

گزارش پیک نوروزی

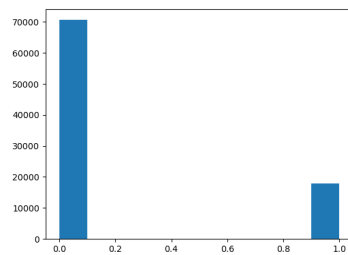
سارا حسینی ۴۰۰۲۲۲۰۲۶

۱ مقدمه

در این گزارش ما روی دیتاست مربوط به ملاقات‌های بیماران با پزشکان، تحلیل و بررسی انجام دادیم. ما یک شبکه عصبی چندلایه را برای تسک باینری کلسیفیکیشن ترین کردیم. همچنین با تغییر هایپرپارامترها و چند تکنیک رگولاریزیشن مثل دراپ اوت یا L1 دقت آن را بررسی کردیم.

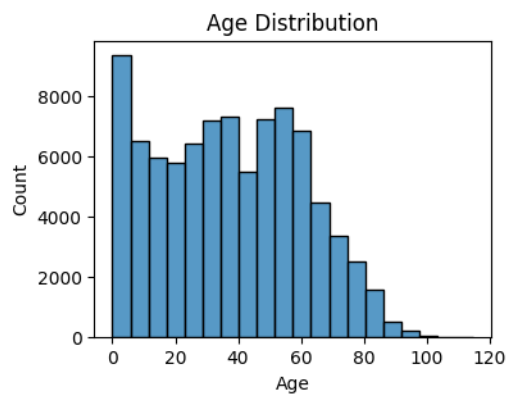
۲ preprocessing

دیتا دارای ۱۱۰ هزار سطر و ۱۴ ستون بود. همچنین هیچ‌سطر تکراری یا مقدار null هم نداشت. ستون آیدی قرار ملاقات با پزشک، برای هر سطر یونیک بود و اطلاعاتی به ما نمیداد پس همانجا حذفش کردیم. متغیرهای مرتبط به بیماریها را به صورت ۰ و ۱ داشتیم و ۲ متغیر تاریخ داشتیم. یکی زمان وقت گرفتن از دکتر (زمانی که بیمار زنگ زده و وقت گرفته) و دیگری زمان قرار ملاقات با دکتر. ما قبل از انجام هر کاری روی دیتا، برای جلوگیری از نشت ترین به تست، دیتا را با نسبت ۸ به ۲ تقسیم کردیم. سپس متغیرهای باینری که به صورت "زن و مرد" یا "بله و خیر" بودند را به عدد درآوردیم یعنی ۰ و ۱. از ستون زمان ملاقات، ساعت ملاقات، سال، ماه و روز را استخراج کردیم. سال و ماه و روز را از ستون زمان وقت گرفتن هم استخراج و در ستونهای جدید ریختیم اما ساعت وقت گرفتن در دیتا مشخص نبود (کلا صفر بود). یک متغیر جدید هم بعنوان فاصله زمانی وقت گرفتن تا زمان ملاقات در واحد روز، استخراج کردیم. نهایتا هم دو ستون اولیه که همه اینها از انان استخراج شد را حذف کردیم. زیرا به فرم پیچیده‌ای ذخیره شده بود و مناسب شبکه نبود.



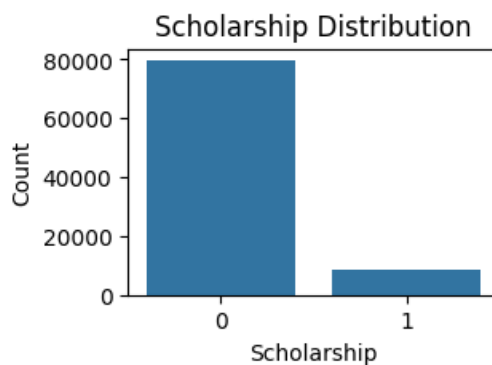
شکل ۱: نمودار توزیع متغیر "حاضر شدن در ملاقات"

شکل ۱ نشان می‌دهد که متغیر تارگت بطور غیر متوازن در دیتا توزیع شده است.

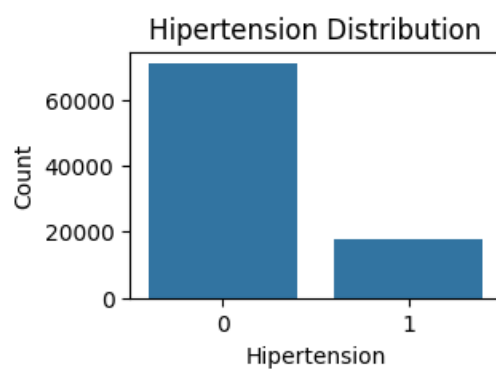


شکل ۲: توزیع متغیر سن

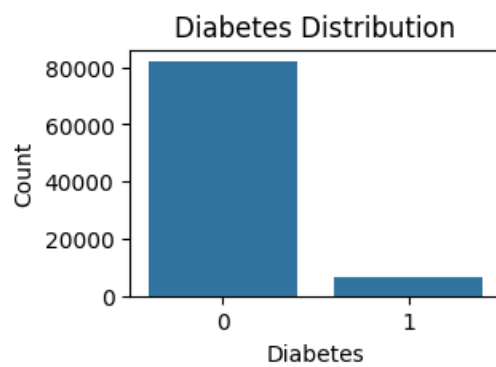
متغیر سن چولگی راست دارد اما مشخص است داده اوتلایر در بازه بالای ۱۰۰ سال هم داشته‌ایم.



شکل ۳: بیشتر افراد scholarship نداشته‌اند

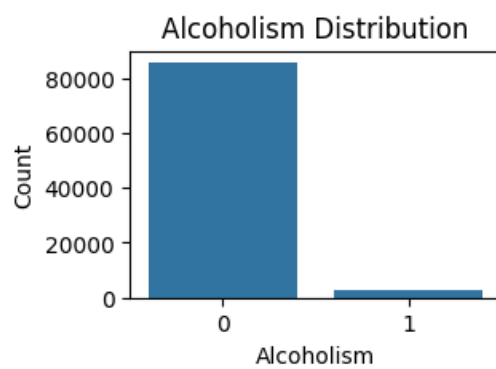


شکل ۴: توزیع فشار خون

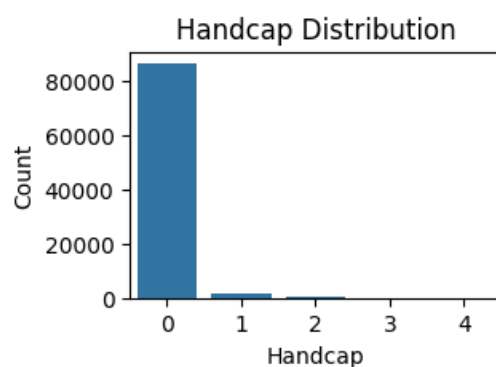


شکل ۵: توزیع دیابت

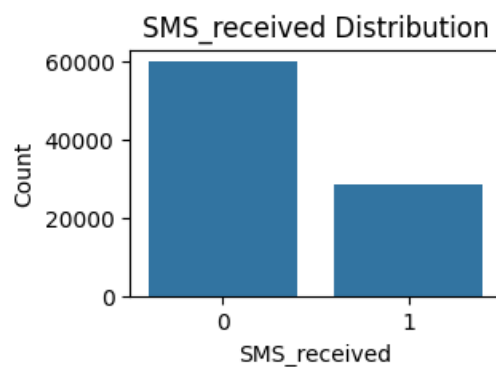
همانطور که در نمودار مشاهده میشود بیشتر افراد بیماری زمینه‌ای داشته‌اند.



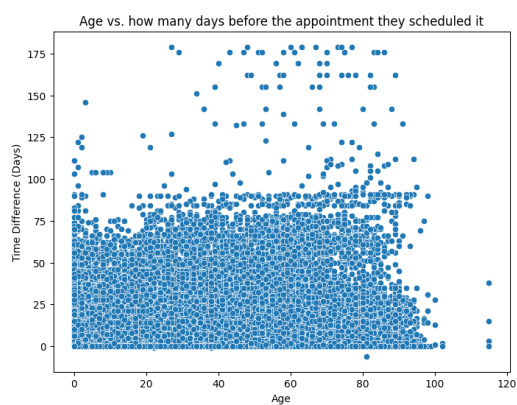
شکل ۶: توزیع الکل خواری



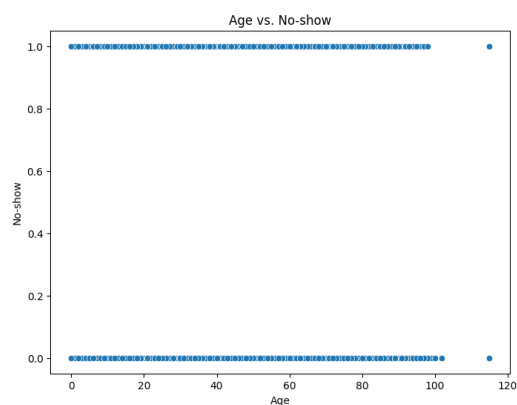
شکل ۷: درجه معلولیت



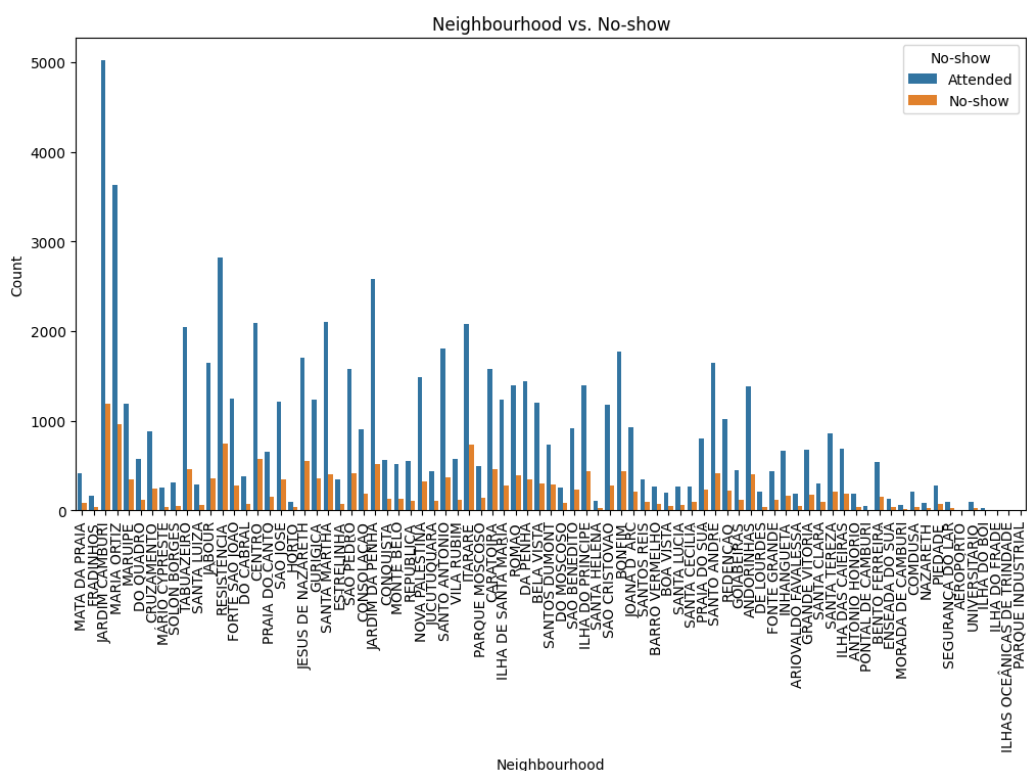
شکل ۸: اینکه افراد قبل وقتشان پیامک دریافت کردند یا نه



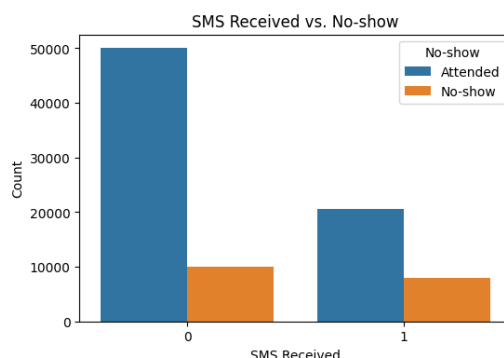
شکل ۹: متغیر سن در برابر متغیر "فاصله بین وقت گرفتن تا قرار ملاقات" رابطه خاصی مشاهده نمیشود.



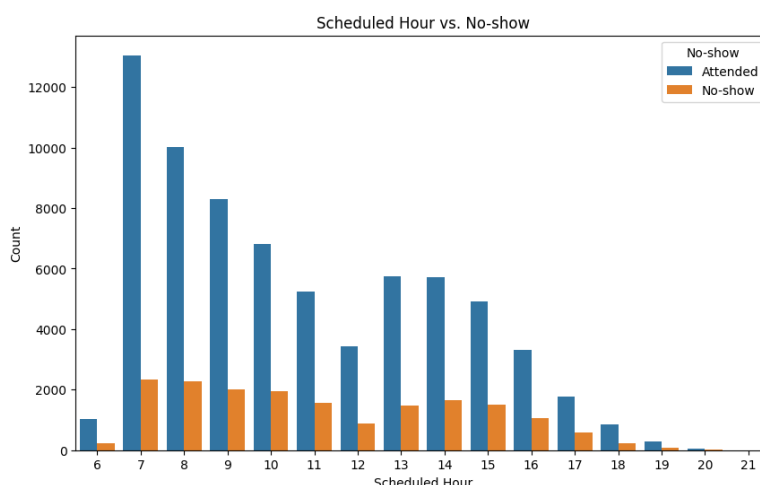
شکل ۱۰: سن در برابر متغیر "حاضر نشدن در مطب". رابطه خاصی نمی‌بینیم.



شکل ۱۱: رابطه بین محل زندگی و در مطب حاضر نشدن. در جفت میله‌های سوم و چهارم از چپ، ستون no show بلندتر است اما تعداد ملاقات‌های رزرو شده بطور کلی در این دو محل بیشتر است پس نمیتوان نتیجه ای گرفت.



شکل ۱۲: اینجا درصد بیشتری از کسانی که پیامک نگرفته‌اند، در مطب حضور یافته‌اند. این بر خلاف انتظار ماست چون انگار دریافت پیامک تاثیری بر حضورشان نگذاشته.

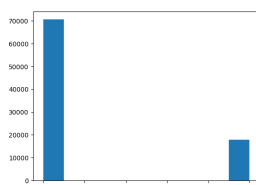


شکل ۱۳: رابطه ساعت مقرر شده ملاقات با حضور یافتن. هر چه به ساعات پایانی روز نزدیک میشویم، احتمال حضور نیافتن بیمار بالا میرود.

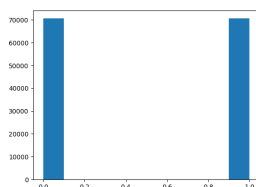
در ادامه، به انکد و سپس اسکیل دیتای ترین و دیتای تست پرداختیم. ستون محل زندگی neighbourhood را one hot encode کردیم که با توجه به داشتن ۸۱ مقدار یونیک، ۸۱ ستون به دیتاست افزود. در اینجا همان انکدوری که روی ترین فیت شده، تست را نیز ترنسفرم میکند و اگر هم مقدار جدیدی در تست بود، کلا صفر میگذارد. در مرحله بعد، با استفاده از minmaxscaler اکثر مقادیر عددی را اسکیل کردیم. نهایتاً دیتاست ما صرفاً شامل عددهایی اسکیل شده است.

۳ مدل

ابتدا تارگت و فیچر را جدا کردیم. برای حل مشکل غیر بالانس بودن دیتاست، با استفاده از SMOTE تارگت خود را اورسمل کردیم.



شکل ۱۴: دیتاست در ابتدا



شکل ۱۵: بعد از اورسمل

با بررسی‌ای که در مورد انواع معماری‌های شبکه انجام دادیم، به این نتیجه رسیدیم که شبکه‌های عریض برای یادگیری ارتباطات مستقیم‌تر و ساده‌تر بین فیچرها و الگوهای کلی مناسبند. همانطور از universal approximation theorem میدانیم، با افزودن تعداد نوروں در هر لایه می‌توانیم توابع پیچیده‌تری را مدل کنیم. مثلاً برای نشان دادن یک تابع n متغیره جبر بولی، متشکل از ترکیب DNF ای از n ورودی، در بدترین حالت که تابعمان یک XOR n متغیره باشد، به 2^{n-1} نوروں نیاز داریم تا با یک شبکه‌ی دو لایه‌ای در عریض‌ترین حالت ممکن، تابع را نشان دهیم. اما اگر تعداد لایه‌ها را محدود نکنیم، می‌توانیم دوه دو از متغیرها XOR بگیریم و در لایه‌ی بعد همینطور ادامه دهیم و با تعداد کمتری نوروں و در $O(n \lg n)$ لایه همان تابع را نشان دهیم.

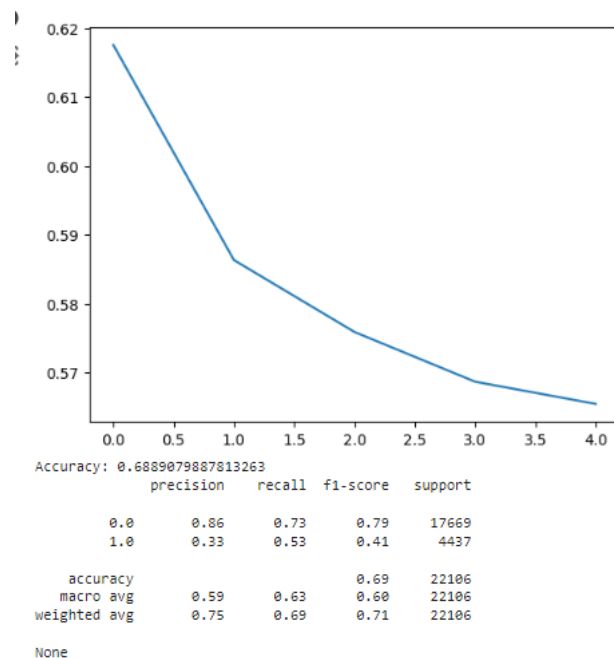
افزودن لایه باعث میشود سریعتر و با محاسبات کمتر بتوانیم همان روابط پیچیده و انتزاعی را یاد بگیریم. افزودن نوروں به هر لایه نیز میتواند باعث بهت شدن یادگیری بشود اما احتمال اورفیت را هم بالا میبرد.

همچنین با بررسی مکانیسم اتشن، فهمیدیم که استفاده از آن، میتواند به شبکه کمک کند بر روی ورودی خاصی بیشتر متمرکز شود و اینگونه سبب بالا بردن دقت مدل و یادگیری شود اما این مکانیسم بیشتر برای دیتای متنی کاربرد دارد پس ما در اینجا از آن استفاده نکردیم. در ادامه ام ال پی خودمان را با ساختارهای عمیق و عریض پیاده کردیم تا ببینیم چه تاثیری روی

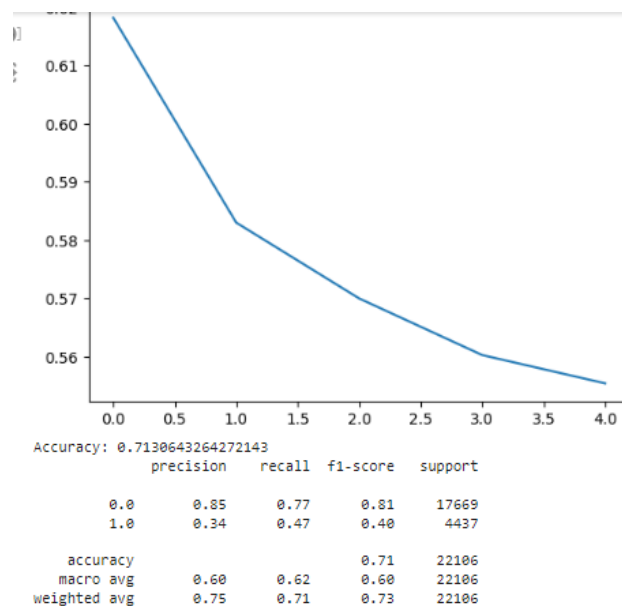
مدل دارند. همچنین مدل را با وزن‌دهی Xavier که باعث همگرایی بهت می‌شود ساختیم و رگولاریزیشن‌های ال ۱، ال ۲ و الاستیک نت هم آزمایش کردیم.

۴ عمق و عرض شبکه

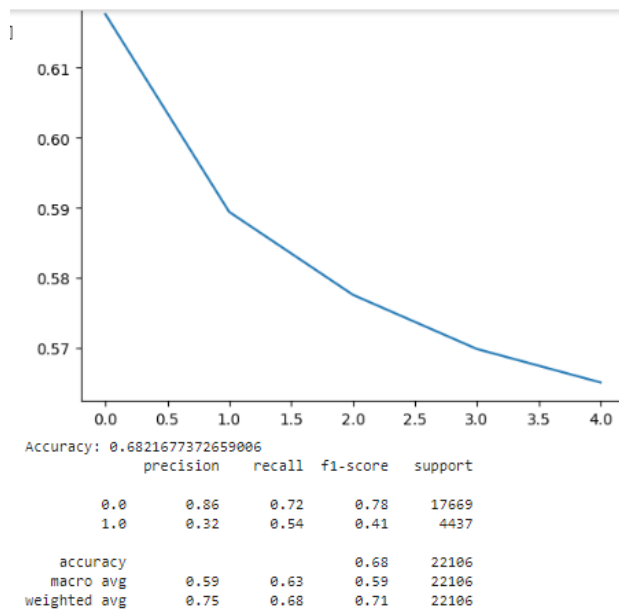
شبکه ۱ ما یک شبکه فولی کانکتد با ۳ لایه و یک اکتیویشن سیگموید در لایه آخر است که به ما خروجی احتمال می‌دهد. عرض هر لایه ۶۴ نورون لرنینگ ریت ۰.۱ و با ۵ اپاک ترین شده است.



سپس، یک لایه کم کردیم و با ۲ لایه آن را اجرا کردیم که نتیجه اکیورسی در طی چندین اجرا، ۷۰٪ شد:

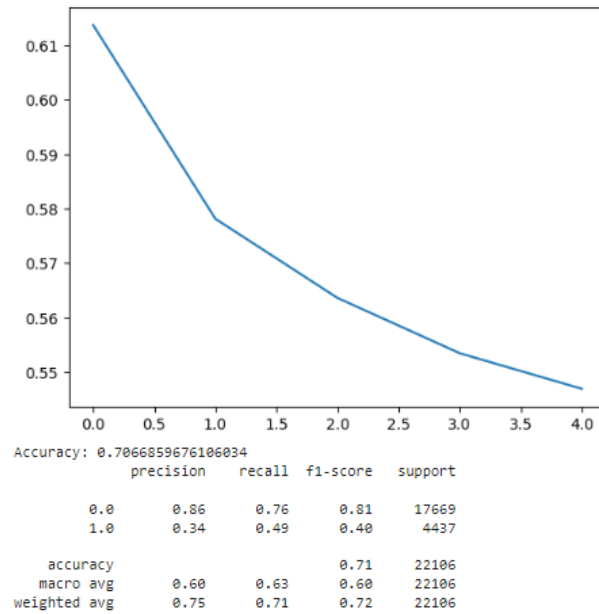


یعنی با این شرایط موجود، تعداد لایه کمتر بهتر جواب داده است. وقتی شبکه را با ۴ لایه ترین کردیم نیز دقت بار اول ۷۱.۰ و بار دوم ۶۹.۰ و بار سوم ۶۸.۰ شد.



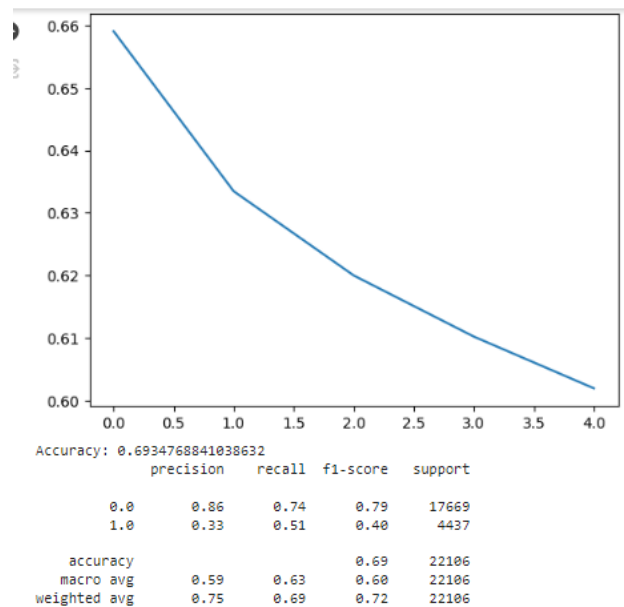
پس شبکه دولایه‌ای را بعنوان شبکه بهتر، برداشتیم. در ادامه، شبکه را با تعداد نورون هیدن ۱۲۰ ترین کردیم و دقت روی تست در اجرای اول به ۶۹.۰، در اجرای دوم به ۷۰.۰ و در اجرای سوم به ۶۹.۰ رسید. با تعداد نورون ۵۰۰، دقت روی تست به ۷۰.۰ رسید. به نظر میرسد افزودن

نورون کمک چندانى نمیکند.

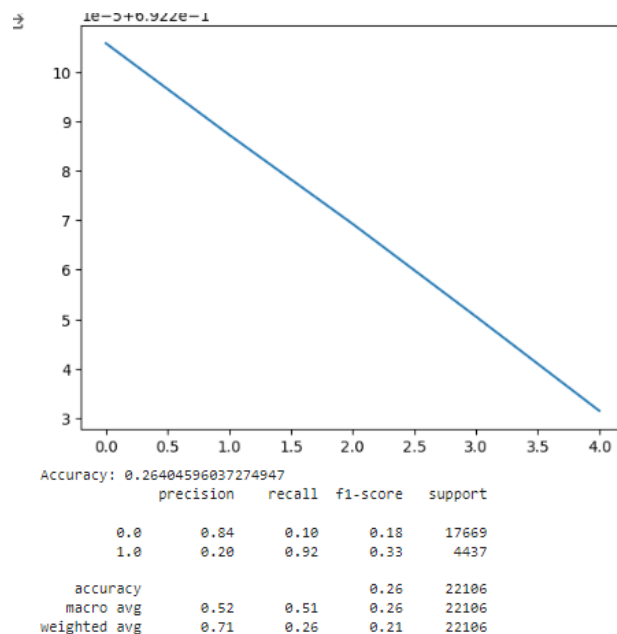


۵ لرنینگ ریت

شبکه ۲ لایه‌ای را با لرنینگ ریت کمتری یعنی ۰.۰۰۱۰۰ آزمودیم و نتیجه دقت ۶۹ شد. اما مشخص است که تابع هزینه با شیب کمتری به اپتیمم حرکت کرده است و همواری آن کمتر شده است.

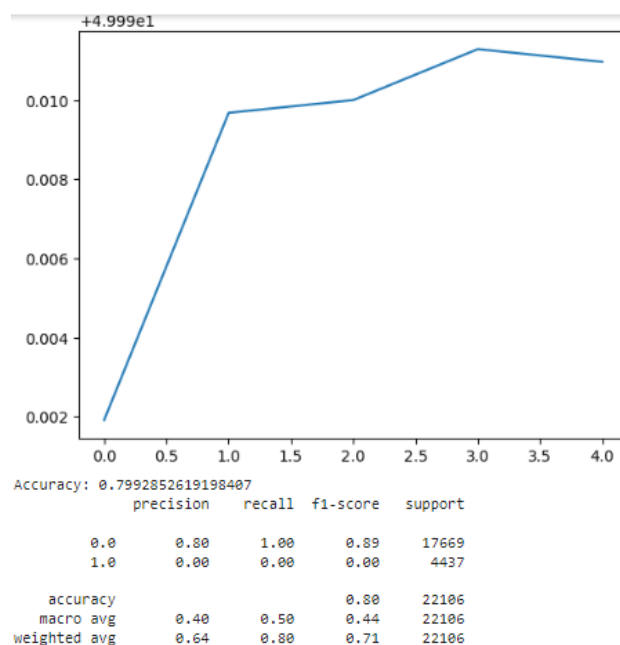


اما لرنینگ ریت 10^{-7} باعث شد دقت مدل به طور زیادی پایین بیاید و به ۲۶.۰ برسد. چرا که قدرت اکسپلوریشن را از مدل گرفته است و گرچه از آدام استفاده میکنیم، همچنان نتیجه بد شده است.



لرنینگ ریت ۵.۰ هم باعث شد که دقت به ۷۹ برسد اما این به دلیل اورفیت شدن است چرا که precision, recall, f1 روی کلاس ۰، به ۰ رسیده‌اند. در تمام مدلها این معیارها روی کلاس

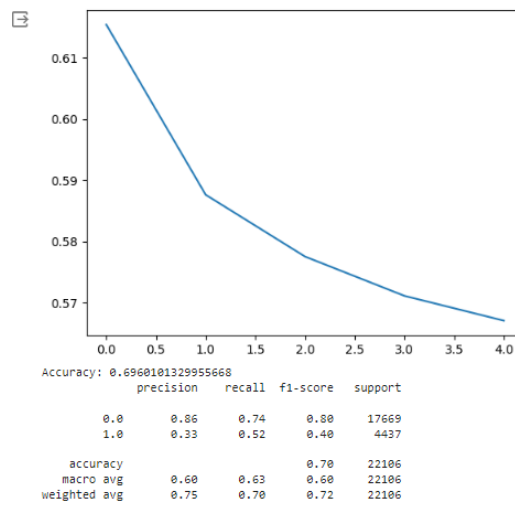
۱، بالا هستند و این هم به دلیل ایمبالانس کلاس هاست.



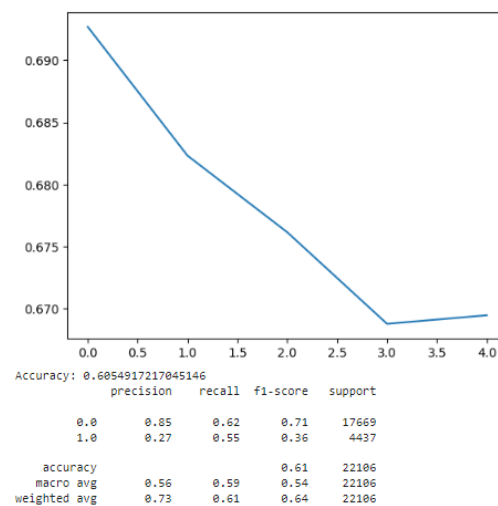
پس لرنینگ ریت ۰.۱۰۰ بهترین ریت بوده است.

۶ سائز بچ

بچ ما در تمام مراحل قبل برابر ۳۲ سمپل بود. اینجا ابتدا دیتای ترین را با بچ سائز ۱۶ ترین کردیم که دقت کمی پایین آمد اما recall کمی بالا رفت. سپس با سائز ۸ شبکه را ترین کردیم. اینجا هم ریکال بالا رفته است.



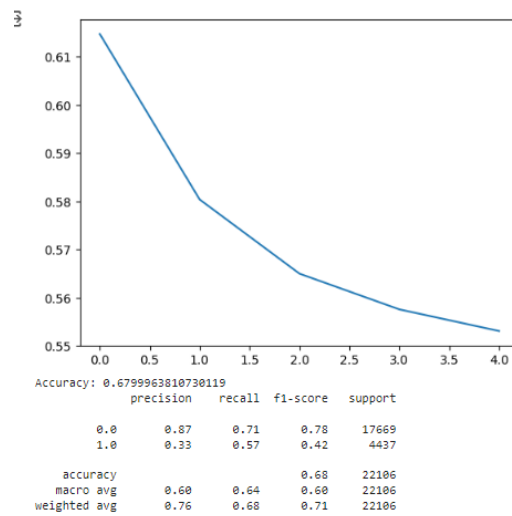
در آخر شبکه را با سایز بچ برابر با کل ترینینگ دیتا ترین کردیم و دقت به ۶۰ رسید.



اینجا ما گرادیان دیسنت کامل داریم و احتمالاً در یک لوکال مینیمم گیر افتاده است و به همین دلیل لاس روی ترین هم بالاتر رفته است. همچنین دقت هم پایین آمده است.

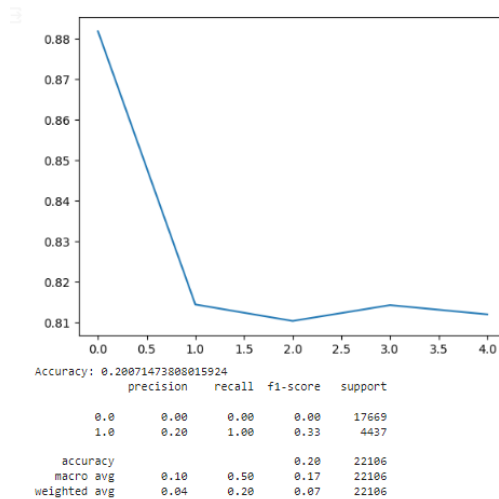
۷ وزن دهی اولیه

مدل با وزندهی xavier مدل را آموزش دادیم.

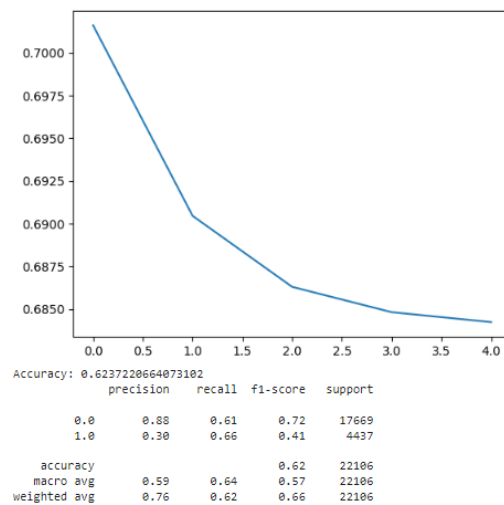


روی دیتای ترین، loss کمتر شد اما اکيورسی پايينتر آمد. اين ميتواند به دليل استفاده ما از رلو باشد. بار ديگر با اکتیویشن سیگموید شبکه را ترین کردیم و نتیجه اکيورسی ۶۹ شد. تکنیک xavier مناسب اکتیویشن‌هایی است که حول صفر متقارن هستند مثل سیگموید یا tanh مناسب است.

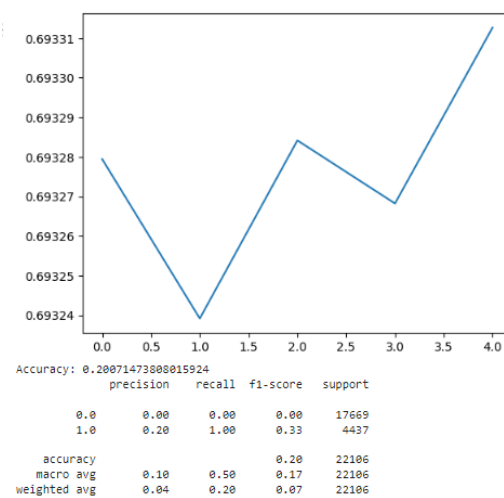
۸ رگولاریزیشن و دراپ اوت



شکل ۱۶: نتایج شبکه با ضریب ال ۱ برابر با 0.1. چون سختگیرانه می‌خواهد وزن‌ها کوچک باشند، شبکه تقریباً چیزی یاد نگرفته

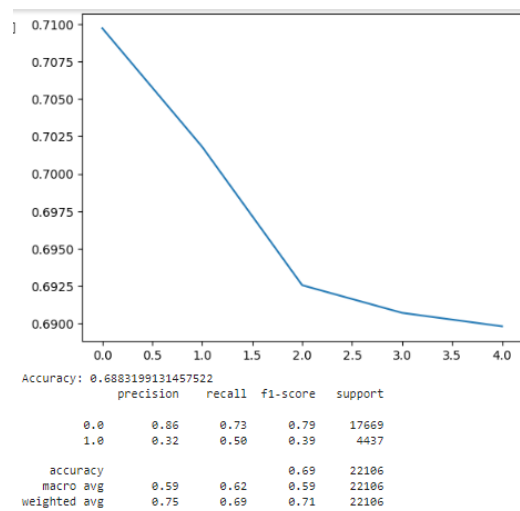


شکل ۱۷: با ضریب ال ۱ برابر با 0.01 بسیار بهتر یاد گرفته اما همچنین از حالت بدون رگولاریز بدتر است. چون شبکه ما خیلی ساده است و کلا وزنهای زیادی ندارد و احتمالا به تمام وزنها برای یادگیری نیاز دارد



شکل ۱۸: شبکه با ال ۲ با ضریب 0.5 که باز هم نتوانسته یاد بگیرد چون پیچیدگی مدل به اندازه مسئله است و نیازی به این حد از پنالتی نداریم

۹ اکتویشن فانکشن



شکل ۱۹: شبکه با اکتویشن فانکشن تانژانت هیپربولیک

۱۰ بهینه‌ساز

شکل ۲۰: شبکه با اپتیمایزر اس جی دی که بدتر عملد کرده و آدام بهتر بوده