به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس شبکه‌های عصبی

استاد صفابخش

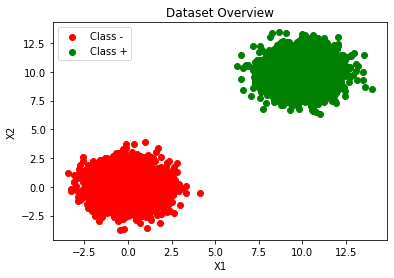
تمرین اول

علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

سوال ۱

برای ایجاد مجموعه‌داده خطی جداپذیر دو توزیع دو بعدی گاوسی با کواریانس یکسان و میانگین‌های و درنظر گرفتم. این مجموعه‌داده با احتمال بسیار بسیار بالا خطی جداپذیر است؛ چراکه فاصله دو مرکز برابر با ده واریانس در هر دو بعد است و در توزیع نرمال 99.9٪ داده‌ها در فاصله‌ای به اندازه چهار انحراف معیار از میانگین قرار گرفته‌اند. مجموعه‌داده می‌تواند چیزی مانند تصویر زیر باشد.



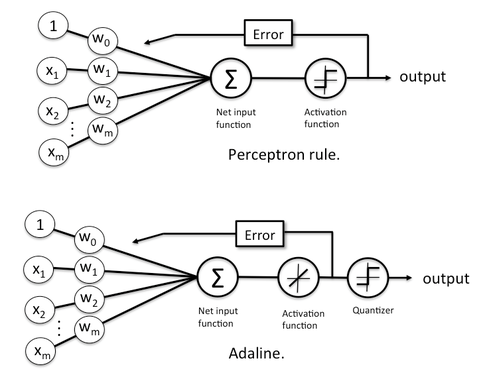
لازم به ذکر است که یک کلاس مثبت با برچسب عددی +۱ و یک کلاس منفی با برچسب عددی ۱- درنظر گرفته شده است.

سوال ۲

باتوجه به آنکه ابعاد مسئله ورودی دوبعدی است، ورودی هر دو واحد پردازشگر هم دوبعدی خواهد بود (). برای مقادیر اولیه وزن‌ها اعداد تصادفی بین 1- تا ۱+ درنظر گرفته‌ام. اما تابع فعالیت دو نورون متفاوت است که جلوتر توضیح داده می‌شود.

تفاوت دو نورون به این شرح است[[1]](#footnote-1)[[2]](#footnote-2):

* پرسپتورون از یک تابع فعالیت استفاده می‌کند که خروجی را باینری (یا به طور مشابه قطبی) می‌کند. این خروجی باینری همان خروجی نهایی نورون است. خطا و درنتیجه بروزرسانی وزن‌ها هم که برای این نورون درنظر گرفته می‌شود برپایه خروجی گسسته خواهد بود. در آدالاین به طور کلاسیک ابتدا یک تابع فعالیت همانی روی ورودی خالص (net input) لحاظ می‌شود که عملا بی‌تاثیر است. خروجی تابع فعال‌سازی این قسمت که خروجی خطی (linear output) نامیده می‌شود برای محاسبه خطا و بروزرسانی وزن‌ها استفاده می‌شود. پس در اینجا تابع خطا برخلاف پرسپتورون پیوسته خواهد بود. نهایتا از یک تابع sign به عنوان تابع آستانه در خروجی خطی نورون استفاده می‌شود تا یک خروجی قطبی (bipolar output) ایجاد شود که این خروجی پیشبینی مدل به ازای داده ورودی است و برای خطا و بروزرسانی وزن استفاده نمی‌شود. خلاصه این تفاوت در تصویر زیر مشخص است:

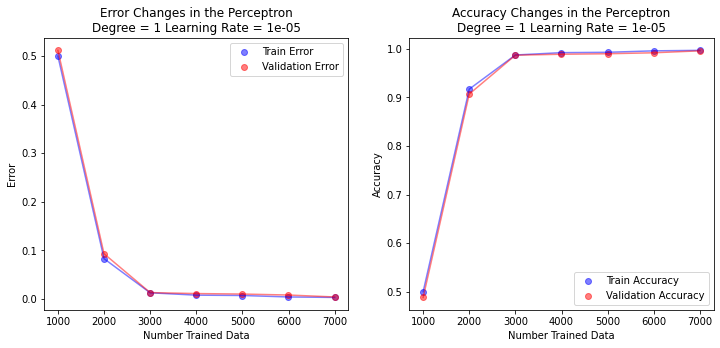


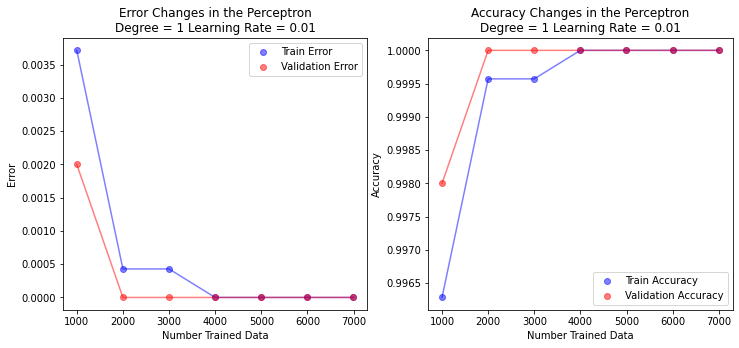
* تفاوت دیگری هم که بین دو نورون وجود دارد برای حالت درجه دو است (بر اساس تدریس کلاس). در نورون پرسپتورون درجه دوم از و استفاده می‌شود. درحالی که در آدالاین مرتبه دوم علاوه بر دو جمله مذکور از هم کمک گرفته می‌شود.

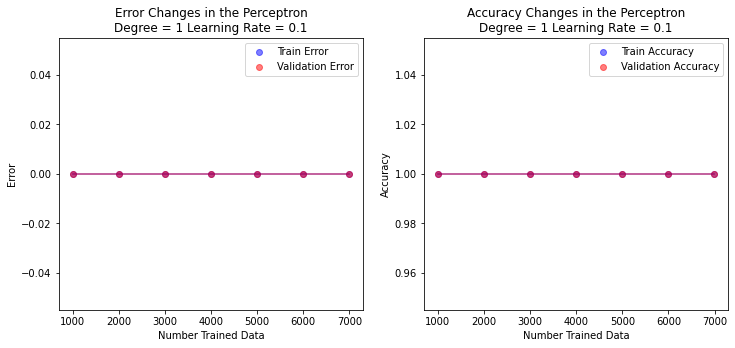
سوال ۳

برای نورون پرسپترون پنج نرخ یادگیری 0.00001، 0.01، 0.1، 1 و 100 را تست کردم که نتایج به ترتیب در ادامه آورده شده است. باتوجه به سرعت همگرایی بالای پرسپترون برای این مجموعه‌داده حتی یک گام هم برای پایین‌ترین نرخ‌های یادگیری کافی به نظر می‌آید. با توجه به نمودار‌ها بهتر است نرخ یادگیری حداقل 0.1 باشد چراکه برای نرخ‌های یادگیری کمتر از آن همگرایی کمی کندتر انجام شده است. لذا برای آزمایش بعدی نرخ یادگیری را روی مقدار 0.1 تنظیم می‌کنم. نهایتا آنکه دقت نهایی تمام نرخ‌های یادگیری در این آزمایش مطلوب است.

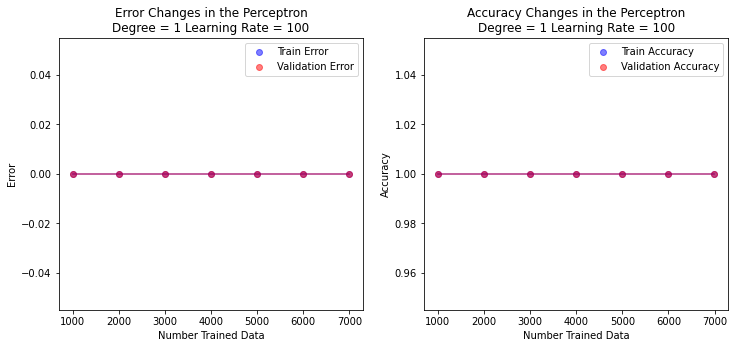
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *نرخ یادگیری* | 0.00001 | 0.01 | 0.1 | 1 | 100 |
| *صحت مجموعه آموزشی* | *0.997* | *1* | *1* | *0.999* | *1* |
| *صحت مجموعه اعتبارسنجی* | *0.996* | *1* | *1* | *1* | *1* |



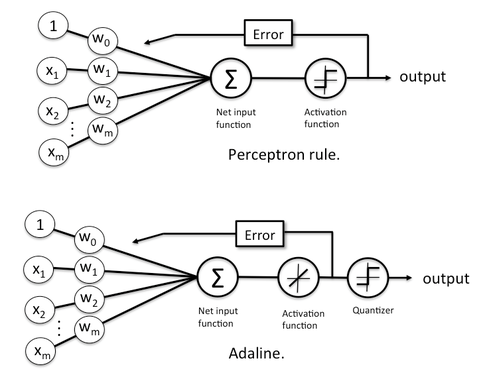








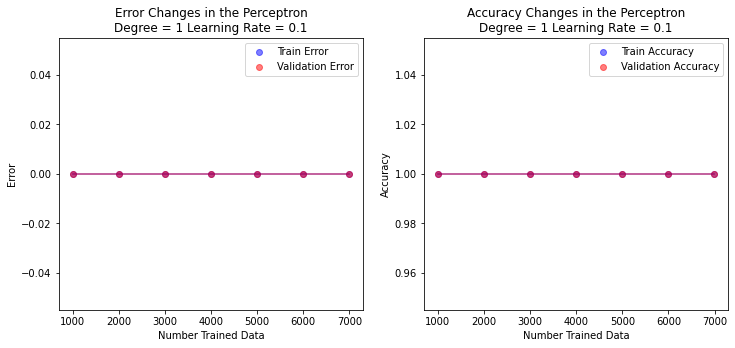
برای تابع فعال‌سازی پرسپترون باید توجه کرد مطابق معماری زیر خروجی تابع فعال‌سازی مستقیما به عنوان خروجی استفاده می‌شود؛ پس حتما باید از تابع فعال‌سازی‌ای کمک گرفت که خروجی قطبی یا باینری داشته باشد.

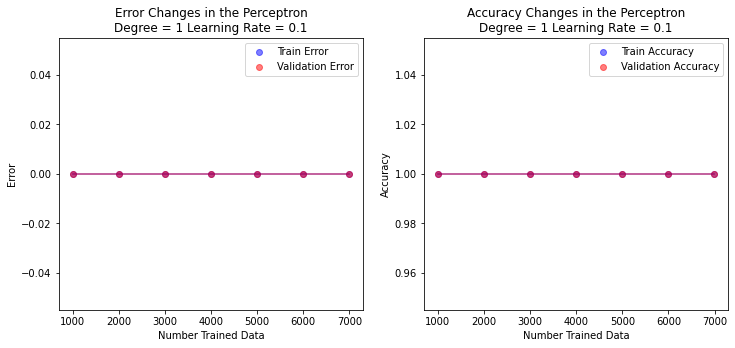


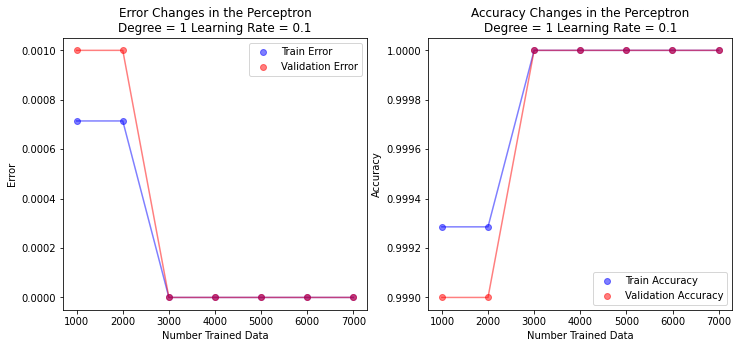
برای این کار از تابع پله زیر با سه مقدار t برابر با -۵، ۰ و +۵ استفاده کردم و نتایج این سه حالت به ترتیب در ادامه آورده شده است. لازم به ذکر است برای بررسی نرخ یادگیری از مقدار t=0 کمک گرفتم.

نتایج عملی تفاوت چشمگیری میان سه مقدار نشان نمی‌دهد. از نظر تئوری هم تفاوت زیادی بین این سه تابع فعال‌سازی وجود ندارد چراکه عملا تفاوت این سه مقدار در بایاس ورودی خالص نمود پیدا می‌کند و تاثیر مهمی در فرآیند آموزش ندارد.

|  |  |  |  |
| --- | --- | --- | --- |
| *تابع فعال‌سازی* | t = -5 | t = 0 | t = +5 |
| *صحت مجموعه آموزشی* | *۱* | *۱* | *۱* |
| *صحت مجموعه اعتبارسنجی* | *۱* | *۱* | *۱* |





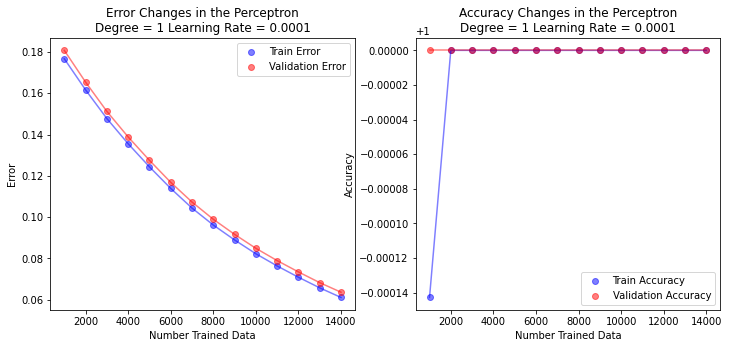


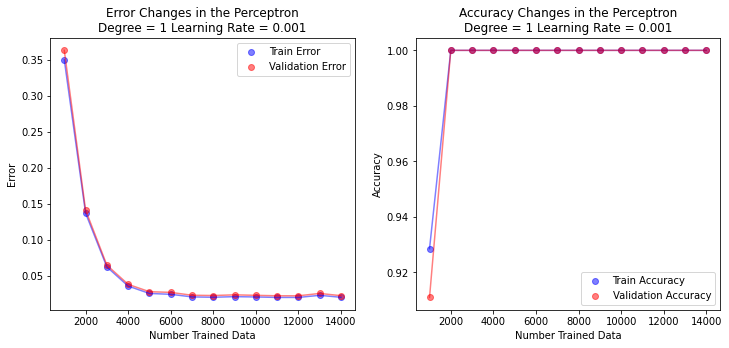
حال نوبت به بررسی نورون آدالاین می‌رسد. برای این نورون استفاده از نرخ‌های یادگیری مشابه با نورون پرسپتورون منجر به خطای زمان اجرا می‌شد و لذا به ناچار از رنج مقادیر کمتر برای نرخ یادگیری استفاده کردم. اعداد ،، ،، به ترتیب مورد بررسی قرار گرفت که نتایج آن به همراه دقت نهایی در ادامه آورده شده است. برای این حالت دو گام آموزش درنظر گرفته شده است. دو نرخ یادگیری و در زمان دو گام آموزش به همگرایی نرسیدند و مشخصا تعداد گام بیشتری برای آموزش نیاز داشتند و دقت نهایی آن‌ها هم به ترتیب 0.44 و 0.51 بوده است. از طرف دیگر برای نرخ یادگیری دقت مدل دائما نوسان داشته است و دقت 0.3 و 1.0 را در زمان کمی همزمان دیده است. این نشان می‌دهد که این نرخ یادگیری برای مدل زیاد است. دو مقدار یادگیری دیگر یعنی و از نظر دقت به سرعت به مقدار کامل همگرا شده‌اند. بین این دو بهترین گزینه به نظر می‌رسد چراکه گزینه دیگر از نظر همگرایی تابع خطا کمی کند بوده است و عملا در دو گام به همگرایی نرسیده است.

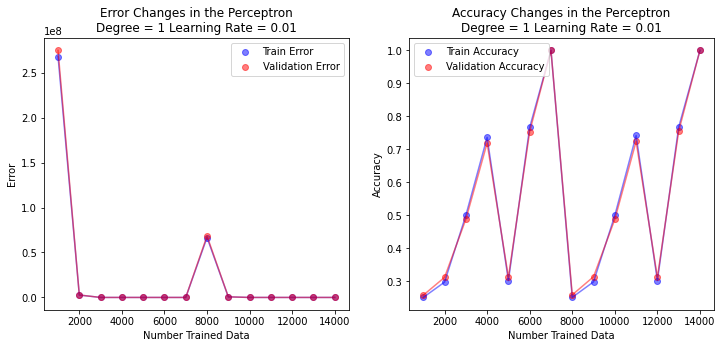
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *نرخ یادگیری* |  |  |  |  |  |
| *صحت مجموعه آموزشی* | *0.436* | *0.505* | *1* | *1* | *0.999* |
| *صحت مجموعه اعتبارسنجی* | *0.441* | *0.513* | *1* | *1* | *1* |





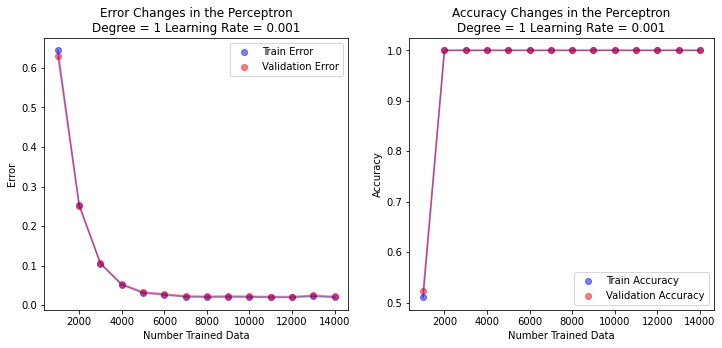


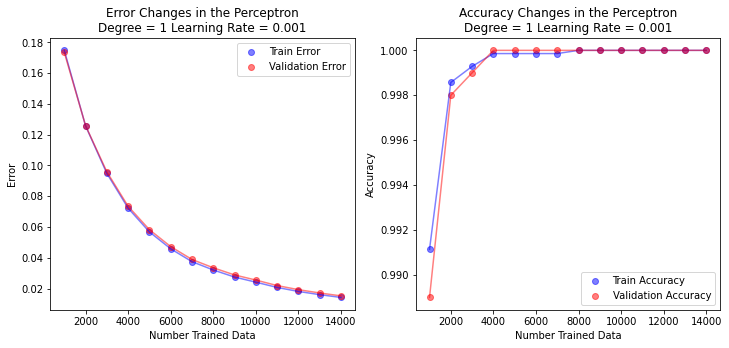


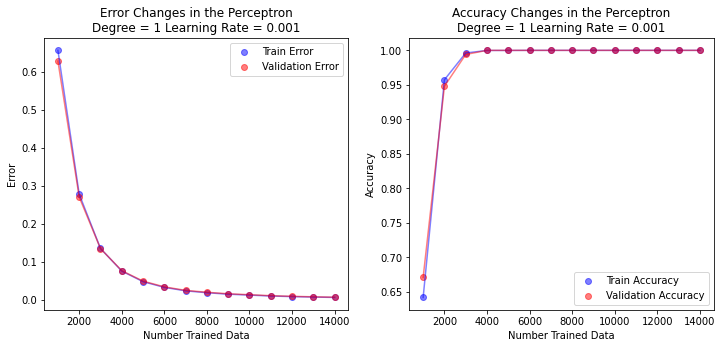


در آزمایش دیگر باید تاثیر تابع فعال‌سازی سنجیده شود. با توجه به نتایج قسمت قبل نرخ یادگیری برای این قسمت برابر با در نظر گرفته شده است. مدل آدالاین به صورت اصلی از یک تابع همانی برای فعال‌سازی استفاده می‌کند. اما می‌توان توابع دیگری را هم استفاده کرد. برای بررسی تاثیر تابع فعال‌سازی علاوه بر تابع همانی، از تابع سیگموید قطبی و تانژانت هایپربولیک استفاده کرده‌ام. لازم به ذکر است برای بررسی تاثیر نرخ یادگیری از تابع همانی کمک گرفته بودم. در ادامه به ترتیب نمودارهای مربوط به تابع همانی، سیگموید و تانژانت هایپربولیک به همراه جدول مقادیر دقت نهایی آورده شده است. با نگاه به نتایج در می‌یابیم هر سه تابع مناسب است ولی تابع سیگموید حداقل در یک بار اجرا و جمع‌آوری نتایج همگرایی کندتری نسبت به دو تابع دیگر داشته است.

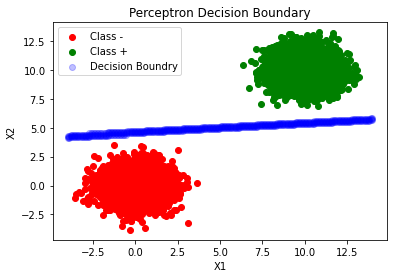
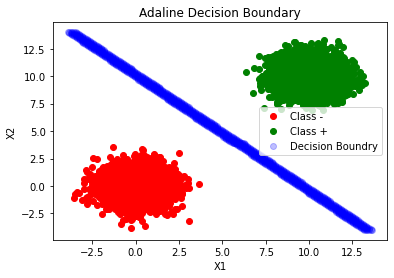
|  |  |  |  |
| --- | --- | --- | --- |
| *تابع فعال‌سازی* | همانی | سیگموید | تانژانت هایپربولیک |
| *صحت مجموعه آموزشی* | *۱* | *۱* | *۱* |
| *صحت مجموعه اعتبارسنجی* | *۱* | *۱* | *۱* |





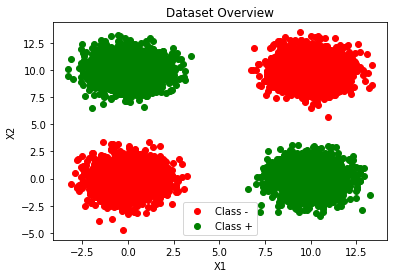


حال که نتایج هر دو نورون ارائه شده است باید بین این دو نورون و برای این مجموعه‌داده یکی از دو نورون را انتخاب کرد. اگر ملاک سرعت همگرایی باشد به نظر می‌آید نورون پرسپترون همگرایی خوبی در تمام حالات داشته است درحالی که برای آدالاین این همگرایی بعضا کمی کندتر بوده است و در برخی از آزمایش‌ها نیاز به حداقل دو گام آموزش احساس می‌شده است. اما در کنار این‌ها باید به یک عامل دیگر هم توجه داشت. در نورون پرسپترون تابع خطا صرفا به درست بودن پیشبینی اهمیت می‌دهد و یک خطای گسسته را ارائه می‌دهد درحالی که در آدالاین خطا پیوسته است و برایش مهم است که مدل با چه اطمینانی (چه فاصله‌ای از خط تصمیم) داده را به یک کلاس تعلق داده است؛ در نتیجه برای نورون پرسپترون وقتی تمام داده‌های آموزشی یک کلاس در یک طرف خط بیافتد و داده‌های دیگر در ظرف دیگر، کار خاتمه بیابد. در حالی که ممکن است این خط مماس به داده‌ها باشد ولی در آدلاین این خط تا حد ممکن از دو کلاس فاصله می‌گیرد. در تصویر زیر خط تصمیم برای هر دو نورون با بهترین تنظیمات نرخ یادگیری آورده شده است. بدیهی است که خروجی آدالاین کیفیت و اطمینان بیشتری دارد. برای همین در مجموع من ترجیح می‌دهم برای این مجموعه‌داده از یک نورون آدالاین استفاده کنم ولی پرسپترون هم بدون مشکل است.



سوال ۴

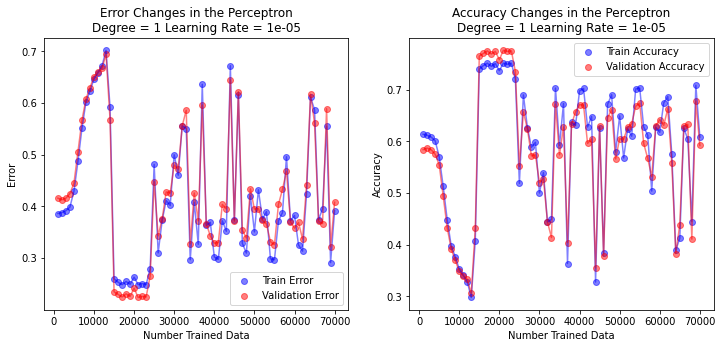
باتوجه به آنکه در صورت سوال به صورت غیرمستقیم درخواست شده است که مجموعه‌داده از نوع XOR باشد، من چهار توزیع گاوسی در چهار نقطه از فضا با کواریانس در نظر گرفتم. میانگین این چهار توزیع عبارت است از ، ،، و . توزیع اول و چهارم برای کلاس و دو توزیع دیگر به کلاس دیگر متعلق است. مشابه استدلال‌های ارائه‌شده در سوال ۱ می‌توان با اطمینان بسیار بالا مطمئن شد که داده‌ها این چهار توزیع باهم ترکیب نخواهند شد. مجموعه‌داده به شکل زیر به وجود می‌آید. بدیهی است که مجموعه خطی جداپذیر نیست ولی استفاده از نورون با درجات بالا ممکن است مشکل را حل کند.

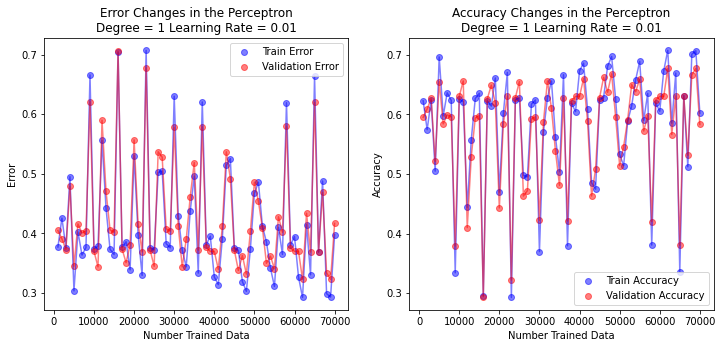


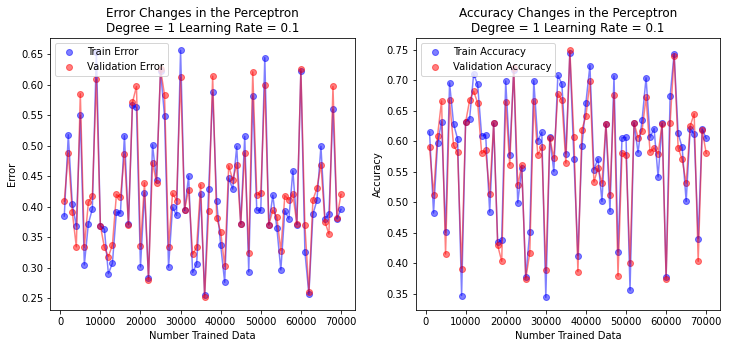
سوال ۵

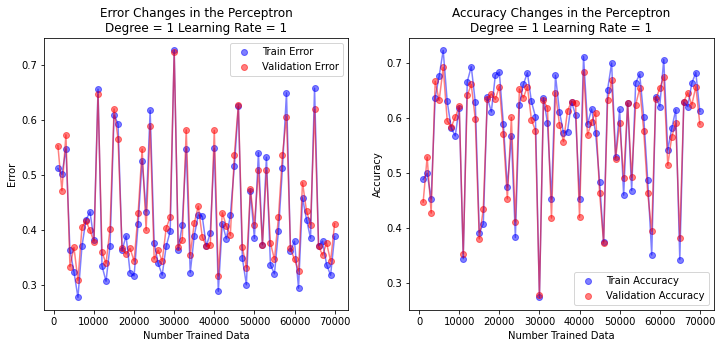
برای پرسپترون مشابه قبل مقادیر نرخ یادگیری 0.00001، 0.01، 0.1، 1 و 100 را در نظر گرفته‌ام. در اینجا باتوجه به نوساناتی که در نتایج دیده شد، من تصمیم گرفتم از 10 گام آموزش استفاده کنم. نمودار‌های خطا و جدول نتایج در ادامه آورده شده است. همانطور که از نتایج بر می‌آید، پرسپترون در مواجه با این مجموعه‌داده دائما در نوسان است و به نقطه مشخصی نمی‌تواند همگرا شود. دقت‌های خروجی هم جالب نیستند. البته باتوجه به آنکه پرسپترون تنها یک خط می‌تواند ترسیم کند انتظار دقت بالای 0.75 ممکن نیست.

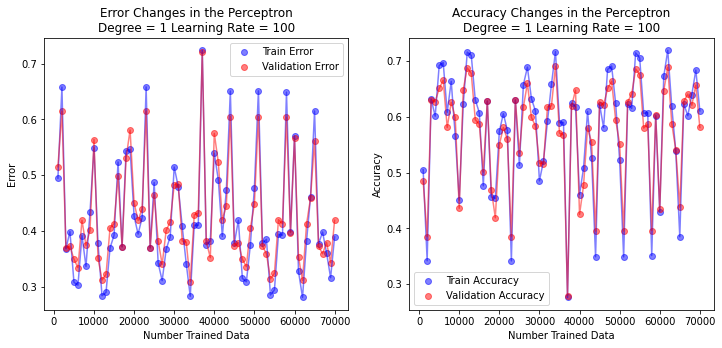
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *نرخ یادگیری* | 0.00001 | 0.01 | 0.1 | 1 | 100 |
| *صحت مجموعه آموزشی* | *0.608* | *0.602* | *0.604* | *0.612* | *0.610* |
| *صحت مجموعه اعتبارسنجی* | *0.592* | *0.583* | *0.580* | *0.589* | *0.581* |





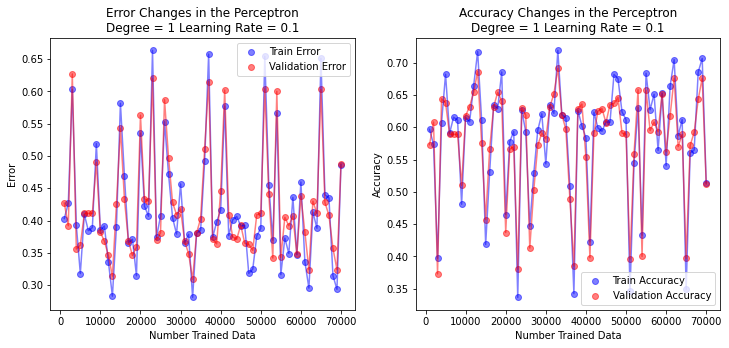


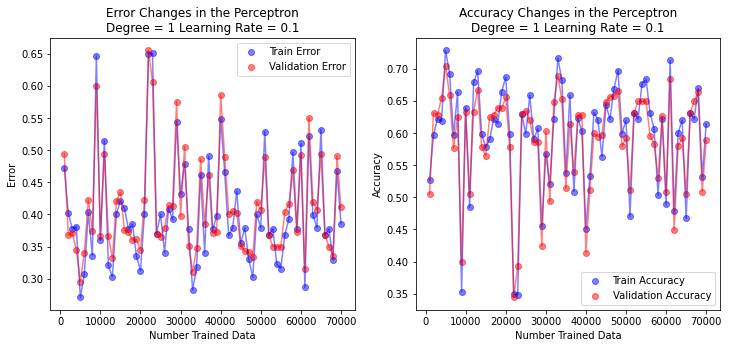


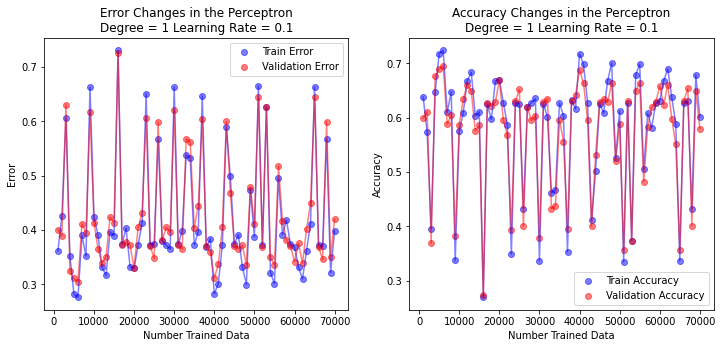


از نظر تئوری بدیهی است که تغییر تابع فعال‌سازی هم نمی‌تواند به پرسپترون کمک کند ولی مشابه با قسمت قبل سه تابع پله را امتحان کردیم که نتایج برای سه مقدار t برابر با ۵-، ۰ و ۵+ به ترتیب در ادامه آورده شده است.

|  |  |  |  |
| --- | --- | --- | --- |
| *تابع فعال‌سازی* | t = -5 | t = 0 | t = +5 |
| *صحت مجموعه آموزشی* | *0.513* | *0.614* | *0.601* |
| *صحت مجموعه اعتبارسنجی* | *0.512* | *0.589* | *0.580* |

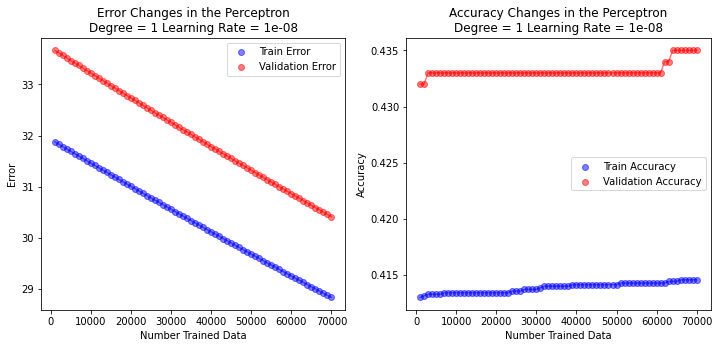


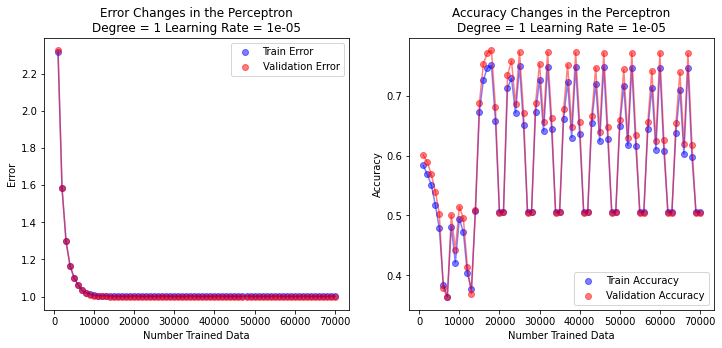


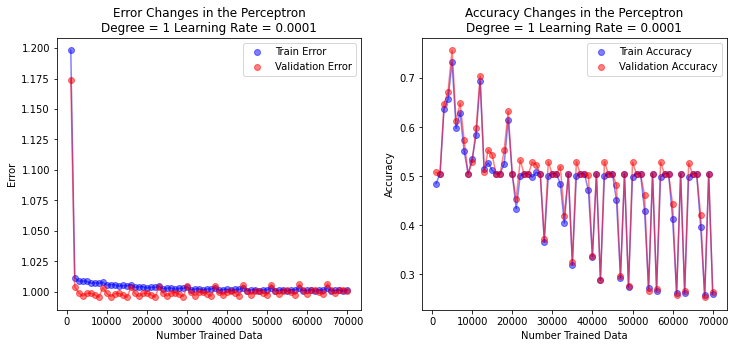


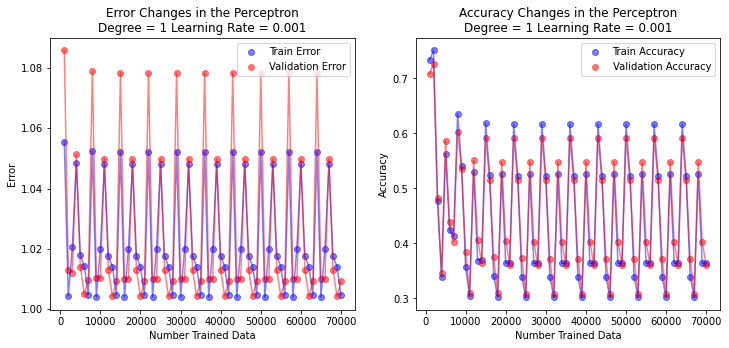
برای نرخ‌های یادگیری آدالاین مطابق با قسمت قبل اعداد ، ، ،، را در نظر می‌گیریم. تعداد گام برای این حالت را ۱۰ درنظر می‌گیرم. به ترتیب نتایج زیر حاصل می‌شود. برای آدالاین هم مشابه با پرسپترون همگرایی دیده نمی‌شود و خروجی بین مقادیر متفاوت نوسان می‌کند. البته در مورد باتوجه به نرخ بسیار پایین آموزش مدل کلا آموزش کندی دارد و نوسانی دیده نمی‌شود.

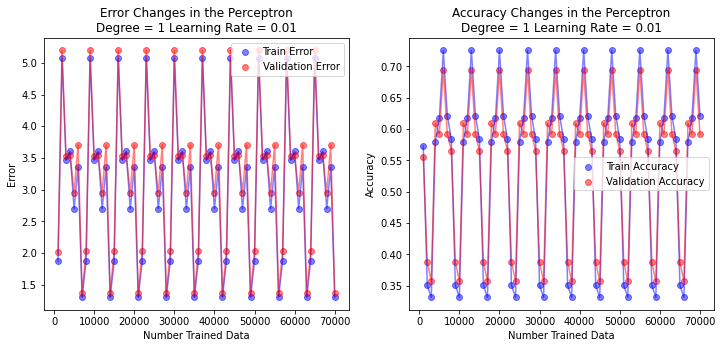
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *نرخ یادگیری* |  |  |  |  |  |
| *صحت مجموعه آموزشی* | *0.414* | *0.505* | *0.260* | *0.364* | *0.621* |
| *صحت مجموعه اعتبارسنجی* | *0.435* | *0.504* | *0.265* | *0.360* | *0.592* |





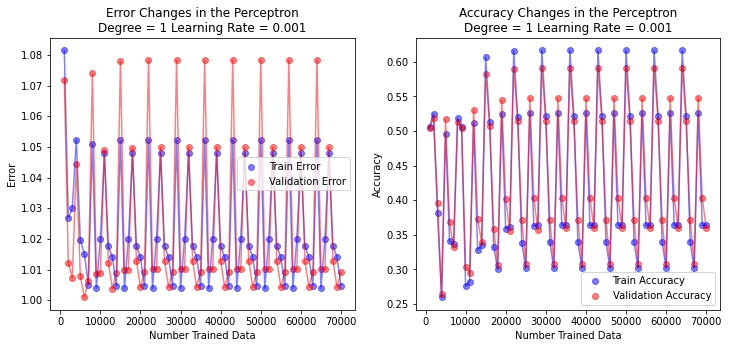


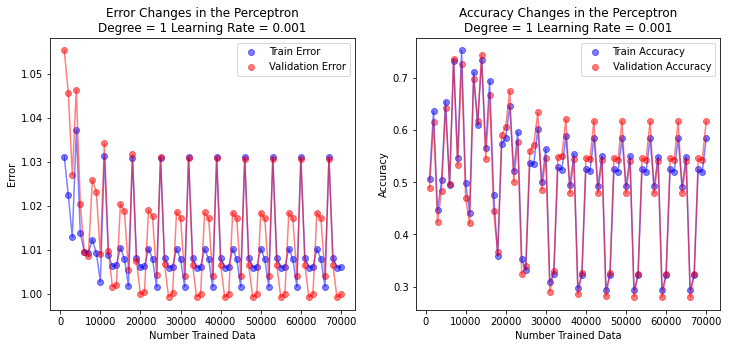


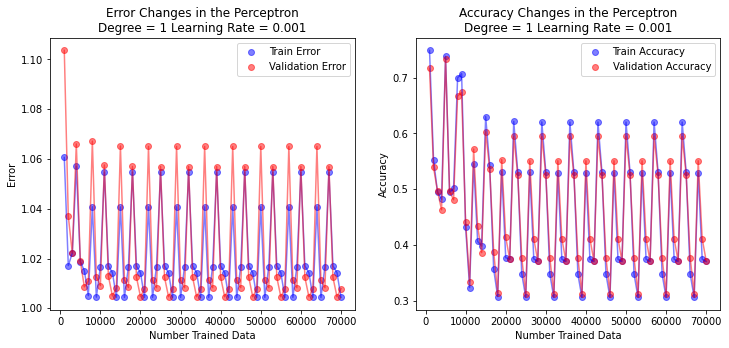


برای تابع فعال‌سازی باز مشابه قبل توابع همانی، سیگموید و تانژانت هایپربولیک را تست کردم که نتایج در ادامه آورده شده است. مطابق انتظار تغییر تابع فعال‌سازی هم نقش مثبتی ندارد. در میان نتایج اگرچه سیگموید از دو تابع دیگر دقت‌های بالاتری دارد ولی باید توجه داشت که نمودار خطای این تابع هم مثل مابقی بوده است و تنها فرقی که وجود داشته این است که موقع ارزیابی در بالای موج نوسان بوده است.

|  |  |  |  |
| --- | --- | --- | --- |
| *تابع فعال‌سازی* | همانی | سیگموید | تانژانت هایپربولیک |
| *صحت مجموعه آموزشی* | *0.364* | *0.584* | *0.370* |
| *صحت مجموعه اعتبارسنجی* | *0.360* | *0.617* | *0.371* |





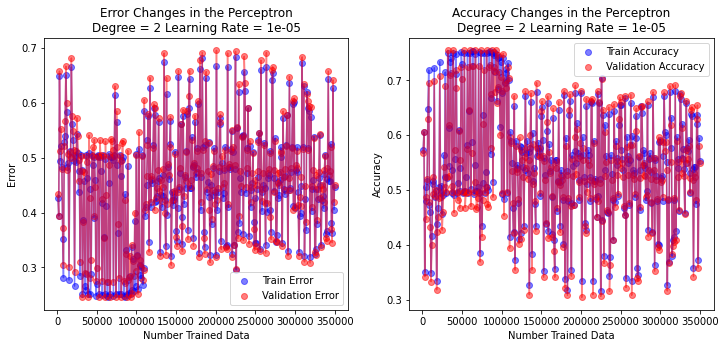


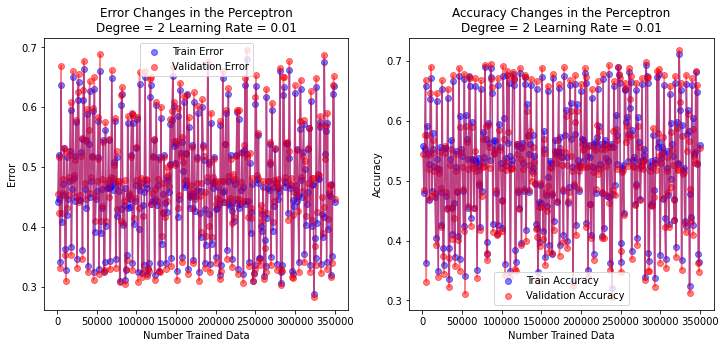
سوال ۶

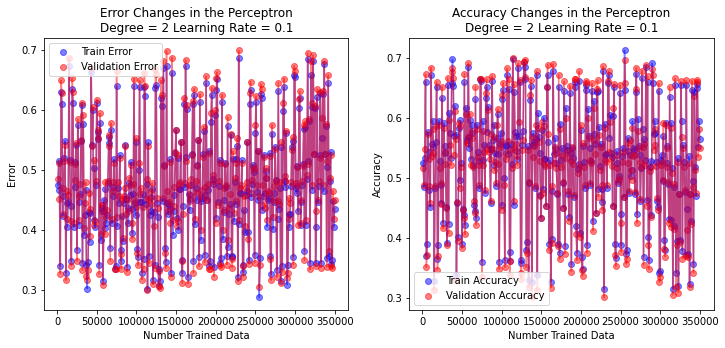
برای پیاده‌سازی نورون‌های درجه دو من دقیقا از معماری موجود در اسلایدهای درس استفاده کردم؛ یعنی برای پرسپترون از جملات 1، ، ، و کمک گرفته‌ام و برای آدالاین علاوه بر پنج جمله فوق از جمله هم استفاده کردم.

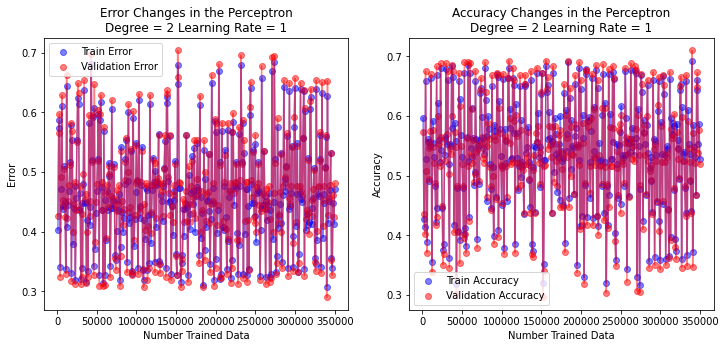
برای پرسپترون درجه دو از نرخ‌های یادگیری 0.00001، 0.01، 0.1، 1 و 100 استفاده کردم. برای مجموعه‌داده XOR و برای این نورون از 50 گام آموزش استفاده کردم تا مطمئن شوم که مدل می‌توانسته است همگرا شود یا خیر. نتایج این پنج حالت به همراه جدول مقادیر دقت نهایی در ادامه آورده شده است.

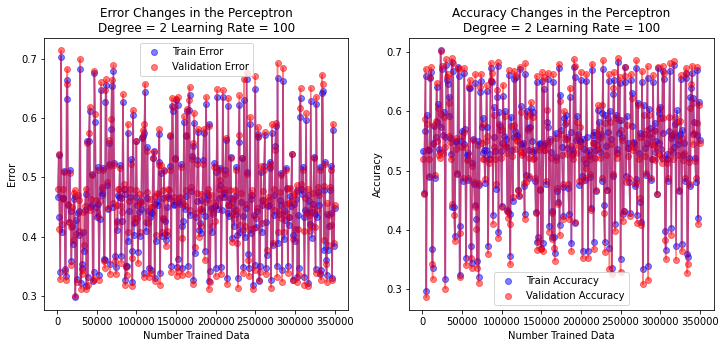
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *نرخ یادگیری* | 0.00001 | 0.01 | 0.1 | 1 | 100 |
| *صحت مجموعه آموزشی* | *0.553* | *0.559* | *0.565* | *0.528* | *0.551* |
| *صحت مجموعه اعتبارسنجی* | *0.549* | *0.554* | *0.550* | *0.519* | *0.546* |



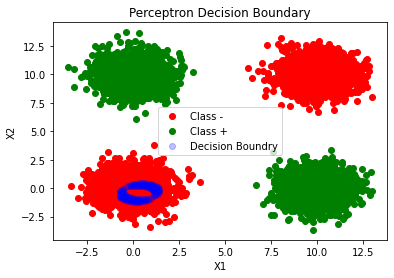
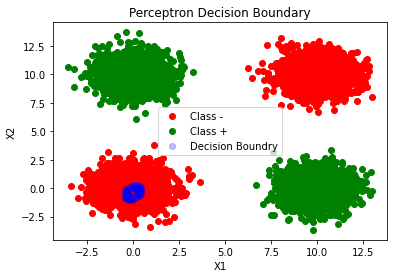








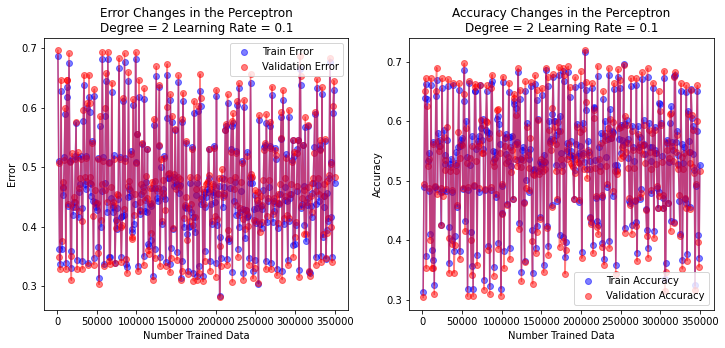
به نظر می‌رسد که پرسپترون درجه دو هم مانند پرسپترون درجه یک ناموفق بوده است اما چرا؟؛ به نظرم نبود جمله باعث این موضوع شده است. اگر این جمله وجود داشت بیضی یا هذلولی جداکننده دو کلاس امکان چرخش داشت ولی الان چنین امکانی وجود ندارد و مرزهایی مانند مرزهای زیر امکان تشکیل دارند. نهایتا در این مجموعه‌داده هم باتوجه به محل قرار گیری داده‌های دو کلاس به مشکل می‌خوریم.

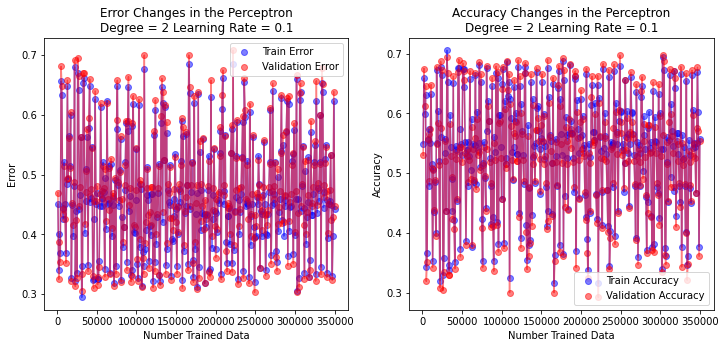


برای بررسی تابع فعال‌سازی مانند پرسپترون درجه یک از تابع پله با سه مقدار ۵-، ۰ و ۵+ استفاده کرده‌ام که نتایج آن در ادامه آورده شده است. مطابق انتظار تغییر تابع فعال‌سازی تغییر مثبتی در نتایج حاصل‌شده نداشت.

|  |  |  |  |
| --- | --- | --- | --- |
| *تابع فعال‌سازی* | t = -5 | t = 0 | t = +5 |
| *صحت مجموعه آموزشی* | *0.568* | *0.526* | *0.557* |
| *صحت مجموعه اعتبارسنجی* | *0.561* | *0.516* | *0.553* |

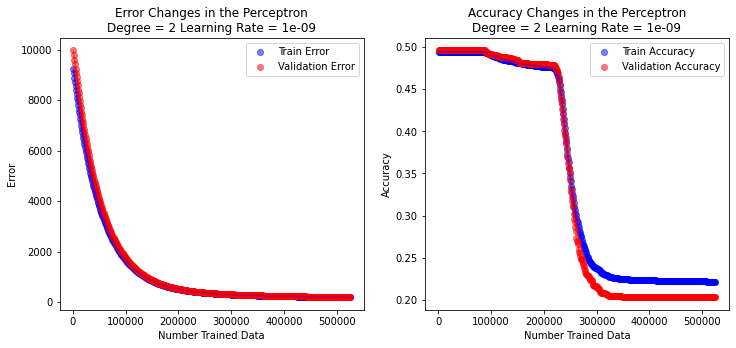


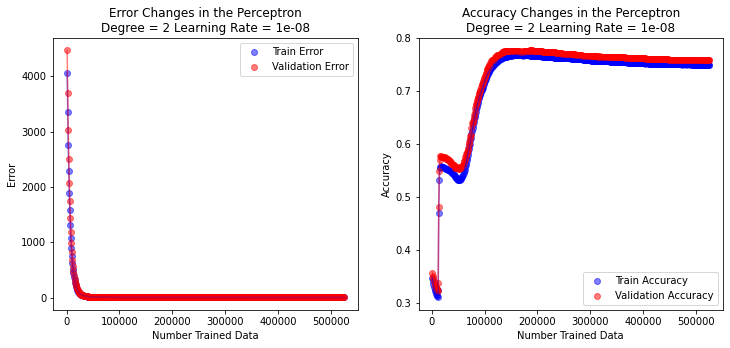


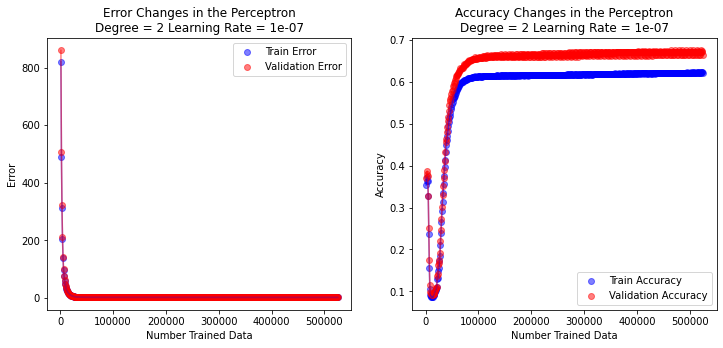


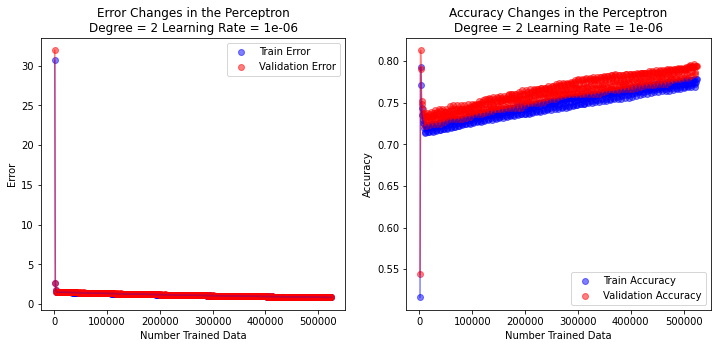
برای آدالاین مرتبه دوم استفاده از برخی نرخ‌های یادگیری مشابه با آدالاین مرتبه اول منجر به خطای زمان اجرا می‌شد که من را مجبور کرد تا از مقادیر کمتری برای این آزمایش کمک بگیرم. اعداد ، ، ،، به عنوان نرخ یادگیری استفاده شد که نتایج مربوط به هر یک به همراه دقت‌های نهایی در ادامه آورده شده است. در نهایت توجه کنید که برای این قسمت از 75 گام آموزش استفاده کردم. با توجه به نتایج به نظر می‌رسد برای به خروجی مناسبی رسیده‌ایم و همگرایی هم در تعداد گام معقولی انجام شده است. اما چنین خروجی مناسب به همراه همگرایی خوب را در سایر نرخ‌های یادگیری مشاهده نمی‌کنیم.

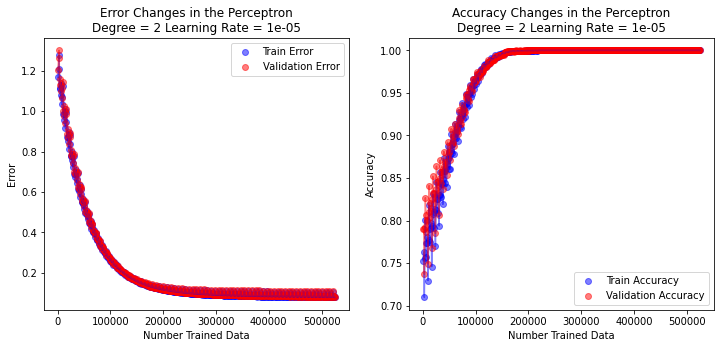
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *نرخ یادگیری* |  |  |  |  |  |
| *صحت مجموعه آموزشی* | *0.221* | *0.749* | *0.622* | *0.778* | *1* |
| *صحت مجموعه اعتبارسنجی* | *0.203* | *0.759* | *0.665* | *0.794* | *1* |





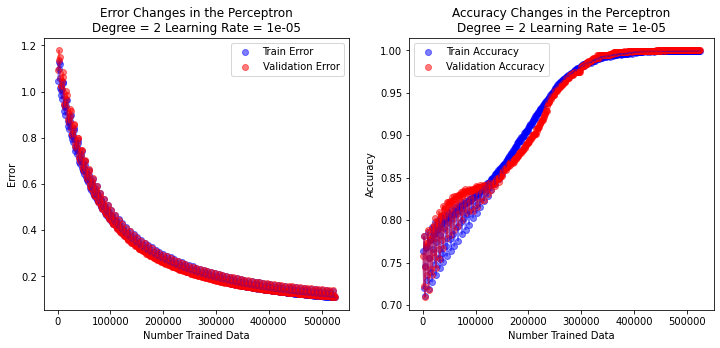


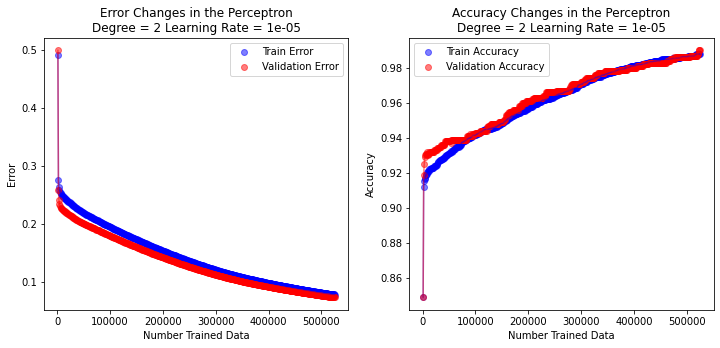


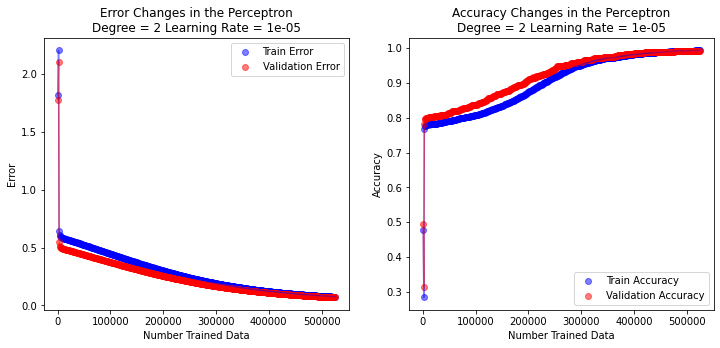


تعداد 75 گام و نرخ یادگیری را برای سه تابع فعال‌سازی همانی، سیگموید و تانژانت هایپربولیک در نظر گرفتیم و نتایج زیر حاصل شد. همگرایی و دقت نهایی هر سه تابع مناسب است ولی به نظرم مانند سوال ۳ در اینجا هم سیگموید کمی متفاوت‌تر از دو تابع دیگر حداقل در یک بار اجرای من بوده است. سیگموید در ابتدا نسبت به دو تابع دیگر سریع‌تر به مقادیر مناسب دقت و خطا رسیده است ولی نهایتا در دقت و خطای نهایی از دو روش دیگر عقب افتاده است.

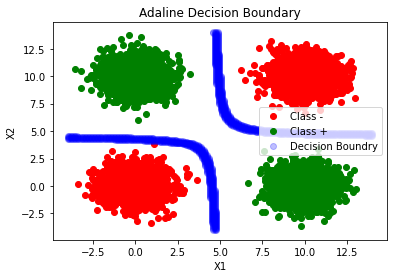
|  |  |  |  |
| --- | --- | --- | --- |
| *تابع فعال‌سازی* | همانی | سیگموید | تانژانت هایپربولیک |
| *صحت مجموعه آموزشی* | *0.999* | *0.987* | *0.993* |
| *صحت مجموعه اعتبارسنجی* | *1* | *0.990* | *0.992* |







آدالاین درجه دو برخلاف آدالاین درجه یک توانست روی مجموعه‌داده XOR‌ به جواب بهینه برسد و از این منظر می‌توان گفت مدل بهتری است. اما باید توجه داشت که این مدل پیچیده‌تر است و زمان اجرای بیشتری دارد کمااینکه به تعداد گام بیشتری هم برای برازش نیاز داشت. در ادامه هم مرز تصمیم آدالاین درجه دو برای یک تنظیم مناسب روی مجموعه‌داده XOR آورده شده است. این درحالی است که پرسپترون درجه دو پیاده‌سازی شده توسط من برای این مجموعه‌داده به دلیل گفته شده برتری چندانی ندارد ولی ممکن است در مجموعه‌های داده دیگر مناسب ظاهر شود.



1. <https://sebastianraschka.com/faq/docs/diff-perceptron-adaline-neuralnet.html> [↑](#footnote-ref-1)
2. <https://datascience.stackexchange.com/q/36368/121138> [↑](#footnote-ref-2)