

گزارش

سارا حسینی ۴۰۰۲۲۲۰۲۶

۱۹ فروردین ۱۴۰۳

۱ مقدمه

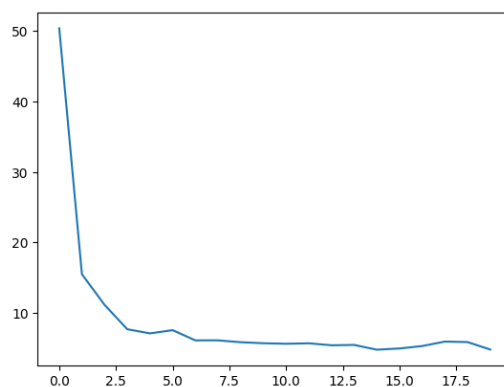
در این مقاله، ما یک بکه عصبی mlp برای تسک رگرسیون تشخیص عملکرد دانش آموزان را پیاده سازی میکنیم و بین هایپرپارامترهای موجود مثل معماری شبکه، اکتیویشن فانکشن مختلف، رگولاریزیشن مختلف و دراپ اوت مختلف، بهترین را پیدا خواهیم کرد.

۲ preprocessing

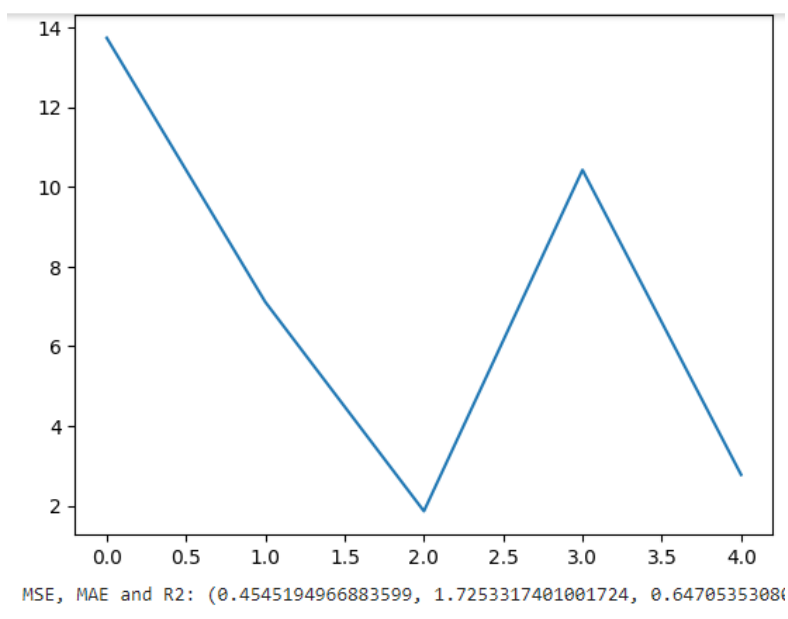
برای این کار، ابتدا eda مختصری روی داده داریم. دیتاست ۳ ستون دارد. هیچ داده تکراری در سطرها ندارد. تمامی ستون ها بجز تعداد غیبت های دانش آموز، کتگوریکال هستند. هیچ جا هم میسینگ و لیو نداریم. ابتدا دیتا را با نسبت ۸:۲ به دو قسمت تقسیم میکنیم برای تست و ترین. سپس اعمال بعدی را جداگانه روی هر بخش اعمال میکنیم. ابتدا داده های باینری را به صورت ۰ و ۱ انکد کردیم. سپس "شغل پدر"، "شغل مادر" "انگیزه" و "سرپرست" را که بیش از دو مقدار داشتند، وان هات انکد کردیم. متغیرهای 'age', 'Medu', 'Fedu', 'traveltime', 'studytime', 'failures', 'famrel', 'freetime', 'goout', 'Dalc', 'Walac', 'health', 'absences' که همگی عددی بودند را نیز با minmax scalar اسکیل کردیم.

۳ مدل

ستون G3 که نمره نهایی فرد بود را بعنوان تارگت از فیچرها جدا کردیم. توزیع داده های تارگت را هم رسم کردیم که تقریباً مشابه نورمال بود. اولین ام ال پی که ساختیم، ۳ لایه دارد با اکتیویشن رلو و دراپ اوت بعد از هر لایه با نرخ 0.2. ۲۰ اپاک ترین کردیم و به برابر 4.7 روی ترین و

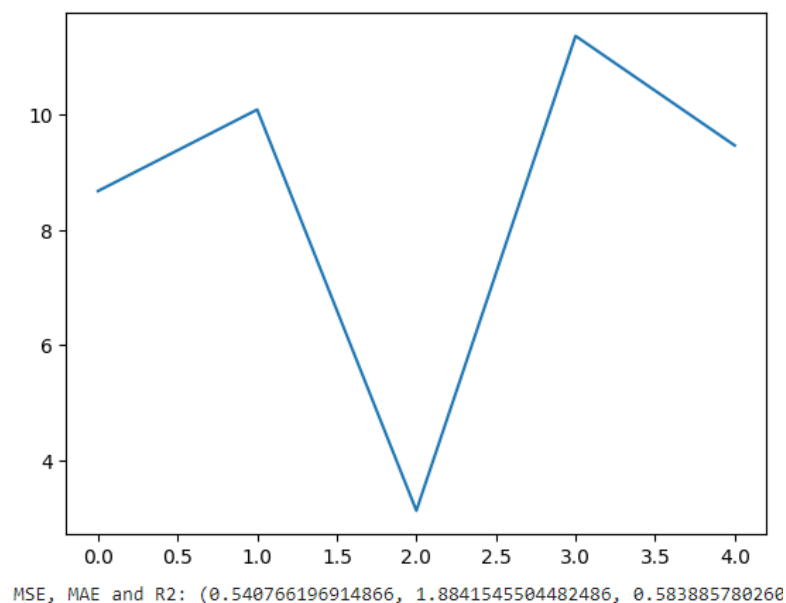


شکل ۱: خطا روی ترین



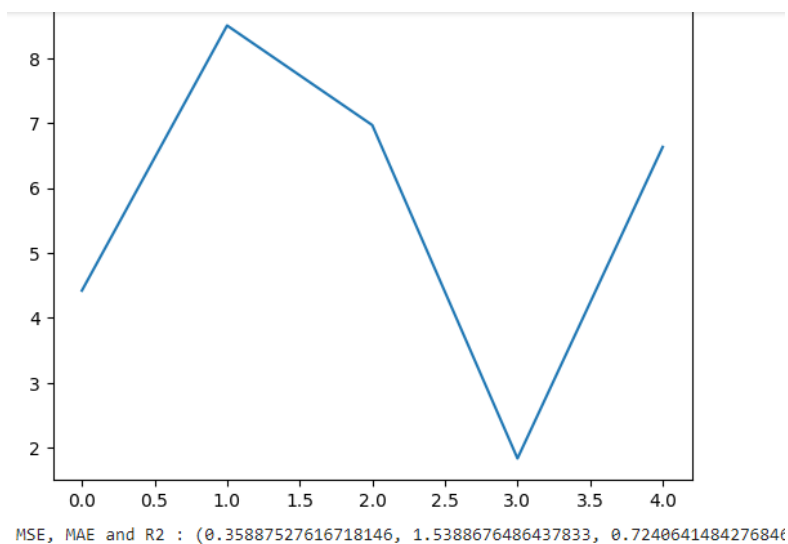
شکل ۲: خطا روی تست

همین مدل را باری دیگر بدون دراپ اوت ترین کردیم و خطای ترین به 1.3 رسید. اما نتایج روی تست بهتر نشد.



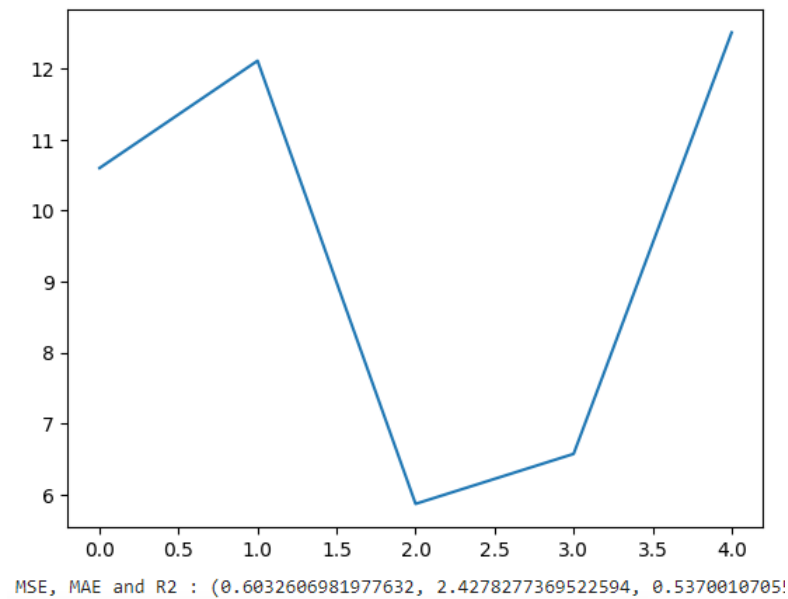
شکل ۳: خطای تست

در مرحله بعد، شبکه دومی با دو لایه طراحی کردیم. آن را نیز ۲۰ اپیاک و با همان در اوت آموزش دادیم و نتیجه خطای ترین ۵.۴ شد.



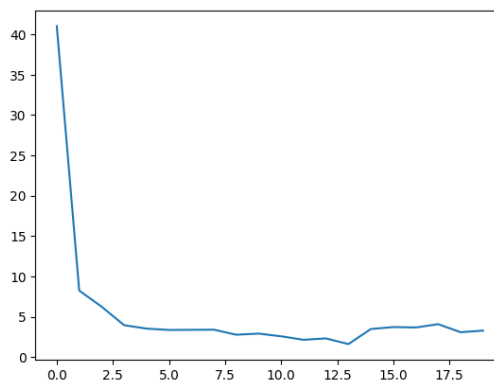
شکل ۴: خطای تست شبکه دوم

با این وجود، مقدار خطای تست از هر دو مدل قبلی کمتر شد و R2 بسیار خوبی بدست



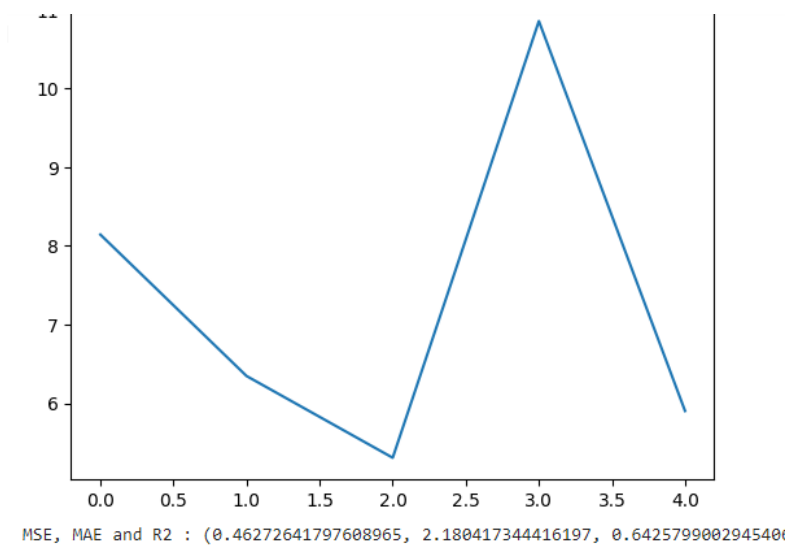
شکل ۵: خطای تست

آوردیم یعنی ۷.۰. در مرحله بعد، شبکه‌ای با ۶ لایه طراحی کردیم. این شبکه نیز خطای ۷.۲ روی ترین داشت و روی تست به صورت زیر عمل کرد: بنظر میرسد افزودن لایه‌ها تاثیر خوبی نداشته است. حالا همین شبکه را بدون دراپ‌اوت امتحان میکنیم. خطای ترین زیاد شد و به ۲.۳ رسید



شکل ۶: خطای ترین

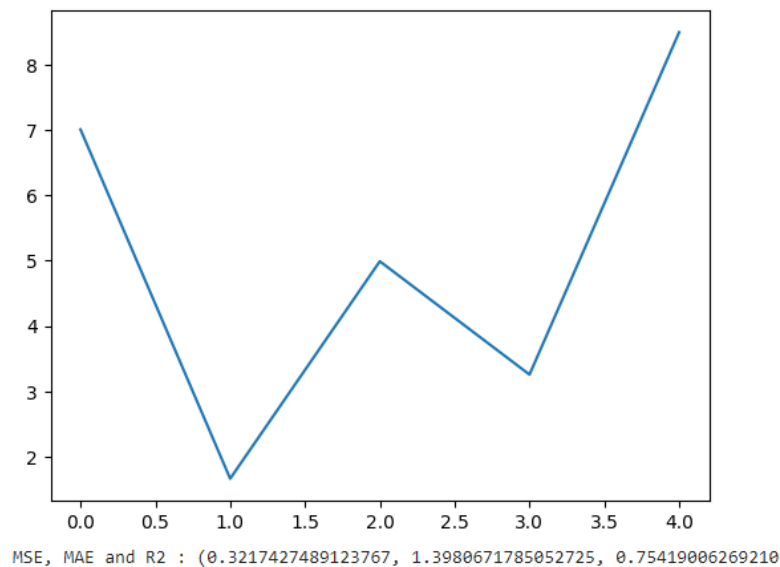
شیب تند این تابع نشان میدهد که به سرعت بهتری به جواب رسیده‌ای نسبت به حالت با دراپ‌اوت.



شکل ۷: خطای تست بدون دراپ‌اوت کمتر و مربع آر بهتر شده است

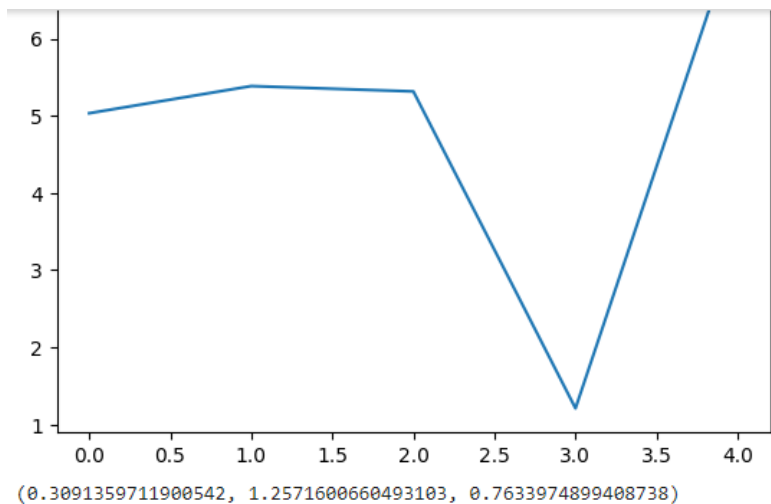
سپس شبکه را با و بدون رگولاریزیشن L1 آموزش دادیم. بدون آن، مربع آر برابر 0.66 و با این تکنیک و با یک ضریب ۵.۰ این معیار به 0.69 رسید. البته نکته این بود که وقتی ضریب ۰.۱۰ بود، مربع آر حدود ۵۵.۰ میشد. یعنی مقدار بهتر برای ضریب رگولاریزر ۵.۹ بوده است. در ۱۰۰ اپیاک، با و بدون L1 شبکه را ترین کردیم با ضریب رگولاریزیشن ۵.۰ اما نتایج فرقی نکرد.

در مراحل بعد، I2 را هم پیاده‌سازی کردیم با ۱۰۰ اپیاک و ضریب ۱.۰ اما نتیجه آر ۲ برابر ۶.۰ شد. اما وقتی هم از L1 با ضریب ۱.۰ و هم I2 با ضریب ۵.۰ استفاده کردیم، نتایج بهبود یافتند و به R2 برابر با ۴۵.۰ رسیدیم با ۱۰۰ اپیاک.



شکل ۸: الستیک نت

تابع دیگری که برای آموزش طراحی کردیم، از یک RandomWeightSampler استفاده میکرد و دیتا را موقع ترین، با احتمالی به مدل میداد که هر داده‌ای را مدل بدتر پیشبینی کرده باشد، با احتمال بالاتری در سمپل انتخاب میشد. این تکنیک نیز باعث بهبود مدل و R2 برابر با ۷۶.۰ شد.



شکل ۹: استفاده از سمپلر خاص

در انتها نیز تابع فعالیت دیگر را نیز با شبکه آموزش دادیم. بهترین نتیجه از بین اینها تابع

سیگموید بود. حتی بهتر از SELU و Tanh.