



گزارش تکلیف دوم درس شبکه‌های عصبی

داریوش حسن‌پور آده

۹۳۰۸۱۶۴

توجه: در کدها نوشته شده داده‌های آموزشی و ارزیابی و تست در داخل یک متغیر به صورت متوالی قرار داده شده‌اند و توسط اندیس‌ها با استفاده از پارامتر $divideParam$ نشان‌گذاری شده‌اند. و همچنین در کلیه قسمت‌های این نوشتار حداکثر تعداد دوره^۱ مقدار ۳۰ در نظر گرفته شده است (در پیروی از تکلیف ۱).

۱ قسمت ۱

کلیه زیرقسمت‌های این قسمت به پیروی از آنچه که در گزارش تکلیف ۱ آمده است شبکه‌ای با ۱ لایه مخفی و با ۱۰۰ نرون در لایه مخفی آزموده شده است.

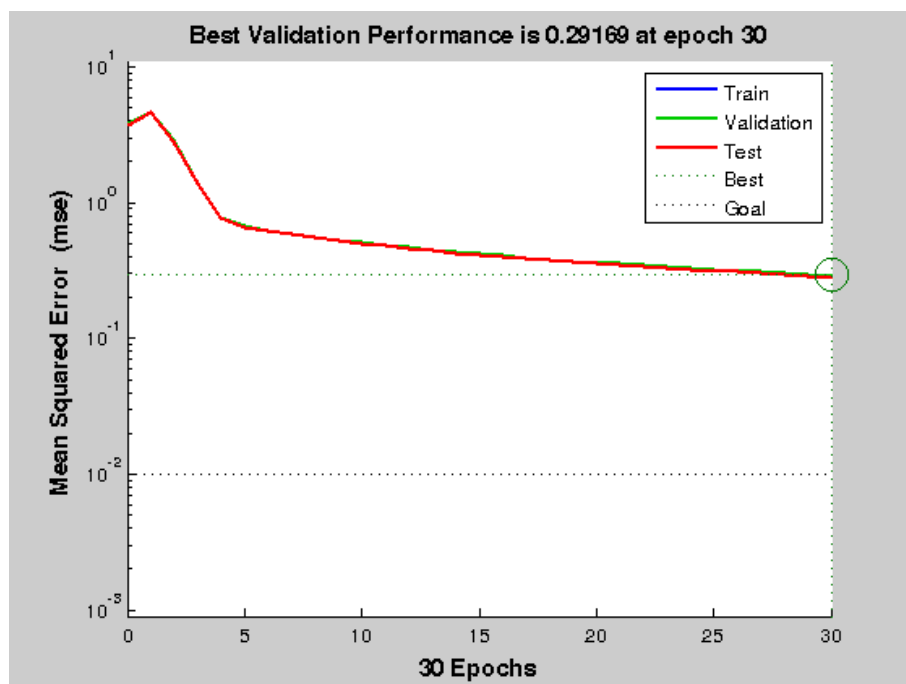
۱.۱ یادگیری دسته‌ای

با استفاده از تابع یادگیری $traingd$ به یادگیری دسته‌ای^۲ داده‌های آموزشی پرداختیم، که بعد از اتمام حداکثر تعداد دوره (۳۰) بهترین مقدار MSE معادل با ۰.۲۹ بدست آمد^۳. که در شکل ۱ نمودار عملکرد شبکه با یادگیری دسته‌ای آورده شده است. که مدت آموزش در این روش ۲.۵ دقیقه شد که بعد از اتمام حداکثر ۳۰ دوره به اتمام رسید.

Epoch^۱

Batch^۲

^۳کلیه نتایج مندرج در این نوشتار چندین بار تست شده و نتایج همیشه در یک بازه از نتایج مندرج در این نوشتار بدست آمده‌اند ولی ما برای نمونه مقدار معمول را در گزارش آورده‌ایم.



شکل ۱: عملکرد شبکه در یادگیری دسته‌ای

۲.۱ یادگیری برخط

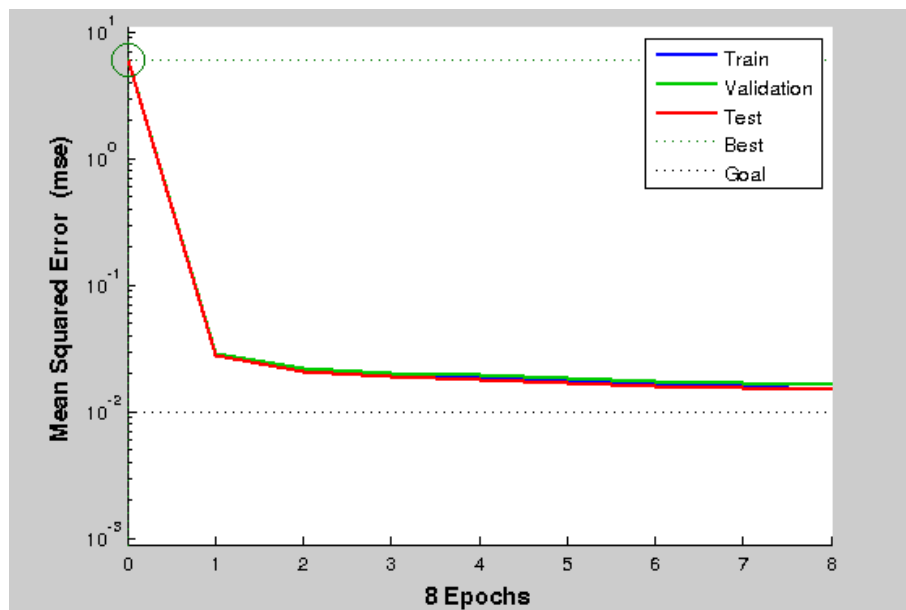
با استفاده از تابع یادگیری *trainc* به یادگیری برخط^۴ داده‌های آموزشی پرداختیم، که بعد از تعداد ۸ دوره^۵ بهترین مقدار *MSE* معادل با ۰.۰۱۵۴ بدست آمد. که در شکل ۲ نمودار عملکرد شبکه با یادگیری برخط آورده شده است. که مدت آموزش در این روش ۴.۲۵ ساعت شد که بعد از اتمام ۸ دوره به اجرای برنامه خاتمه دادیم. در مقایسه‌ی این نتایج این قسمت با قسمت یادگیری دسته‌ای می‌توان گفت که آموزش این قسمت مدت زمان بیشتری نسبت به قسمت اول نیاز داشت و دوم اینکه سرعت همگرایی این قسمت زیادتر از قسمت اولی بود زیرا در این قسمت با این حال بعد از ۸ دوره به صورت دستی آموزش را متوقف کردیم ولی در همین ۸ دوره به میزان کارایی ۰.۰۱۵۴ رسیده بود (یعنی در موقع قطع کردن فقط ۰.۰۰۵ واحد فاصله داشت تا مقدار هدف آن هم با ۲۲ دوره باقی مانده) و همچنین در نمودار *MSE* این روش یادگیری دیده می‌شود که به صورت یک نواخت در هر دوره کاهش پیدا کرده است ولی در قسمت یادگیری دسته‌ای افت و خیز مشاهده می‌شود.

۳.۱ یادگیری دسته‌ای با استفاده از مومنتم

با استفاده از تابع یادگیری *traindm* به یادگیری دسته‌ای با استفاده از مومنتم داده‌های آموزشی پرداختیم، که بعد از تعداد ۷ دوره بهترین مقدار *MSE* معادل با ۰.۲۸۷ بدست آمد که در نهایت بعد از ۶ بازافزایش خطا در داده‌های ارزیاب برنامه به آموزش خود خاتمه داد. که در شکل ۴ نمودار عملکرد شبکه با یادگیری دسته‌ای با استفاده از مومنتم آورده شده است. که مدت آموزش در این روش تقریباً ۲ دقیقه شد که بعد از اتمام ۷ دوره

^۴Online

^۵به علت زیاد بودن مدت آموزش، بعد از ۸ دوره یادگیری را متوقف کردم - ۸ دوره نزدیک به ۴ ساعت زمان برد.



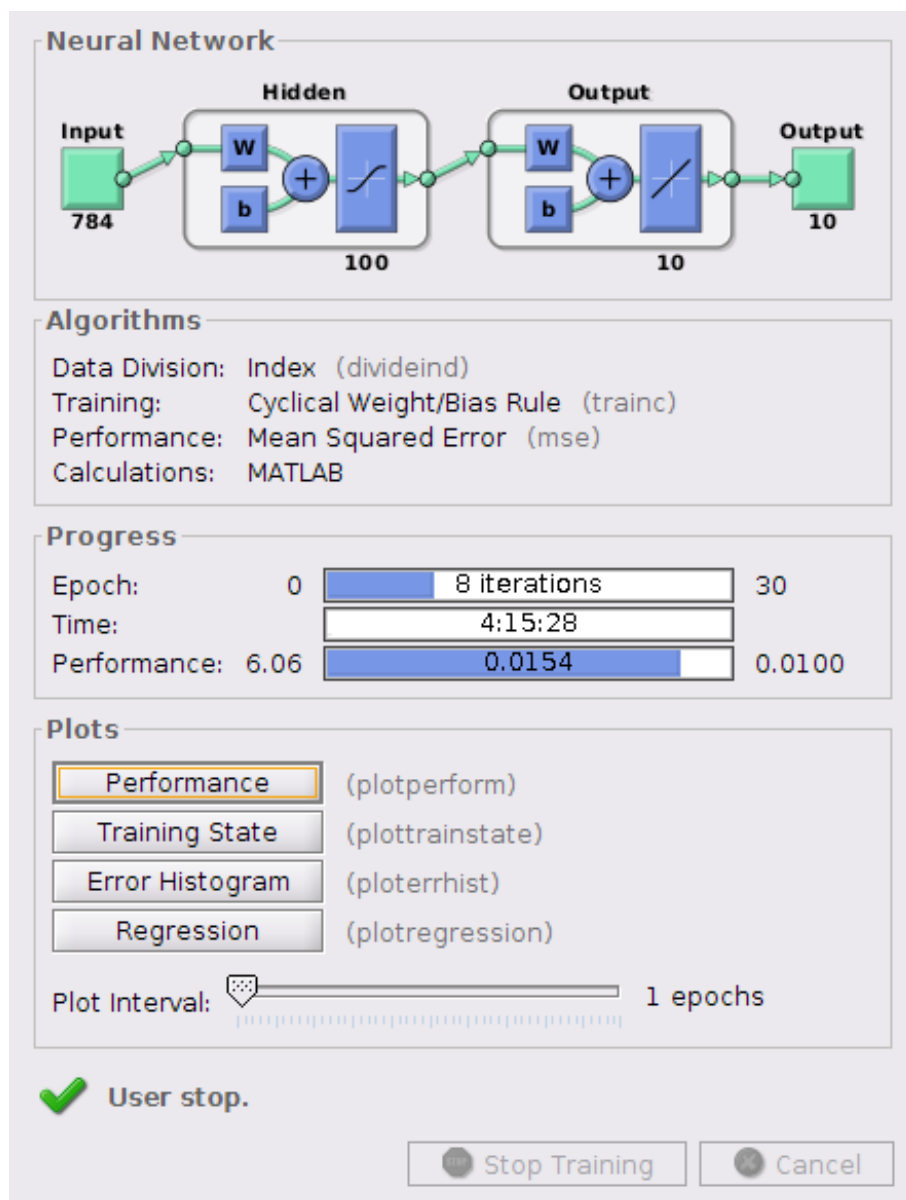
شکل ۲: عملکرد شبکه در یادگیری برخط

برنامه به اجرای خود خاتمه داده است (به علت افزایش خطا در داده‌های ارزیاب). در مقایسه‌ی این نتایج این قسمت با قسمت یادگیری دسته‌ای می‌توان گفت که آموزش این قسمت برعکس قسمت اول ابتدا بهبود یافته ولی بعد از اولین دوره خطای آن افزایش یافته ولی به علت استفاده از مونتیم برنامه به کندی توانست این افزایش در خطا را کند و رو به کاهش نهد به همین علت برنامه بعد از ۶ دوره که توانست خطای خود را بروی داده‌های آموزشی بهتر از بهترین خطای بدست آمده کند (در دوره‌ی ۱) به اجرای خود خاتمه داد ولی با این حال نیز در حدود ۱٪ خطایی بهتر از آنچه که در قسمت ۱ بدست آمد را داشت.

۲ قسمت ۲

برای این منظور تعدادی از شبکه‌های یک لایه، دو لایه و سه لایه را آزمودیم که نتایج هر لایه به صورت نمودارهای در جدول ۲ بدست آمد مقادیر MSE جدول ۲ در شکل ۶ آمده است که به وضوح از بین شبکه‌های موجود بهترین نتیجه را از شبکه با ID برابر با ۱۱ و دومین بهترین نتیجه را در شبکه با ID ۳ بدست آمد که به ترتیب دارای ترکیبات $۷۸۴ \times ۵۰ \times ۲۰ \times ۱۰$ و $۷۸۴ \times ۲۰ \times ۱۰$ می‌باشند - همان‌طور که از نتایج پیدا است تغییرات در ترکیبات شبکه (چه در تعداد لایه و چه در تعداد نورون‌ها) باعث بهبود یا بدتر شدن نتایج می‌شود و از نتایج مشاهده می‌شود که افزایش تعداد نورون‌ها یا تعداد لایه‌ها لزوماً باعث بهبود خروجی شبکه نمی‌شود.^۶

^۶ این پاسخ و جدول ۲ و ۶ جواب هر دو قسمت، قسمت دوم را به صورت یکجا داده است.



شکل ۳: پنجره‌ی اجرایی آموزش شبکه برای یادگیری برخط - بعد از ۴.۲۵ ساعت و ۸ دوره به صورت دستی برنامه متوقف کردیم و عملکردی معادل با ۰.۰۱۵۴ بدست آوردیم.

۳ قسمت ۳

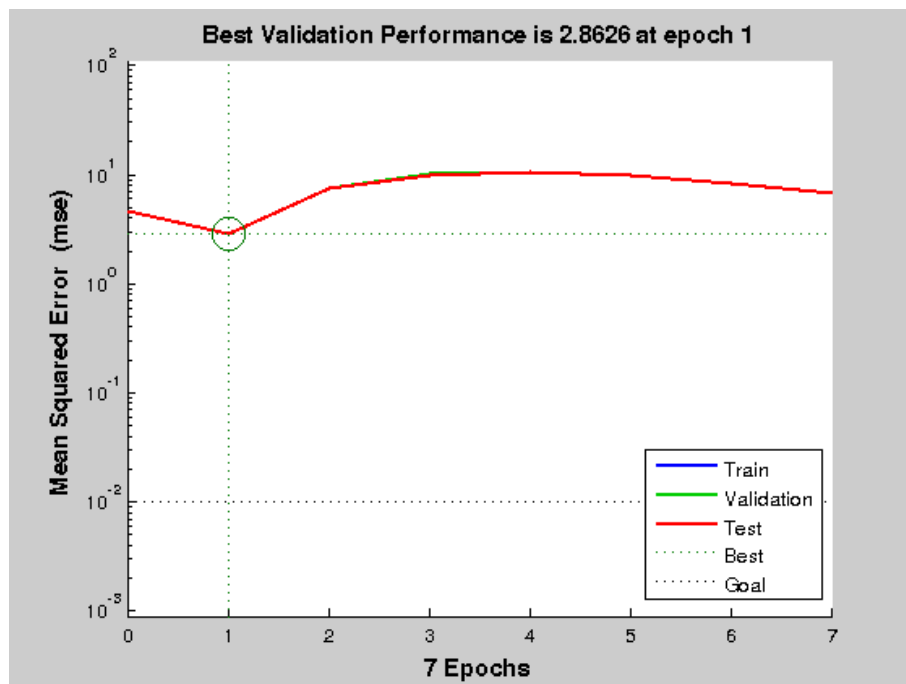
در جدول زیر نتایج روش‌های بهینه‌سازی و استفاده از تابع هزینه *Cross-Entropy* آورده شده است. همان‌طور که مشاهده می‌شود نتایج ردیف‌های ۱ و ۲ نسبت به کلیدی نتایج بدست آمده در این گزارش بهتر بودند ولی با روش‌هایی که از تابع هزینه *Cross-Entropy* استفاده شده است هیچ‌یک نتیجه‌ی مطلوبی نسبت به آنچه که در جدول ۲ آمده است نداشته‌اند. همچنین دو روش بهینه‌سازی *trainbr* & *trainlm* کلاً خطای کمبود حافظه (بروی سیستم با ۸ گیگابایت رم) دادند!!

ID	Layer#	Net. Format	MSE
1	1	784×10	14.98
2	2	$784 \times 10 \times 10$	0.188
3	2	$784 \times 20 \times 10$	0.184
4	2	$784 \times 25 \times 10$	0.204
5	2	$784 \times 50 \times 10$	0.279
6	2	$784 \times 75 \times 10$	0.287
7	2	$784 \times 100 \times 10$	0.291
8	2	$784 \times 150 \times 10$	0.326
9	2	$784 \times 200 \times 10$	0.319
10	2	$784 \times 300 \times 10$	0.268
11	3	$784 \times 50 \times 20 \times 10$	0.165
12	3	$784 \times 200 \times 100 \times 10$	0.223
13	4	$784 \times 50 \times 20 \times 20 \times 10$	0.401

جدول ۱: مقایسه‌ی مقدار MSE در ترکیبات مختلف شبکه

ID	Train Func.	Cross Entropy	Net. Format	MSE
1	trainscg	✗	$784 \times 100 \times 10$	0.075
2	trainrp	✗	$784 \times 100 \times 10$	0.097
3	trainbr	✗	$784 \times 100 \times 10$	Memory Overflow!!
4	trainlm	✗	$784 \times 100 \times 10$	Memory Overflow!!
5	traingd	✓	$784 \times 100 \times 10$	3.604
6	trainscg	✓	$784 \times 100 \times 10$	0.500
7	trainrp	✓	$784 \times 100 \times 10$	4.130

جدول ۲: مقایسه‌ی مقدار MSE در ترکیبات مختلف شبکه

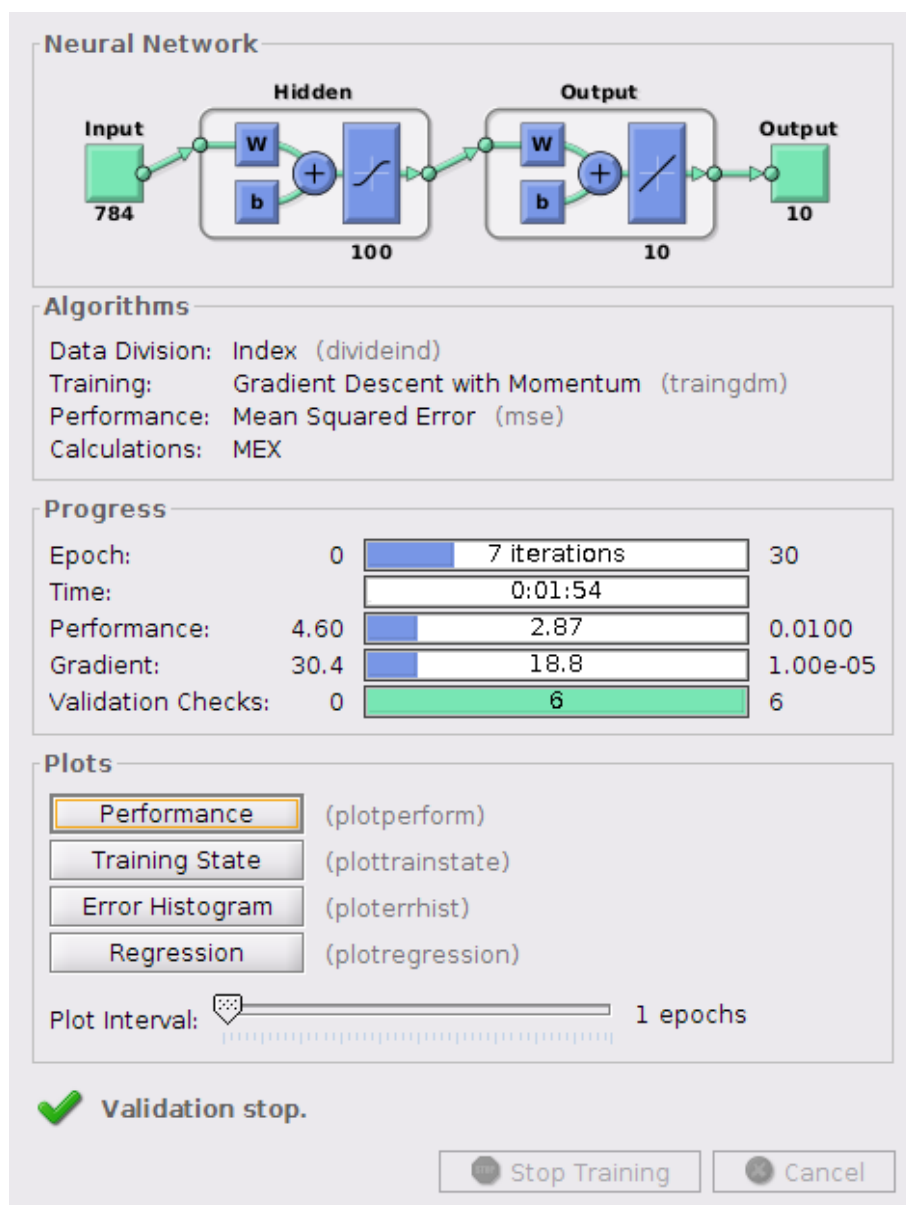


شکل ۴: عملکرد شبکه در یادگیری برخط

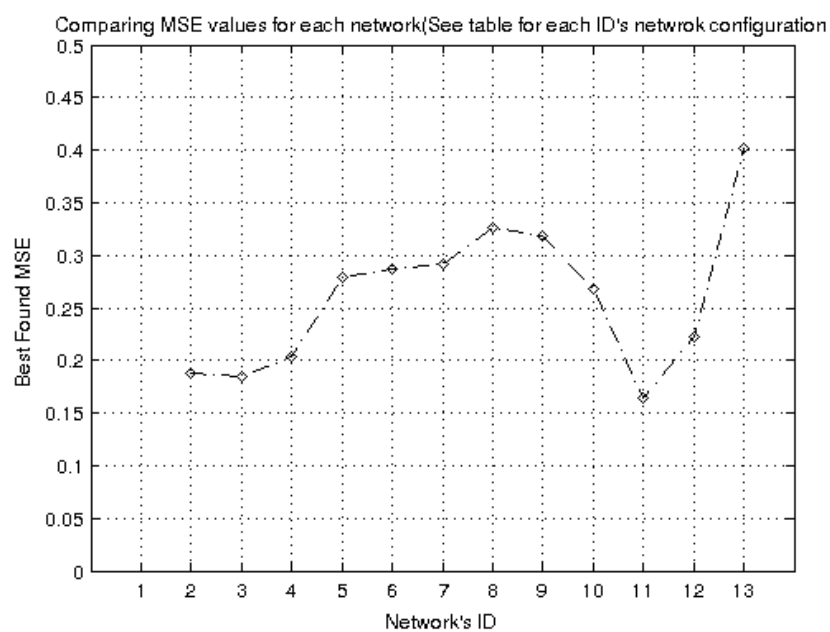
۴ جمع‌بندی

در این گزارش عملکرد روش‌های مختلف و توابع هزینه‌ی مختلف مورد بررسی قرار گرفت و در طی آزمایشات متوجه شدیم که افزایش لایه‌ها و/یا تعداد نوروها همیشه باعث بهتر شدن نتایج نمی‌شود و برای دیتاست MNIST بهترین نتیجه‌ای که توانستیم به ازای ۳۰ دوره‌ی آموزشی بدست بیاوریم با استفاده از تابع بهینه‌سازی *trainseq* با ترکیب $10 \times 100 \times 784$ (جدول ۲) می‌باشد.

توجه شود که در این تکلیف به پیروی برخی از پارامترهایی که در اجرای تکلیف اول در نظر گرفته شد بود (از جمله حداکثر تعداد دوره‌ها و همچنین ترکیب شبکه) در سراسر گزارش (به جز قسمت دوم که ماهیت سوال ایجاب می‌کرد) از آن پارامترها استفاده شده است. همچنین کلیه‌ی نتایج آزمایشات در پوشه‌ی *logs* به صورت دسته‌بندی شده به ازای هر سوال همراه کد آمده است که در صورت تمایل می‌توانید آن‌ها را هم چک کنید. در مورد توضیح در مورد کد هم که خیلی واضح است - داده‌ها لود می‌شود و سپس یک شبکه‌ی *feedforward* بر اساس داده‌ها و فرمت لایه‌های پنهان ایجاد شده و سپس با مشخص کردن نوع تابع بهینه‌سازی و تابع هزینه به آموزش شبکه می‌پردازیم.



شکل ۵: پنجره‌ی اجرایی آموزش شبکه برای یادگیری دسته‌ای با استفاده از مومنتم - همانطور که مشاهده می‌شود برنامه بعد از مواجه با ۶ دوره‌ی متوالی در افزایش خطا بروی داده‌های ارزیاب به اجرای خود خاتمه داده است.



شکل ۶: مقایسه‌ی مقادیر MSE شبکه‌های جدول ۲ (بر اساس ID) به صورت نمودار - به منظور مقایسه‌ی بهتر MSE شبکه با ID با مقدار ۱ که بدترین MSE را دارد در نظر نگرفته شده است.