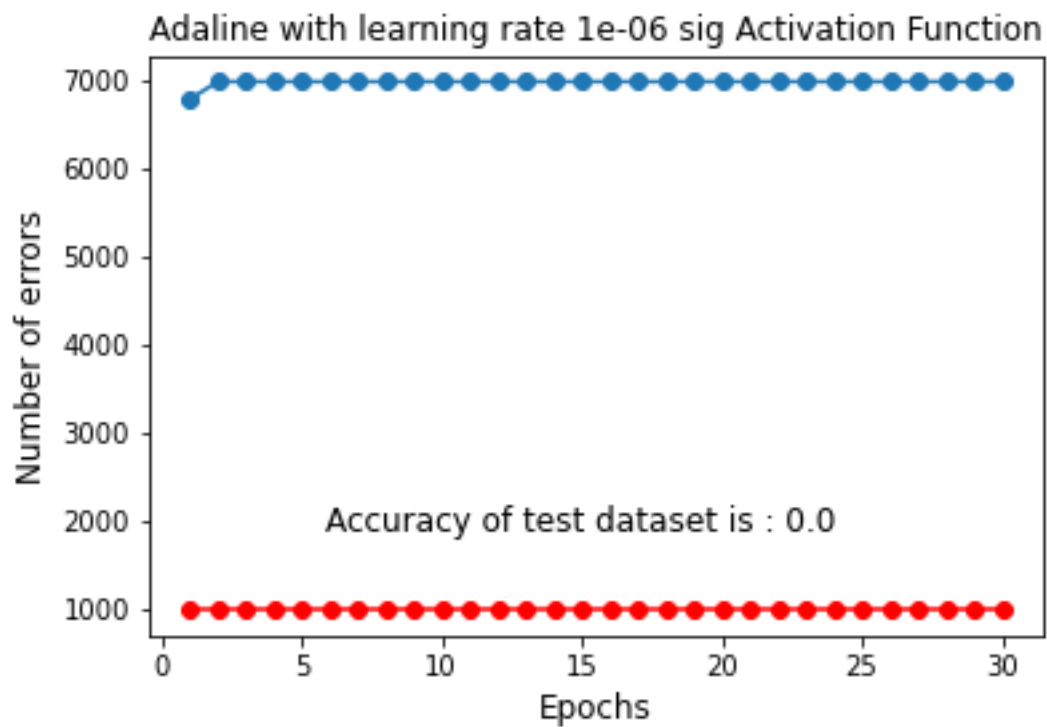
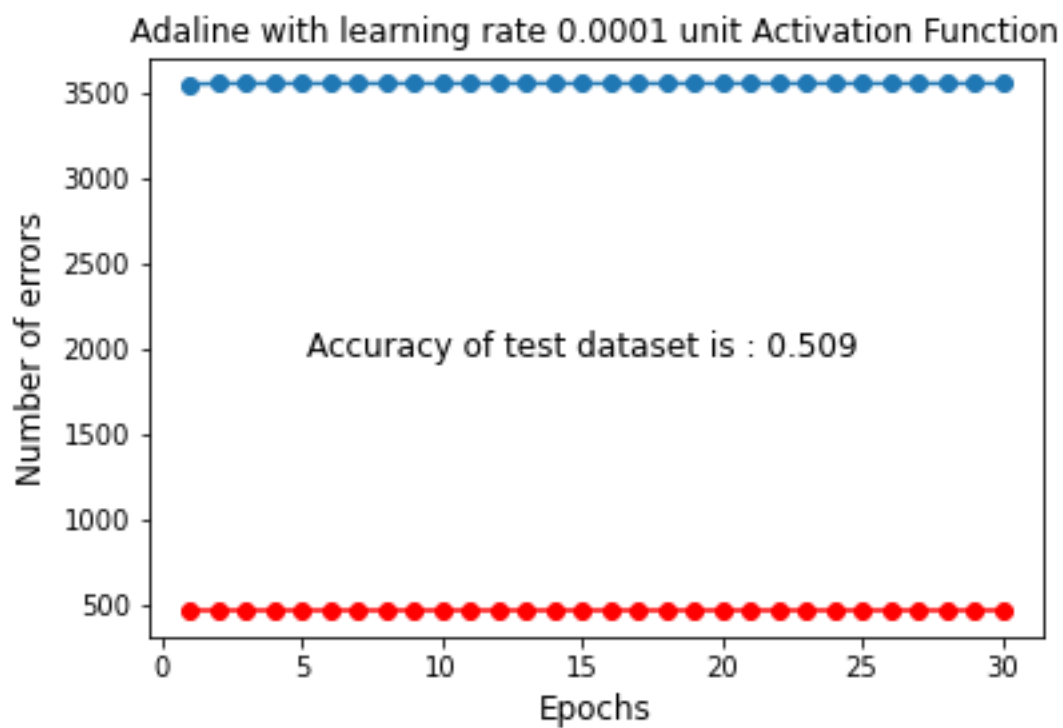


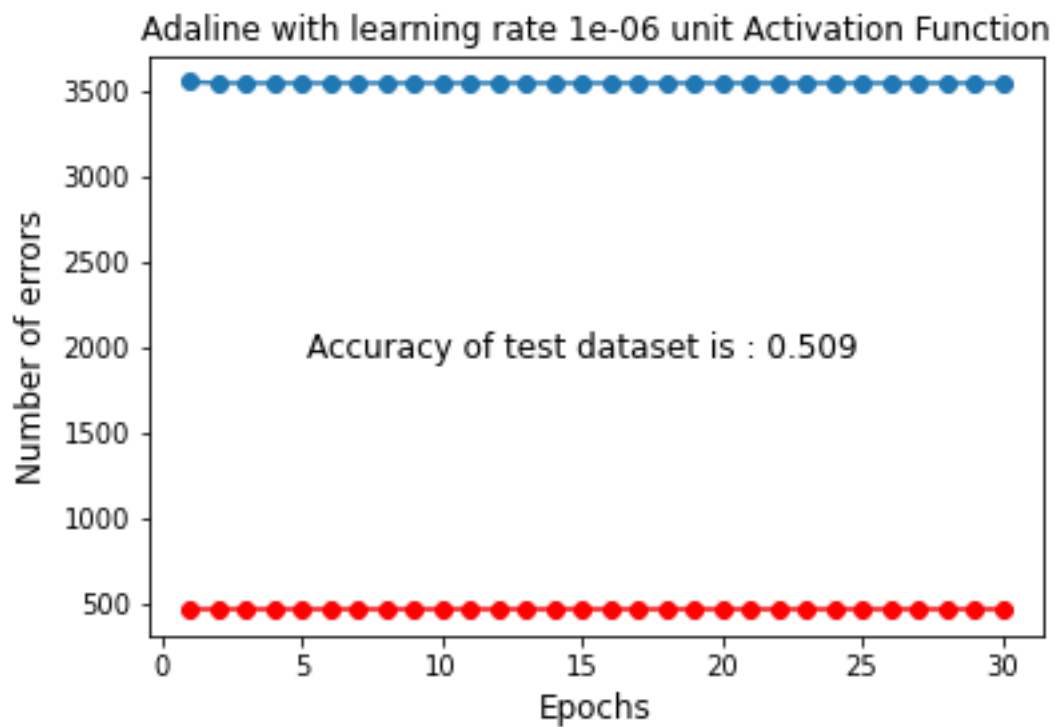
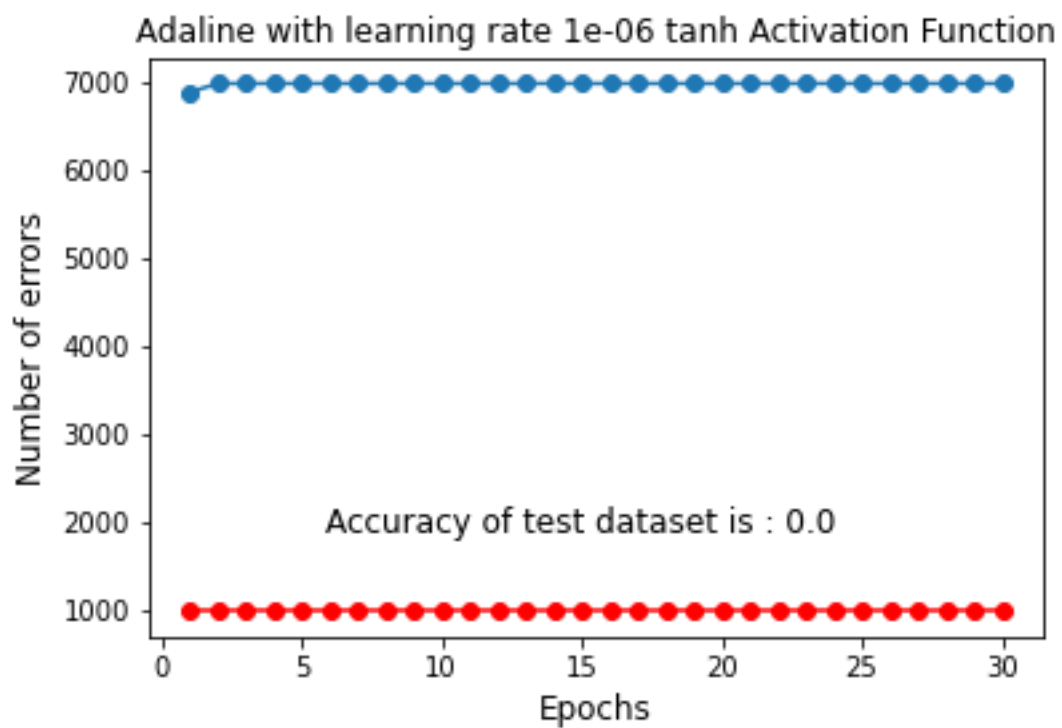




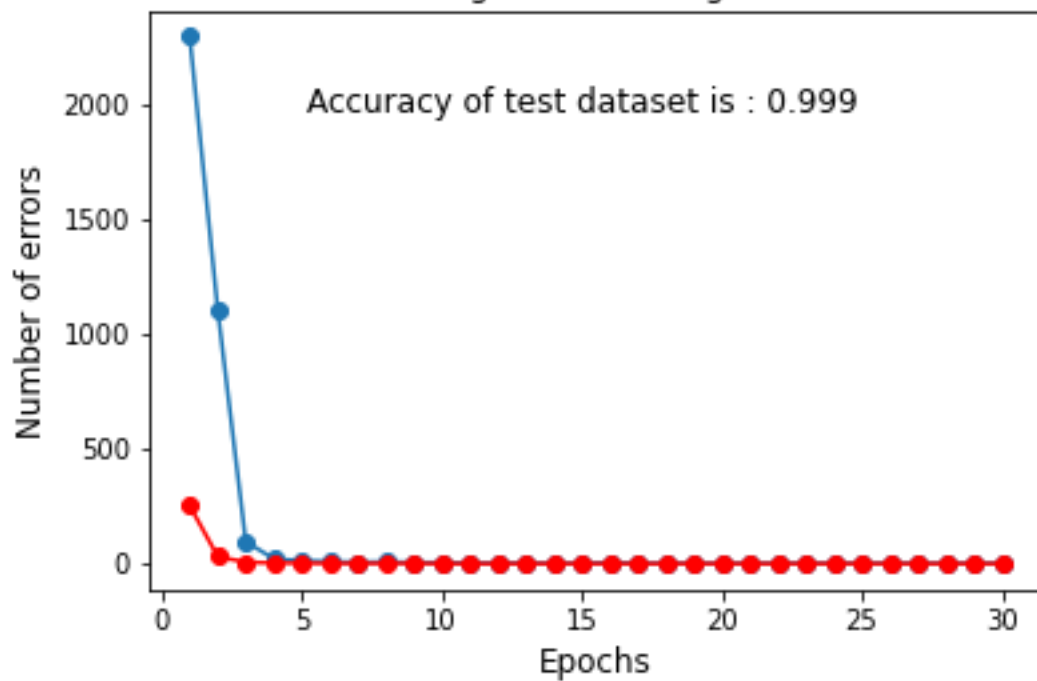




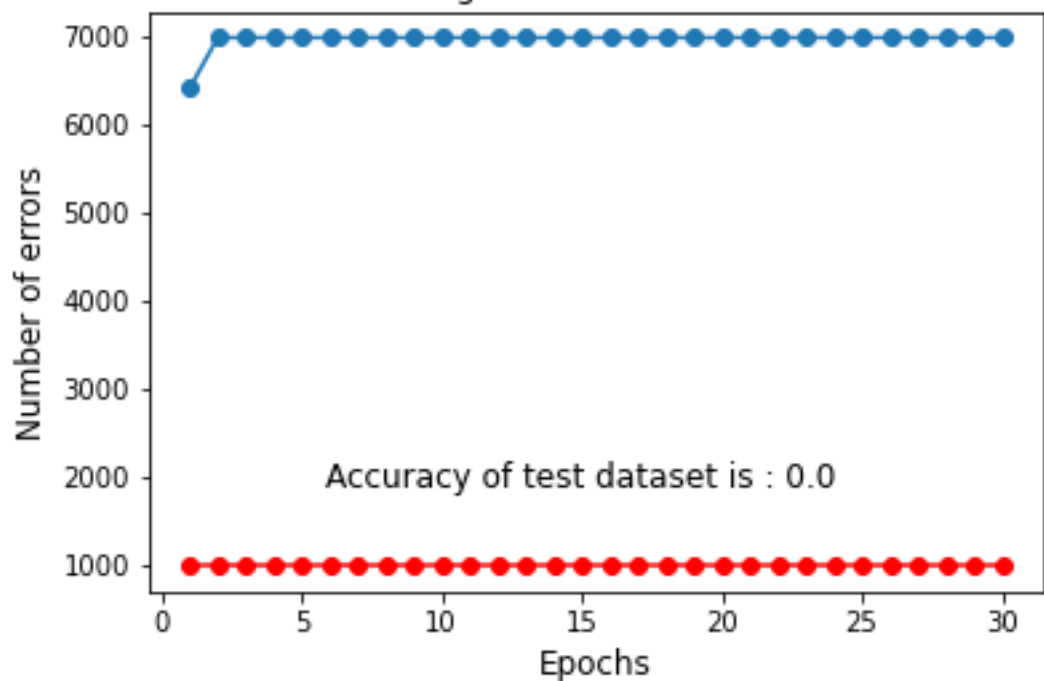


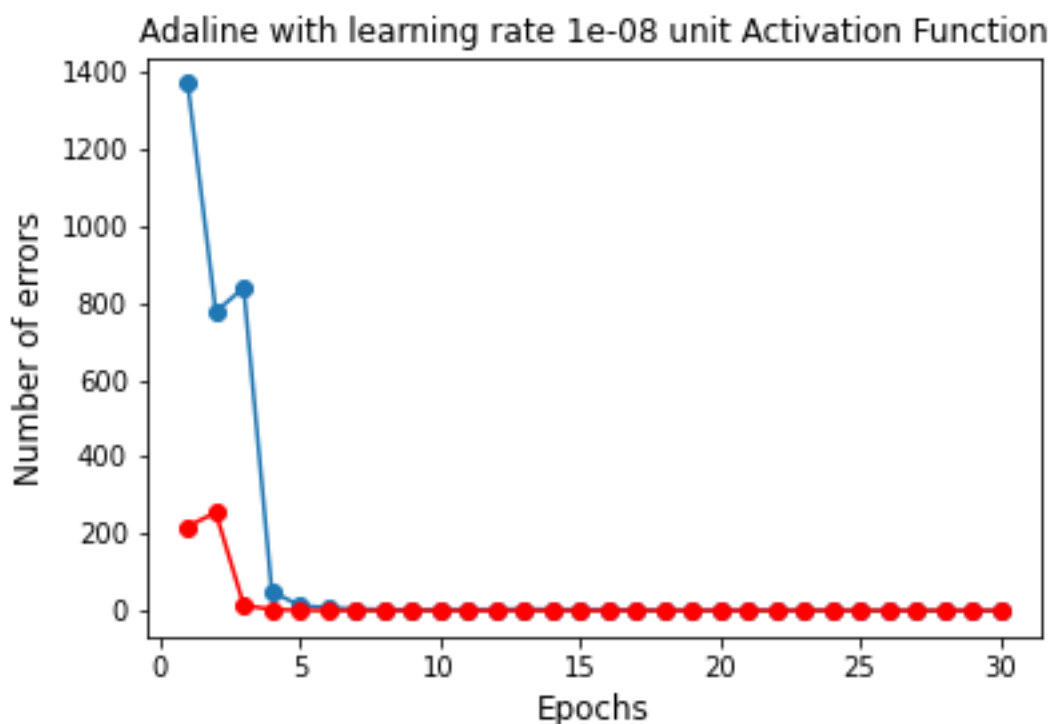


Adaline with learning rate 1e-08 sig Activation Function



Adaline with learning rate 1e-08 tanh Activation Function





شبکه های درجه یک پرسپترونی و ادالاین نمیتوانند به صورت کاملاً درست و مناسب داده های XOR را کلاس بندی کنند و تعداد زیاد از داده ها به صورت اشتباه کلاس بندی میشوند زیرا نمیتوانند به صورت خطی کلاس بندی کنند. در بیشتر حالت ها نیز داده های آموزش و اعتبارسنجی به همگرایی نمیرسند. درصد دقت نیز در طول آزمایشات ۰ بوده است.

در طرف مقابل و با مشاهده آزمایشات درجه ۲ میتوان به این نتیجه رسید که میتوان با این شبکه ها داده ها را به صورت مناسب دسته بندی کرد. برای مثال در شبکه پرسپترونی با نرخ یادگیری ۰/۰۰۰۰۱ و تابع فعالیت unit و sigmoid دقت به ۱۰۰ بسیار نزدیک است و همچنین داده های آموزش و اعتبارسنجی به همگرایی میرسند و در شبکه ادالاین نیز با نرخ یادگیری 1e-08 و تابع فعالیت sigmoid دقت به ۱۰۰ بسیار نزدیکی است و داده ها نیز به همگرایی رسیده اند.

همچنین در شبکه ادالاین بهترین نتیجه برای sigmoid ، unit ، در نرخ یادگیری 1e-08 صورت میگیرد و در مورد تابع فعالیت tanh ، در هیچ کدام از نرخ های یادگیری نتیجه مناسبی ندارد و به طول کلی برای شبکه ادالاین به هیچ عنوان مناسب نیست.

در مورد شبکه پرسپترونی ، در نرخ یادگیری ۰/۰۰۱ برای تابع فعالیت های unit , sigmoid , tanh بهترین حالت است زیرا مقدار دقت داده های تست برابر ۱ میشود و داده ها به همگرایی میرسد.

با کاهش مقدار نرخ یادگیری مقدار دقت کاهش میابد. با افزایش نیز تا حد ۱ مقادیر مناسب هستند اما بزرگتر از این مقدار مناسب نیست.

شبکه ادالاین در بهترین حالت در ۷ epoch به همگرایی رسیده و به دقت ۱ میرسد در حالتی که تابع فعالیت آن unit باشد. در شبکه پرسپترون در بهترین حالت در ۰/۰۰۱ در حالتی که تابع فعالیت آن unit است ، در ۷ epoch به همگرایی میرسد. هر دو شبکه به راحتی میتوانند کلاس بندی را انجام دهند اما شبکه پرسپترون در حالت های مختلفی میتواند مسئله را حل کند پس شبکه پرسپترون بهتر است.

از نظر بررسی تابع های فعالیت ، تابع فعالیت unit در هر دو شبکه به خوبی میتواند مسئله را حل کند و بهتر است. پس از آن تابع sigmoid است که در بعضی حالت ها پاسخ های خوبی برای مسئله ایجاد میکند. و در انتها تابع tanh قرار دارد که در شبکه ادالاین به طور قطعی میتوان گفت برای این شبکه مناسب نیست. در مورد شبکه پرسپترون نیز نسب به ۲ تابع دیگر دقت کمتری ایجاد میکند.

با تشکر فراوان

مهدی فیروزبخت

۴۰۰۱۳۱۰۲۷