



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

پروژه دوم درس رایانش عصبی

نگارش
سیدمهدی میرفندرسکی
مدرس
دکتر رضا صفابخش
آبان ۱۴۰۱

فهرست مطالب

۱	سوال اول	۲
۲	سوال دوم	۵
۳	سوال سوم	۵
۴	سوال چهارم	۱۲
۵	سوال پنجم	۱۶
۶	سوال ششم	۲۲
۷	سوال هفتم	۲۵

در مواردی که گراف شبکه خواسته شده است، از دو طریق این کار انجام شده است. نمای کلی گراف خروجی تنسوربرد گذاشته شده است. در این ابزار با کلیک بر روی هر قسمت اطلاعات جامعی از شبکه بدست می‌آید. که به عنوان مثال معماری لایه‌ها با تابع دیگری در گزارش آورده شده است.

۱ سوال اول

۱.۱ تخمین مقادیر گم‌شده

روش‌های متعددی برای این قسمت وجود دارد که در ادامه به معرفی برخی از آن‌ها می‌پردازیم:

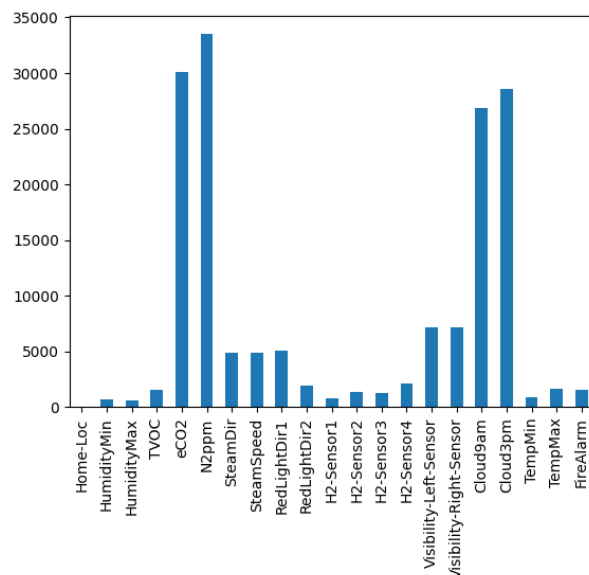
- حذف سطر یا ستون‌ها: در این روش سطر و ستون‌هایی که شامل مقادیر گم‌شده هستند با شروطی می‌توانند حذف شوند.
 - حذف سطر یا ستون‌هایی که درصد قابل توجهی از آن داده گم‌شده است.
 - حذف یا نگهداری سطر یا ستون‌هایی که از یک مقدار حد آستانه بیشتر یا کمتر مقادیر گم‌شده دارند.
 - همچنین می‌توان بر اساس زیر مجموعه‌ای از سطر یا ستون‌ها این مقادیر را حذف کرد.
 - اما بعضی اوقات این مقادیر محدود هستند و ترجیح بر آن خواهد بود که سطرها یا ستون‌ها حذف نشوند. یک روش برای این کار پر کردن این مقدار با مقادیر ثابت خواهد بود که برای هر نوع داده‌ای می‌تواند مورد استفاده قرار گیرد.
 - روش دیگر مورد استفاده برای پر کرده داده‌ها، استفاده از میانگین، میانه یا مد خواهد بود که می‌تواند مفید واقع شود.
 - روش دیگر استفاده از مقدار قبلی یا بعدی در یک ستون مشخص است (مفید برای داده‌های سری زمانی).
- برای قسمت پیش پردازش به طور کلی بدین صورت عمل کردیم که اگر سطر بیش از ۱۴ مقدار گم‌شده داشت، حذف شد. بعد از آن میانگین را برای مقادیر گم‌شده عددی و بیشترین برچسب هر ستون برای مقادیر گم‌شده دسته‌بندی را جایگذاری کردیم. در شکل تعداد مقادیر گم‌شده قبل از حذف و در شکل تعداد مقادیر گم‌شده مشاهده می‌شود. همچنین در شکل مقادیر جایگذاری شده مشاهده می‌شود.

۲.۱ نرمال‌سازی

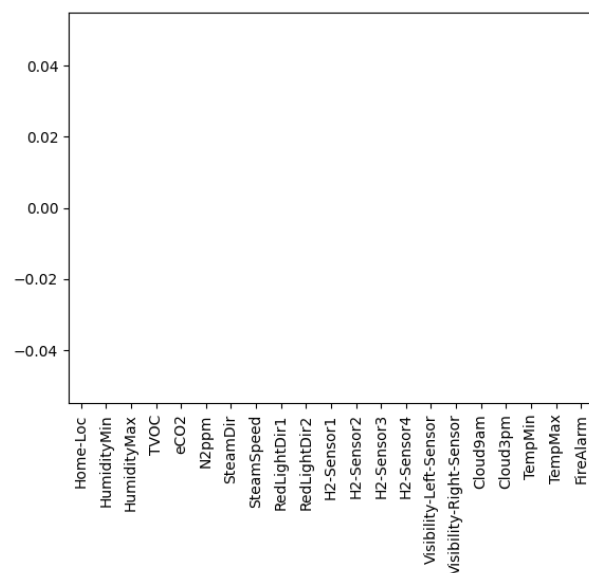
در آزمایش‌های متعدد دیده شده است که مدل‌ها در برابر مجموعه داده‌ای که نرمال‌سازی شده است، عملکرد بهتری از خود نشان داده‌اند. به طور کلی این اتفاق زمانی می‌افتد که مقیاس ویژگی‌های عددی یکسان نباشد. به عبارت دیگر، اگر یک ویژگی داشته باشیم که مقدار زیادی از ویژگی دیگر بزرگتر باشد، به طور خودکار ویژگی بزرگتر تاثیر بیشتری در مدل خواهد داشت (مثل درآمد و سن). با نرمال‌سازی تاثیر این دو یکی خواهد شد. پس هدف انتقال تمام ویژگی‌های عددی به یک محدوده خاص است. در این مجموعه داده به عنوان مثال Visibility ها مقادیر بزرگتری دارند.

۳.۱ تبدیل ویژگی گسسته به عددی

- Integer Encoding
- One Hot Encoding
- Learned Embedding



شکل ۱: تعداد مقادیر گم‌شده برای هر ویژگی قبل از این پیش پردازش



شکل ۲: تعداد مقادیر گم‌شده برای هر ویژگی بعد از انجام این قسمت

در این قسمت از روش One Hot Encoding استفاده شد. در این روش هر برچسب به یک بردار دودویی تبدیل می‌شود. برای هر داده بررسی می‌شود که آیا این برچسب موجود است یا خیر. پس به ازای هر برچسب ستون (ویژگی) جدید به مجموعه داده اضافه می‌شود. این روش بدین دلیل انتخاب شد که تضمین می‌کند که مدل به بعضی برچسب‌ها ارزش متفاوتی نمی‌دهد. به عنوان مثال، مقدار ۸ بزرگتر از مقدار ۱ است، اما این باعث نمی‌شود که ۸ از ۱ مهم‌تر باشد. همین امر در مورد کلمات صادق است: ارزش جهت‌ها باهم برابر است.

```
##### Mode of categ cols #####
SteamDir ==> W
RedLightDir2 ==> SE
FireAlarm ==> No
Home-Loc ==> Hobart
RedLightDir1 ==> N
##### Avg of numerical cols #####
HumidityMin ==> 12.217642283622624
HumidityMax ==> 23.28001711848649
TVOC ==> 2.287303890043821
eCO2 ==> 5.458218952175569
N2ppm ==> 7.613094153064326
SteamSpeed ==> 39.92479734651808
H2-Sensor1 ==> 14.045199574870098
H2-Sensor2 ==> 18.692535063476345
H2-Sensor3 ==> 68.71124200655797
H2-Sensor4 ==> 51.4309511015583
Visibility-Left-Sensor ==> 1017.6638844327515
Visibility-Right-Sensor ==> 1015.2686167070258
Cloud9am ==> 4.433639172582966
Cloud3pm ==> 4.5024140524540845
TempMin ==> 17.009222505865512
TempMax ==> 21.716109165380505
```

شکل ۳: مقادیر جایگذاری شده برای مقادیر گم‌شده

۴.۱ حذف داده‌های پرت

داده پرت، داده ایست که در یک نمونه تصادفی از یک جامعه، فاصله‌ای غیرعادی با مقادیر دیگر دارد. شناسایی داده‌های پرت دارای اهمیت است زیرا ممکن است نشان‌دهنده داده‌های بد باشد، و مدل را دچار انحراف کند. راه‌های شناسایی داده‌های پرت مبتنی بر سه روش هستند:

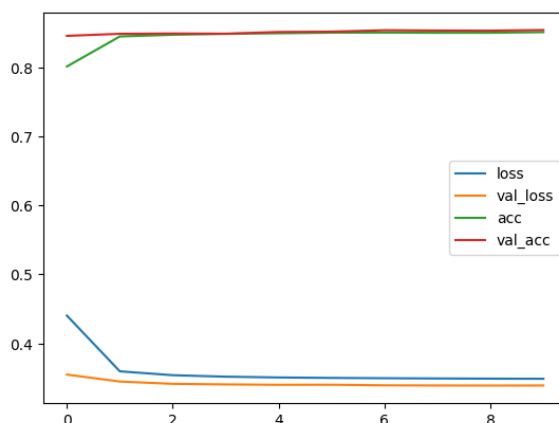
- روش‌های آماری: شناسایی با تجزیه و تحلیل بصری داده‌های تک متغیره با استفاده از Boxplots، نمودارهای پراکنده می‌تواند به یافتن مقادیر پرت در داده‌ها کمک کند. با فرض توزیع نرمال، ۶۸ درصد داده‌ها در یک طول انحراف استاندارد، ۹۵ درصد در طول دو انحراف استاندارد و ۷.۹۹ درصد در سه انحراف استاندارد از میانگین قرار می‌گیرند. با این فروض با در نظر گرفتن فاصله بین چارک اول و سوم، داده‌های با فاصله کمتر از تفریق این فاصله از چارک اول و همچنین بیشتر از جمع این فاصله با چارک سوم را می‌توان به عنوان داده پرت در نظر گرفت.
- روش‌های مجاورت: روش‌های مبتنی بر مجاورت، تکنیک‌های خوشه‌بندی را برای شناسایی خوشه‌ها در داده‌ها و یافتن مرکز هر خوشه به کار می‌گیرند. روش بدین صورت است که یک آستانه ثابت می‌شود و فاصله هر نقطه داده از مرکز خوشه ارزیابی می‌شود و سپس نقاط داده پرت حذف می‌شود. یکی از چالش‌ها تعیین آستانه خواهد بود. دو نوع مبتنی بر فاصله و چگالی (kmean) را داراست.
- روش‌های تصویری: روش‌های پروجکشن از تکنیک‌هایی مانند PCA برای مدل‌سازی داده‌ها در یک زیرفضای با ابعاد پایین‌تر با استفاده از همبستگی‌های خطی استفاده می‌کنند. پس از آن، فاصله هر نقطه داده تا صفحه‌ای که متناسب با فضای فرعی است محاسبه می‌شود. سپس می‌توان از این فاصله برای یافتن نقاط پرت استفاده کرد.

اما شناسایی داده‌های پرت با kmean بدین صورت خواهد بود که با استفاده از یک مقدار برای k بعد از خوشه‌بندی، فاصله هر داده از مرکز را حساب کرده و داده‌های بیشتر از یک آستانه مشخص را به عنوان داده پرت در نظر گرفته و آن حذف می‌شود. در این بخش مقدار ۲ برای k و آستانه ۷۰ تعداد ۴۳۸ نمونه حذف شدند.

۲ سوال دوم

شبکه عمیقی که برای این بخش باید طراحی شود، یک تک پرسپترون خواهد بود. ورودی این تک پرسپترون تمام ویژگی‌ها و خروجی آن هم برچسب مورد نظر خواهد بود. اگر یک تک پرسپترون درجه اول داشته باشیم، میتوان مسائل جداپذیر خطی را حل کند. این بدان دلیل است که تنها یک ضریب در ویژگی‌ها ضرب خواهد شد و با یک مقدار ثابت جمع خواهد شد. بدین ترتیب یک مدل خطی خواهیم داشت. درواقع با یک شبکه تک لایه می‌توان به این دانش رسید.

همانطور که مشاهده می‌شود صحت از ۸۵ درصد بیشتر نمی‌شود. این مسئله بوضوح جداپذیر خطی نیست. به بیان دیگر پرسپترون نتوانست به ۱۰۰ درصد صحت برسد.



شکل ۴: خطا و صحت پرسپترون

۳ سوال سوم

در این قسمت ۹۶ شبکه عصبی (شامل صفر لایه پنهان تا ۷ لایه پنهان) آموزش داده شد (حدود ۵۰ دقیقه). فرض‌هایی که در تمام این شبکه‌های عصبی یکسان بود به صورت زیر است:

ای‌پاک	اندازه دسته	تابع فعالیت خروجی	تابع فعالیت پنهان	نسبت داده‌های اعتبارسنجی	نسبت داده‌های آموزشی
۱۰	۶۴	sigmoid	relu	۱۰	۷۰

اندازه دسته بهتر بود برابر ۱۶ یا ۳۲ قرار داده می‌شد اما آموزش شبکه‌های عصبی بسیار زمان بر بود، به همین دلیل اندازه دسته برابر ۶۴ گذاشته شد ب این حال آموزش ۹۶ شبکه حدود ۵۰ دقیقه زمان برد.

همچنین تعداد نوروں لایه‌های پنهان بر اساس لیستی شامل ۶۴، ۱۲۸، ۱۹۲ و ۲۵۶ بود. در شبکه‌های عصبی با یک و دو لایه تمام حالت‌ها در نظر گرفته شدند، اما در شبکه‌های سه لایه تا ۷ لایه به ۱۵ انتخاب تصادفی از تمام حالت‌های ممکن پسندیده شد. در ادامه جدولی شامل تمام شبکه‌ها به همراه معماری و همچنین صحت داده‌های آموزشی و اعتبارسنجی آورده شده است:

صحت مجموعه اعتبارسنجی	صحت مجموعه آموزش	معماری شبکه	تعداد لایه پنهان	شماره
85.22	85.11	-	۰	۱
85.99	87.58	۶۴	۱	۲

۳	۱	۱۲۸	88.12	86.01
۴	۱	۱۹۲	88.58	86.07
۵	۱	۲۵۶	88.51	86.13
۶	۲	۶۴-۶۴	88.59	86.14
۷	۲	۱۲۸-۶۴	88.63	86.04
۸	۲	۱۹۲-۶۴	88.83	85.93
۹	۲	۲۵۶-۶۴	88.97	85.82
۱۰	۲	۶۴-۱۲۸	90.06	85.25
۱۱	۲	۱۲۸-۱۲۸	89.67	85.86
۱۲	۲	۱۹۲-۱۲۸	89.97	85.52
۱۳	۲	۲۵۶-۱۲۸	90.86	84.47
۱۴	۲	۶۴-۱۹۲	90.76	85.65
۱۵	۲	۱۲۸-۱۹۲	91.23	85.45
۱۶	۲	۱۹۲-۱۹۲	91.64	85.4
۱۷	۲	۲۵۶-۱۹۲	91.93	85.03
۱۸	۲	۶۴-۲۵۶	91.28	85.28
۱۹	۲	۱۲۸-۲۵۶	91.27	85.76
۲۰	۲	۱۹۲-۲۵۶	92.15	85.06
۲۱	۲	۲۵۶-۲۵۶	92.4	85.03
۲۲	۳	۱۲۸-۱۹۲-۱۹۲	93.4	83.87
۲۳	۳	۲۵۶-۶۴-۶۴	89.5	85.58
۲۴	۳	۱۹۲-۲۵۶-۱۲۸	93.12	84.15
۲۵	۳	۱۲۸-۶۴-۱۹۲	91.78	84.85
۲۶	۳	۱۹۲-۱۲۸-۶۴	90.2	85.58
۲۷	۳	۱۹۲-۱۹۲-۱۹۲	93.5	84.36
۲۸	۳	۶۴-۲۵۶-۱۹۲	93.19	84.87
۲۹	۳	۲۵۶-۶۴-۱۹۲	92.48	84.73
۳۰	۳	۶۴-۱۲۸-۱۲۸	91.65	85.25
۳۱	۳	۲۵۶-۶۴-۱۲۸	91.06	85.01
۳۲	۳	۱۹۲-۶۴-۲۵۶	92.23	84.44
۳۳	۳	۱۹۲-۱۲۸-۲۵۶	93.23	84.76
۳۴	۳	۱۹۲-۱۲۸-۱۹۲	93.33	84.32
۳۵	۳	۱۹۲-۲۵۶-۶۴	90.69	85.09
۳۶	۳	۶۴-۱۲۸-۱۹۲	92.28	85.07

۳۷	۴	۲۵۶-۱۲۸-۱۹۲-۲۵۶	94.16	84.36
۳۸	۴	۶۴-۱۹۲-۱۹۲-۱۹۲	93.43	84.85
۳۹	۴	۱۲۸-۶۴-۶۴-۶۴	89.64	85.28
۴۰	۴	۲۵۶-۱۲۸-۶۴-۶۴	89.83	85.01
۴۱	۴	۱۹۲-۶۴-۱۲۸-۱۹۲	92.82	84.94
۴۲	۴	۲۵۶-۱۹۲-۱۲۸-۲۵۶	93.33	84.76
۴۳	۴	۶۴-۶۴-۶۴-۱۲۸	91.04	85.43
۴۴	۴	۲۵۶-۱۲۸-۱۹۲-۲۵۶	94.08	84.32
۴۵	۴	۱۹۲-۱۹۲-۶۴-۱۲۸	91.69	84.73
۴۶	۴	۲۵۶-۱۲۸-۲۵۶-۱۹۲	93.32	84.66
۴۷	۴	۱۹۲-۱۹۲-۱۲۸-۱۹۲	93.48	84.2
۴۸	۴	۲۵۶-۲۵۶-۲۵۶-۱۲۸	93.18	84.51
۴۹	۴	۱۹۲-۱۹۲-۱۹۲-۱۹۲	93.97	84.04
۵۰	۴	۶۴-۱۲۸-۱۲۸-۶۴	90.02	85.16
۵۱	۴	۱۲۸-۱۹۲-۲۵۶-۲۵۶	93.97	84.67
۵۲	۵	۱۲۸-۲۵۶-۲۵۶-۱۹۲-۲۵۶	93.84	85.09
۵۳	۵	۶۴-۲۵۶-۱۹۲-۶۴-۲۵۶	92.47	84.5
۵۴	۵	۱۹۲-۶۴-۱۹۲-۶۴-۱۹۲	92.14	84.94
۵۵	۵	۶۴-۱۲۸-۱۲۸-۱۲۸-۱۹۲	93.05	84.48
۵۶	۵	۱۲۸-۶۴-۱۲۸-۶۴-۱۲۸	91.29	84.48
۵۷	۵	۶۴-۱۲۸-۱۹۲-۱۹۲-۶۴	90.75	85.15
۵۸	۵	۶۴-۱۲۸-۲۵۶-۲۵۶-۲۵۶	94.47	83.75
۵۹	۵	۶۴-۶۴-۲۵۶-۲۵۶-۱۲۸	92.26	83.98
۶۰	۵	۱۲۸-۱۲۸-۲۵۶-۶۴-۶۴	90.33	85.03
۶۱	۵	۱۹۲-۱۹۲-۲۵۶-۱۲۸-۱۹۲	93.49	84.6
۶۲	۵	۱۹۲-۱۹۲-۶۴-۲۵۶-۱۲۸	92.12	85.92
۶۳	۵	۶۴-۱۲۸-۲۵۶-۶۴-۱۲۸	91.21	84.66
۶۴	۵	۱۹۲-۲۵۶-۲۵۶-۲۵۶-۱۹۲	94.05	83.86
۶۵	۵	۱۹۲-۱۲۸-۶۴-۲۵۶-۲۵۶	93.91	84.55
۶۶	۵	۱۲۸-۶۴-۱۹۲-۲۵۶-۶۴	91.04	85.19
۶۷	۶	۱۲۸-۶۴-۱۹۲-۶۴-۱۲۸-۲۵۶	93.04	85.27
۶۸	۶	۶۴-۱۹۲-۱۹۲-۶۴-۲۵۶-۶۴	90.73	84.97
۶۹	۶	۱۹۲-۲۵۶-۱۹۲-۶۴-۲۵۶-۱۹۲	92.31	84.72
۷۰	۶	۲۵۶-۲۵۶-۲۵۶-۲۵۶-۱۹۲-۱۹۲	93.75	84.48

۷۱	۶	۲۵۶-۲۵۶-۱۲۸-۲۵۶-۱۲۸-۶۴	90.67	85.09
۷۲	۶	۱۲۸-۱۹۲-۱۲۸-۱۲۸-۱۹۲-۶۴	90.14	85.31
۷۳	۶	۱۹۲-۶۴-۶۴-۱۹۲-۱۲۸-۶۴	90.64	84.72
۷۴	۶	۱۹۲-۲۵۶-۶۴-۱۲۸-۶۴-۱۹۲	91.71	85.18
۷۵	۶	۶۴-۱۹۲-۲۵۶-۲۵۶-۱۲۸-۶۴	90.41	85.42
۷۶	۶	۲۵۶-۱۲۸-۱۲۸-۱۲۸-۲۵۶-۶۴	91.0	84.39
۷۷	۶	۶۴-۱۲۸-۶۴-۶۴-۲۵۶-۲۵۶	93.63	84.15
۷۸	۶	۲۵۶-۶۴-۶۴-۱۲۸-۶۴-۱۲۸	91.58	85.12
۷۹	۶	۱۲۸-۲۵۶-۶۴-۶۴-۱۲۸-۲۵۶	92.9	85.01
۸۰	۶	۶۴-۶۴-۶۴-۲۵۶-۱۹۲-۶۴	90.85	84.75
۸۱	۶	۶۴-۲۵۶-۶۴-۱۲۸-۶۴-۱۲۸	91.3	85.31
۸۲	۷	۲۵۶-۲۵۶-۱۹۲-۱۲۸-۲۵۶-۱۹۲-۱۹۲	92.47	85.18
۸۳	۷	۱۹۲-۱۹۲-۲۵۶-۱۲۸-۶۴-۲۵۶-۲۵۶	94.05	84.87
۸۴	۷	۱۹۲-۱۲۸-۱۹۲-۶۴-۱۲۸-۱۲۸-۶۴	90.2	84.97
۸۵	۷	۱۹۲-۱۹۲-۲۵۶-۶۴-۶۴-۱۹۲-۱۲۸	92.0	84.12
۸۶	۷	۱۲۸-۶۴-۶۴-۱۹۲-۶۴-۱۲۸-۱۲۸	91.24	85.71
۸۷	۷	۱۹۲-۶۴-۱۲۸-۱۹۲-۱۲۸-۲۵۶-۶۴	90.69	85.76
۸۸	۷	۱۹۲-۱۲۸-۱۲۸-۱۲۸-۲۵۶-۲۵۶-۱۹۲	93.35	84.58
۸۹	۷	۱۲۸-۲۵۶-۱۲۸-۱۹۲-۲۵۶-۲۵۶-۱۹۲	93.29	84.2
۹۰	۷	۱۹۲-۱۲۸-۲۵۶-۶۴-۱۲۸-۲۵۶-۱۹۲	92.84	84.85
۹۱	۷	۱۹۲-۱۹۲-۲۵۶-۱۲۸-۱۹۲-۱۲۸-۲۵۶	93.04	84.73
۹۲	۷	۱۲۸-۱۲۸-۶۴-۱۹۲-۶۴-۱۲۸-۲۵۶	92.4	84.11
۹۳	۷	۲۵۶-۲۵۶-۱۹۲-۱۲۸-۲۵۶-۱۹۲-۱۹۲	93.44	84.38
۹۴	۷	۱۹۲-۱۹۲-۱۹۲-۱۲۸-۶۴-۲۵۶-۱۲۸	92.26	85.53
۹۵	۷	۶۴-۱۹۲-۱۲۸-۶۴-۱۹۲-۶۴-۲۵۶	92.38	84.78
۹۶	۷	۱۹۲-۱۲۸-۱۹۲-۲۵۶-۱۹۲-۱۲۸-۶۴	90.54	85.43

تفسیر متعددی می‌توان بر روی این جدول ارائه داد. به طور کلی با افزایش لایه‌ها به علت پیچیده شدن مدل (افزایش پارامترها)، بیش برآزش اتفاق می‌افتد (افزایش صحت آموزش و کاهش صحت اعتبارسنجی). بیشترین صحت مجموعه اعتبارسنجی مربوط به شبکه‌های ۶ و ۵ و بعد از آن به شبکه‌های ۴، ۷ و ۳ مربوط می‌شود. تمام این شبکه‌ها مربوط به شبکه‌های با ۱ و ۲ لایه پنهان هستند. البته در موارد انتخابی اختلاف مقادیر صحت آموزشی و اعتبارسنجی نیز لحاظ شده‌است که کم باشد. پس با موارد بیشتری را برای این شبکه‌های با ۱ و ۲ لایه لحاظ می‌کنیم. در ادامه نتایج آموزش ۲۱ شبکه با فرض‌های اولیه آورده شده است. بدیهی است که موارد مشترک با جدول قبل ممکن است نتایج متفاوتی داشته باشند.

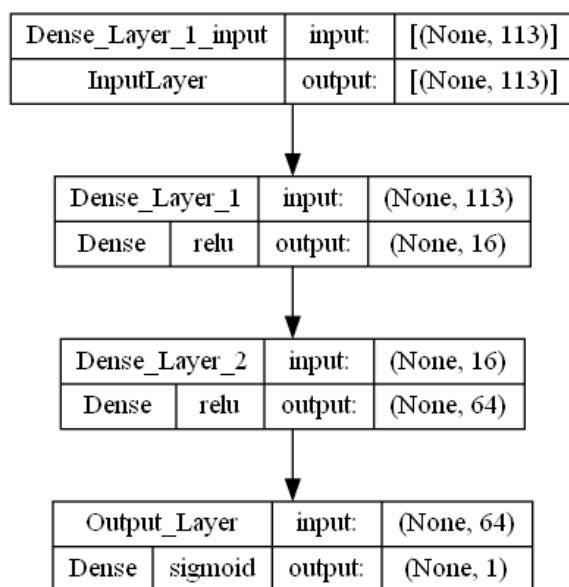
صحت مجموعه اعتبارسنجی	صحت مجموعه آموزش	معماری شبکه	تعداد لایه پنهان	شماره
86.25	86.68	۱۶	۱	۱

۲	۱	۳۲	86.9	86.11
۳	۱	۶۴	87.57	86.02
۴	۱	۱۲۸	88.13	85.99
۵	۱	۲۵۶	88.79	85.82
۶	۲	۱۶-۱۶	86.85	86.1
۷	۲	۳۲-۱۶	86.75	86.13
۸	۲	۶۴-۱۶	86.94	86.71
۹	۲	۱۲۸-۱۶	86.94	85.73
۱۰	۲	۱۶-۳۲	87.61	85.98
۱۱	۲	۳۲-۳۲	87.41	86.22
۱۲	۲	۶۴-۳۲	87.25	86.25
۱۳	۲	۱۲۸-۳۲	87.42	86.25
۱۴	۲	۱۶-۶۴	88.26	86.16
۱۵	۲	۳۲-۶۴	88.41	86.01
۱۶	۲	۶۴-۶۴	88.76	85.83
۱۷	۲	۱۲۸-۶۴	88.81	85.91
۱۸	۲	۱۶-۱۲۸	89.35	85.99
۱۹	۲	۳۲-۱۲۸	89.37	85.28
۲۰	۲	۶۴-۱۲۸	89.62	85.65
۲۱	۲	۱۲۸-۱۲۸	90.06	85.27

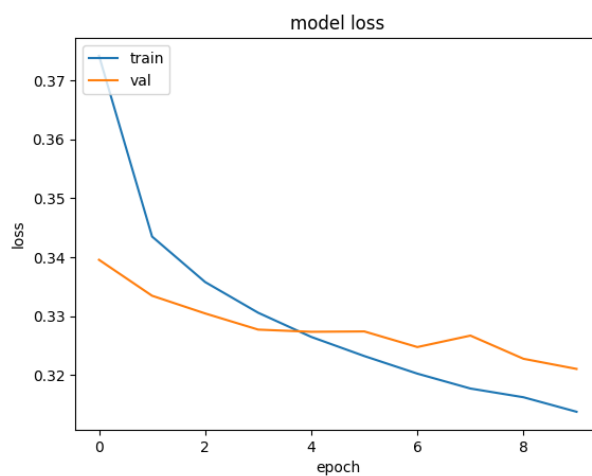
از این تعداد شبکه عصبی، موارد زیر بهترین نتایج اعتبارسنجی را ارائه کردند:

شماره	تعداد لایه پنهان	معماری شبکه	صحت مجموعه آموزش	صحت مجموعه اعتبارسنجی
۸	۲	۶۴-۱۶	86.94	86.71
۱۳	۲	۱۲۸-۳۲	87.42	86.25
۱۲	۲	۶۴-۳۲	87.25	86.25
۱	۱	۱۶	86.68	86.25

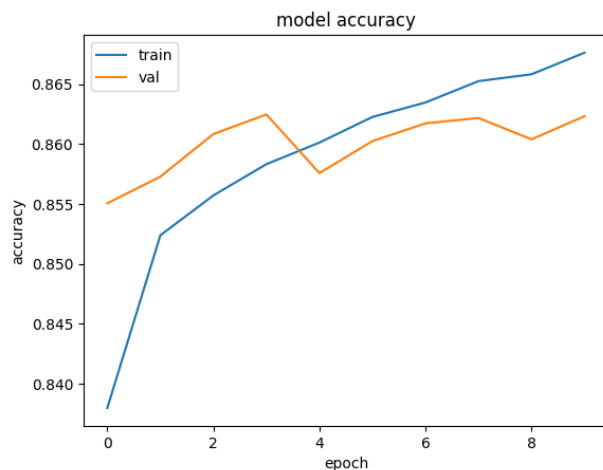
توجه در تمام جداول بالا در قسمت معماری شبکه، ترتیب لایه‌ها از راست به چپ است. پس انتخاب ما شبکه عصبی با دو لایه پنهان با تعداد نورون ۱۶ و ۶۴ خواهد بود. معماری مدل نهایی نیز به صورت زیر است: در ادامه دو نمودار صحت و خطا بر حسب ای‌پاک (تا ۱۰ ای‌پاک) را برای دو مجموعه اعتبارسنجی و آموزشی می‌بینیم. گراف شبکه نهایی با ابزار تنسور بورد نیز تولید شد که به شرح زیر است.



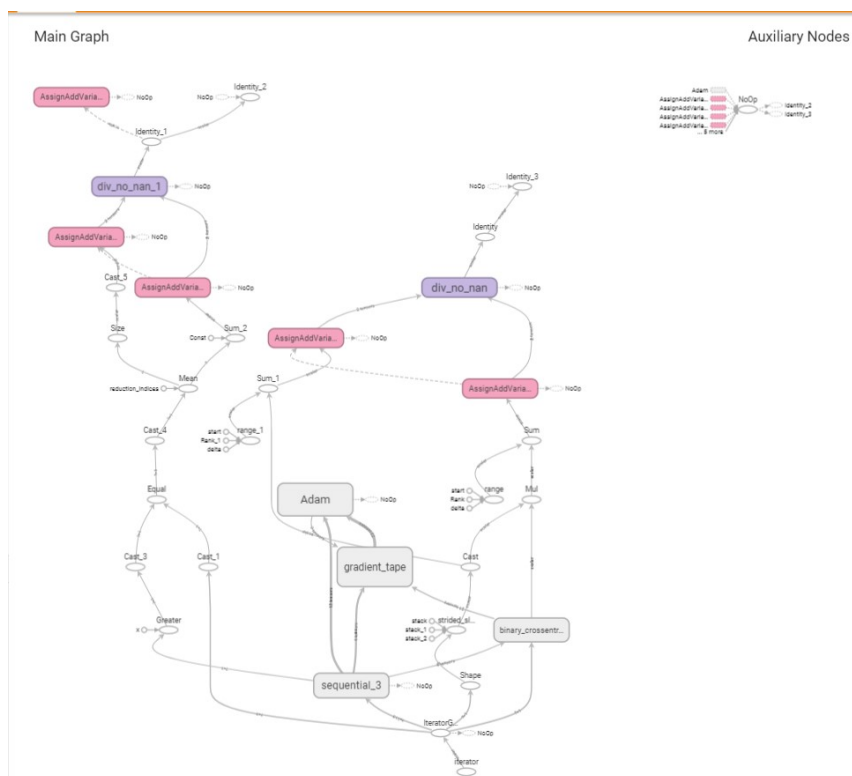
شکل ۵: معماری مدل نهایی



شکل ۶: نمودار خطا بر اساس ای پاک



شکل ۷: نمودار صحت بر اساس ای پاک



شکل ۸: گراف شبکه نهایی با ابزار تنسور بورد

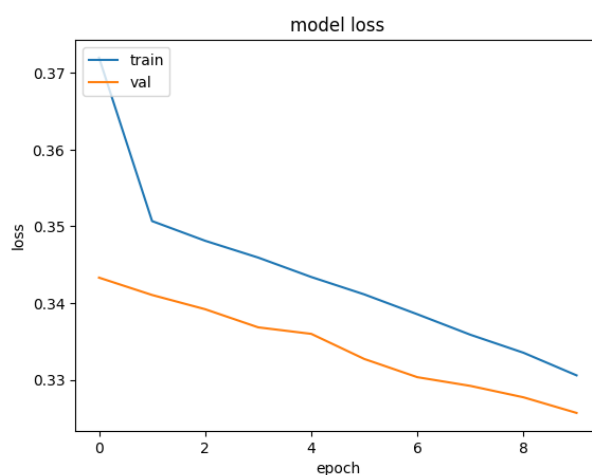
۴ سوال چهارم

توابع فعالیت با ویژگی غیرخطی بودن نقشی اساسی در شبکه‌های عصبی ایفا می‌کنند. این غیرخطی بودن به شبکه‌های عصبی اجازه می‌دهد تا نمایش‌ها و توابع پیچیده‌ای را آموزش دهند که با یک مدل رگرسیون خطی ساده امکان‌پذیر نیست. درواقع اگر توابع فعالیت نباشند، پیچیده‌ترین شبکه‌ها قابلیت مدل شدن با یک پرسپترون را خواهند داشت. حال نوع این توابع فعالیت برای غیرخطی کردن خروجی بسیار حائز اهمیت است. درواقع این انواع مختلف است که منجر به خروجی‌های متفاوت خواهد شد. همچنین برروی سرعت آموزش مدل هم تاثیرگذار خواهد بود.

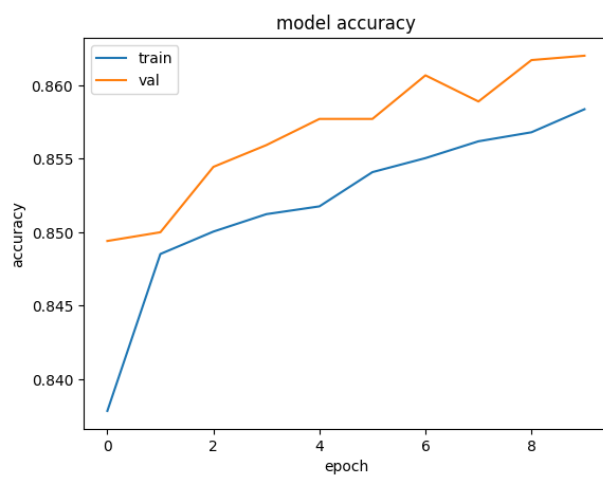
در این قسمت ۳ نوع تابع فعالیت برای شبکه حاصل سوال ۳ (۱۶ و ۶۴ نورون برای لایه‌های اول و دوم) در نظر گرفته شد (با فرض‌های اولیه). که خروجی نهایی در جدول زیر آمده است.

صحت مجموعه اعتبارسنجی	صحت مجموعه آموزش	معماری شبکه	تعداد لایه پنهان	تابع
86.2	85.96	۶۴-۱۶	۲	tanh
85.71	85.51	۶۴-۱۶	۲	sigmoid
85.83	86.94	۶۴-۱۶	۲	relu

همانطور که مشاهده می‌شود خروجی تا ای‌پاک ۱۰ اختلاف چندانی ندارد اما روند آموزش شبکه‌ها جالب است. نمودارهای صحت و خطا برای تابع tanh { به صورت زیر می‌باشد.

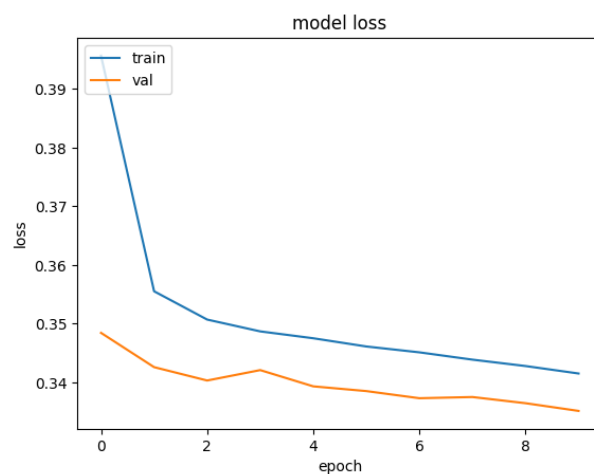


شکل ۹: نمودار خطا بر اساس ای‌پاک (tanh)

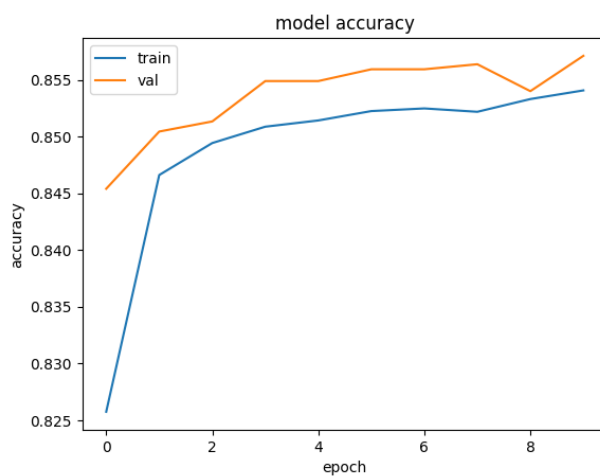


شکل ۱۰: نمودار صحت بر اساس ای‌پاک (tanh)

نمودارهای صحت و خطا برای تابع sigmoid به صورت زیر می باشد.

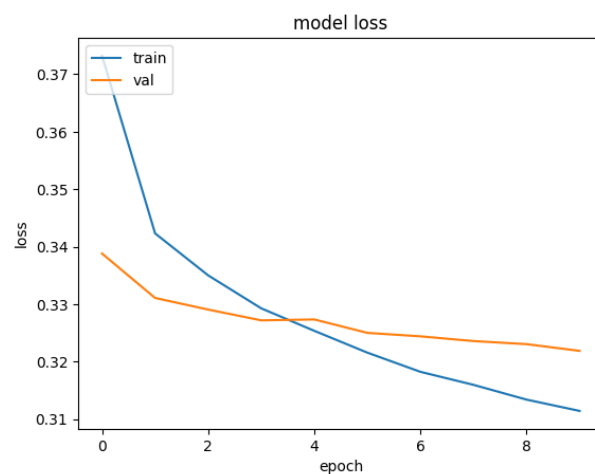


شکل ۱۱: نمودار خطا بر اساس ای پاک (sigmoid)

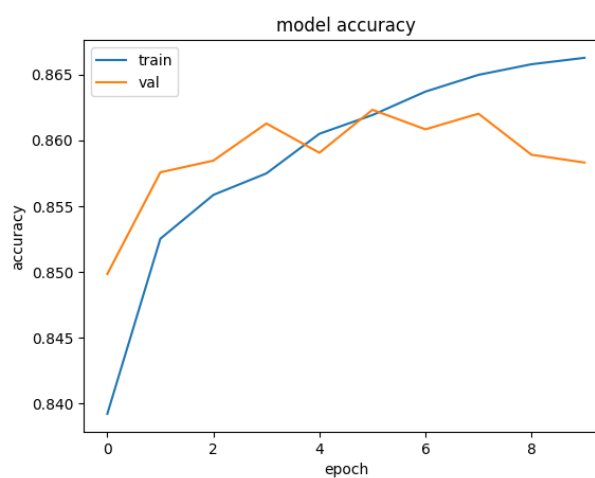


شکل ۱۲: نمودار صحت بر اساس ای پاک (sigmoid)

نمودارهای صحت و خطا برای تابع relu به صورت زیر می باشد.



شکل ۱۳: نمودار خطا بر اساس ای پاک (relu)



شکل ۱۴: نمودار صحت بر اساس ای پاک (relu)

همانطور که دیده شد، نمودارهای tanh و sigmoid در صورت ادامه دادن نتایج جالبی خواهند داد. بدین صورت که تا ای پاک ۱۰ حتی مقدار صحت اعتبارسنجی بیشتر است و هر دو صعودی هستند. این تفاوت طبیعتا ناشی از تفاوت در نوع توابع فعالیت است. بدین صورت که برای یک شبکه ممکن است یک تابع فعالیت برای لایه‌های پنهانی مناسب باشد ولی برای دیگری نوع دیگری. اما در هر صورت تا ای پاک ۱۰ بهترین عملکرد مربوط به tanh است. البته ممکن است با اجرایی دیگر نتایج متفاوت شود. به عنوان مثال با اجرایی دیگر تا ای پاک ۲۰ sigmoid بهترین بود.

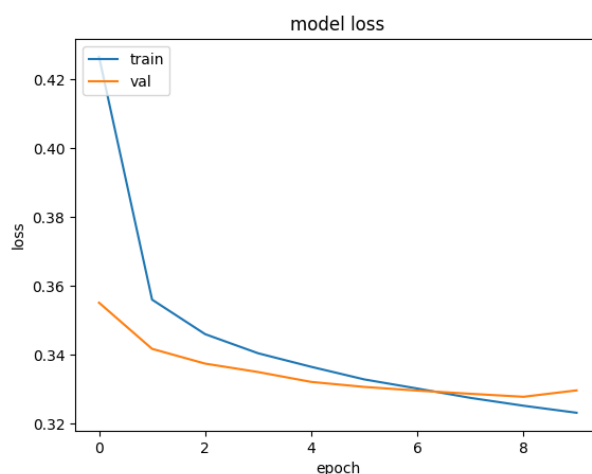
۵ سوال پنجم

اندازه دسته در روند آموزش مدل بسیار مهم است. با توجه به آزمایشات انجام شده به این نتیجه رسیدند که افزایش اندازه دسته عملکرد را کاهش می‌دهد (افزایش خطا و کاهش صحت). همچنین گفته شده است که وقتی اندازه دسته افزایش داده می‌شود، باید نرخ یادگیری برای جبران آن تنظیم شود. درواقع بیان می‌شود که تاثیر افزایش اندازه دسته عملکرد مانند کاهش نرخ یادگیری خواهد بود. اما از طرفی با اندازه‌های دسته بزرگ به روند آموزش سرعت داده می‌شود، گاهی اوقات امکان افزایش ۲ برابری در زمان محاسبات وجود دارد.

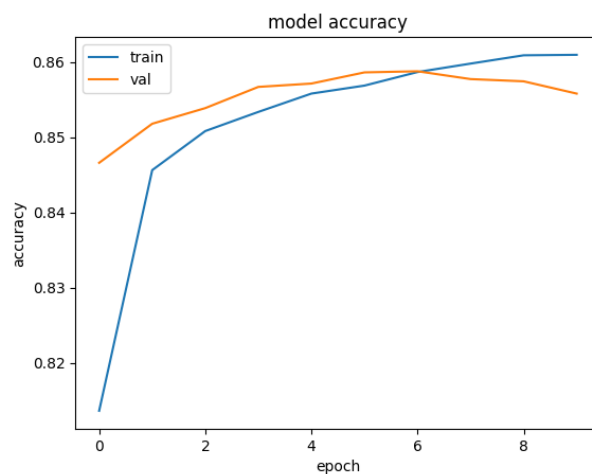
نتایج تا ۱۰ پیمایش با اندازه‌های دسته ۲۵۶، ۶۴ و ۱۶ با فرض‌های اولیه سوال سه به صورت زیر خواهد بود:

صحت مجموعه اعتبارسنجی	صحت مجموعه آموزش	معماری شبکه	تعداد لایه پنهان	اندازه دسته
85.58	86.33	۶۴-۱۶	۲	۲۵۶
86.2	86.83	۶۴-۱۶	۲	۶۴
86.1	87.34	۶۴-۱۶	۲	۱۶

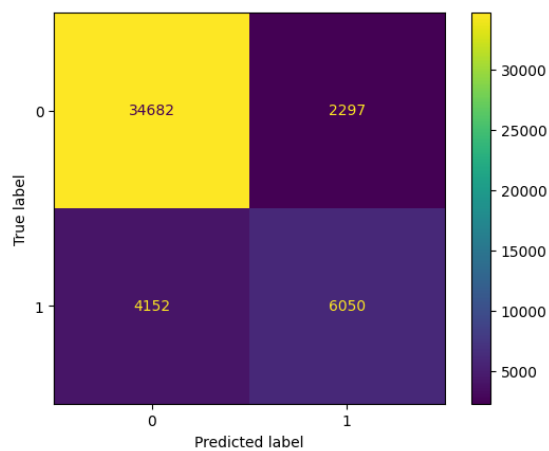
نمودارهای صحت و خطا برای مجموعه آموزشی و اعتبارسنجی و همچنین ماتریس درهم ریختگی برای مجموعه آموزشی برای اندازه دسته ۲۵۶ به شرح زیر می‌باشد:



شکل ۱۵: نمودار خطا بر اساس ای پاک (۲۵۶)

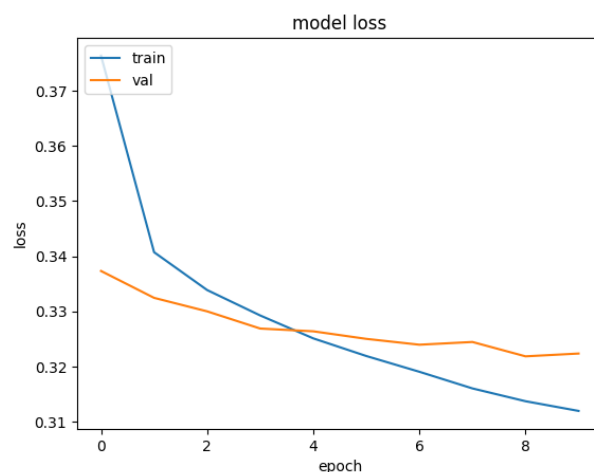


شکل ۱۶: نمودار صحت بر اساس ای پاک (۲۵۶)

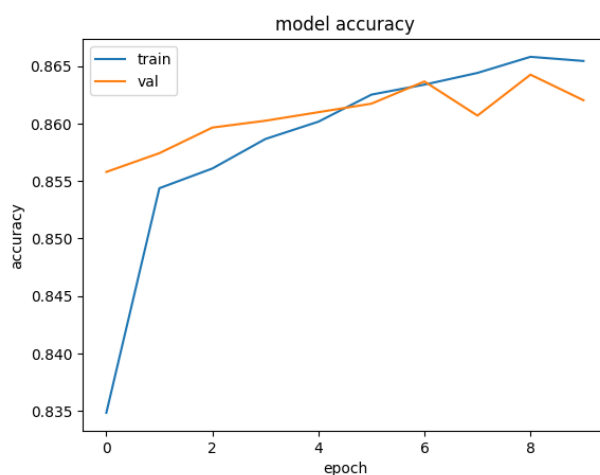


شکل ۱۷: ماتریس درهم‌ریختگی بر اساس ای پاک (۲۵۶)

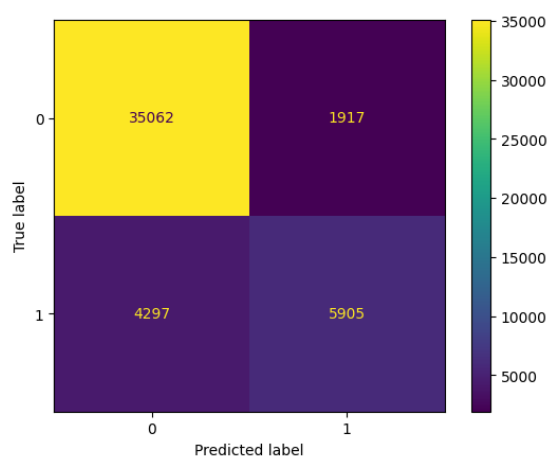
نمودارهای صحت و خطا برای مجموعه آموزشی و اعتبارسنجی و همچنین ماتریس درهم ریختگی برای مجموعه آموزشی برای اندازه دسته ۶۴ به شرح زیر می‌باشد:



شکل ۱۸: نمودار خطا بر اساس ای‌پاک (۶۴)

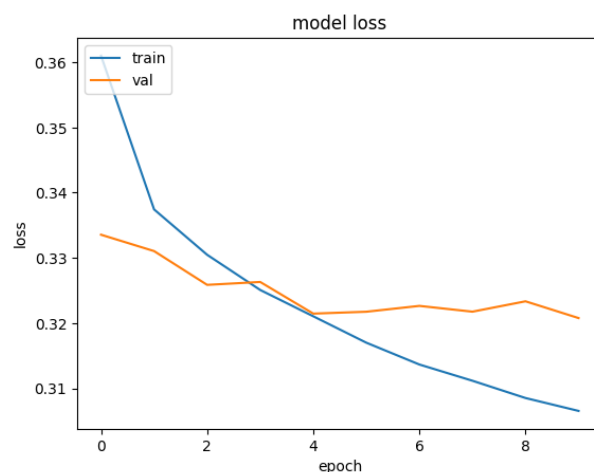


شکل ۱۹: نمودار صحت بر اساس ای‌پاک (۶۴)

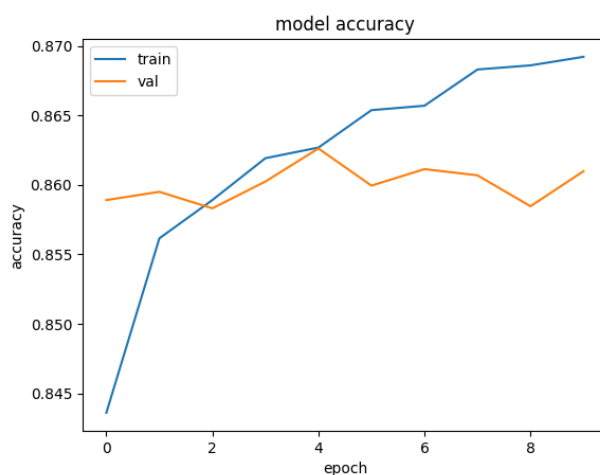


شکل ۲۰: ماتریس درهم‌ریختگی بر اساس ای‌پاک (۶۴)

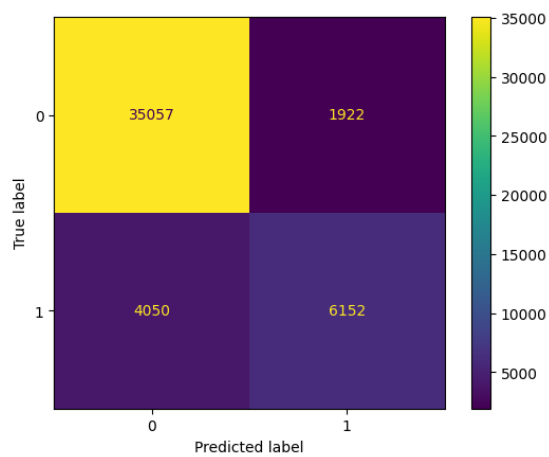
نمودارهای صحت و خطا برای مجموعه آموزشی و اعتبارسنجی و همچنین ماتریس درهم ریختگی برای مجموعه آموزشی برای اندازه دسته ۱۶ به شرح زیر می باشد:



شکل ۲۱: نمودار خطا بر اساس ای پاک (۱۶)



شکل ۲۲: نمودار صحت بر اساس ای پاک (۱۶)



شکل ۲۳: ماتریس درهم‌ریختگی بر اساس ای‌پاک (۱۶)

همانطور که در جدول جمع‌بندی اندازه دسته مشاهده می‌شود، برای مجموعه آموزش بیش‌ترین مقدار صحت مجموعه آموزشی برای کمترین مقدار اندازه دسته است. این اتفاق در ماتریس‌های درهم‌ریختگی نیز قابل مشاهده است. اما تفسیر نمودارهای صحت و خطا نیز جالب است: همانطور که مشاهده می‌شود به ازای مقادیر مختلف اندازه دسته، ای‌پاک‌هایی که بعد از کاهش صحت مجموعه اعتبارسنجی را داریم متفاوت هستند. این اتفاق با توجه به برابری تاثیر نرخ یادگیری و اندازه دسته است، کاملاً صادق است.

۶ سوال ششم

بیش‌برازش زمانی اتفاق می‌افتد که مدل دارای واریانس بالایی باشد، به‌عنوان مثال، مدل در داده‌های آموزشی به خوبی عمل می‌کند اما در مجموعه تست و اعتبارسنجی عملکرد مناسبی ندارد. به بیان دیگر، مدل وابستگی زیادی به داده‌های آموزشی پیدا می‌کند. این وابستگی به قدری زیاد است که مانع عملکرد مطلوب در مجموعه‌های آموزش ندیده می‌شود. بیش‌برازش را می‌توان با بررسی صحت و خطا اعتبارسنجی شناسایی کرد. به عنوان مثال خطا اعتبارسنجی معمولاً تا نقطه‌ای کاهش می‌یابد و بعد از آن شروع به افزایش می‌کند. که از این نقطه به بعد مدل دچار بیش‌برازش می‌شود. یکی از راه‌های ایجاد بیش‌برازش، تولید یک مدل بسیار پیچیده‌تر از نیاز مسئله است. همچنین اگر داده‌های آموزشی کم باشند، این اتفاق خواهد افتاد. تکنیک‌های مختلفی برای جلوگیری از بیش‌برازش وجود دارد:

• Drop out

• Regularization

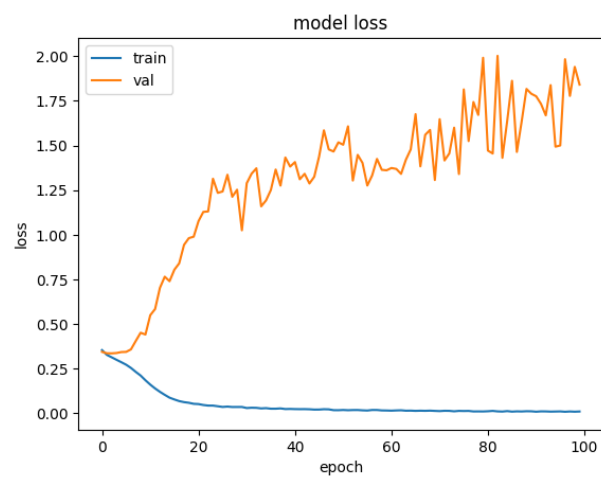
• Early stopping

البته تکنیک‌های دیگری نیز وجود دارند، که در اینجا به ۳ تکنیک بسنده شد.

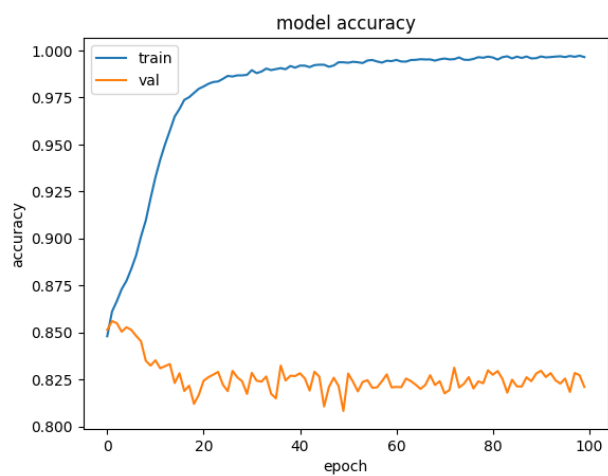
در واقع برای رسیدن به بیش‌برازش باید مدل آموزش داده شود که پارامترهای بسیار زیادی داشته باشد، تا صحت بالا و خطای کمی بر روی مجموعه آموزش و در طرف دیگر صحت کم و خطای بالا برای مجموعه تست و اعتبارسنجی بدست بیاید. برای طرحی این قسمت ما از نتایج تولید شده در سوال ۳ استفاده می‌کنیم. معماری‌ای انتخاب می‌شود که بیشترین صحت مجموعه آموزشی و کمترین صحت مجموعه اعتبارسنجی را دارا باشد. شبکه ۵۸ با ۵ لایه مخفی انتخاب شد. البته می‌توان شبکه‌های با لایه بیشتر نیز انتخاب کرد، اما همین معماری برای این بخش کافی خواهد بود. همچنین برای رخ دادن بیش‌برازش مقدار ای‌پاک را تا ۱۰۰ افزایش می‌دهیم.

صحت مجموعه تست	صحت مجموعه آموزش	معماری شبکه	تعداد لایه پنهان	اندازه دسته
82.1	99.8	۶۴-۱۲۸-۲۵۶-۲۵۶-۲۵۶	۵	۶۴

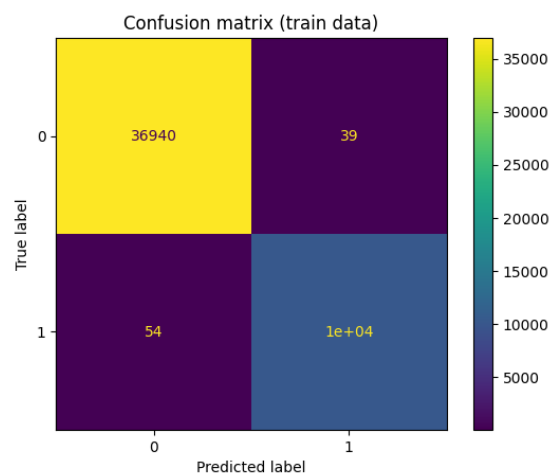
نمودارهای صحت و خطا برای مجموعه آموزشی و تست و همچنین ماتریس درهم‌ریختگی برای مجموعه آموزشی و تست به شرح زیر می‌باشد:



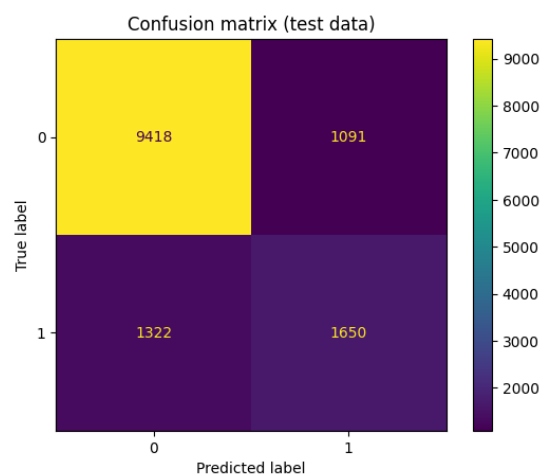
شکل ۲۴: نمودار خطا بر اساس ای پاک



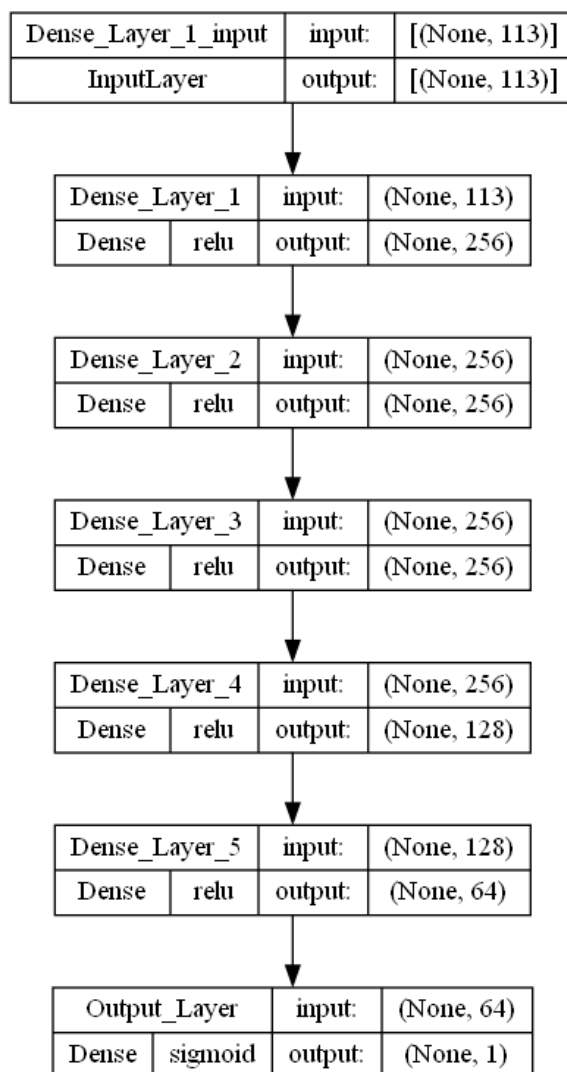
شکل ۲۵: نمودار صحت بر اساس ای پاک



شکل ۲۶: ماتریس درهم‌ریختگی مجموعه آموزش بر اساس ای‌پاک



شکل ۲۷: ماتریس درهم‌ریختگی مجموعه تست بر اساس ای‌پاک

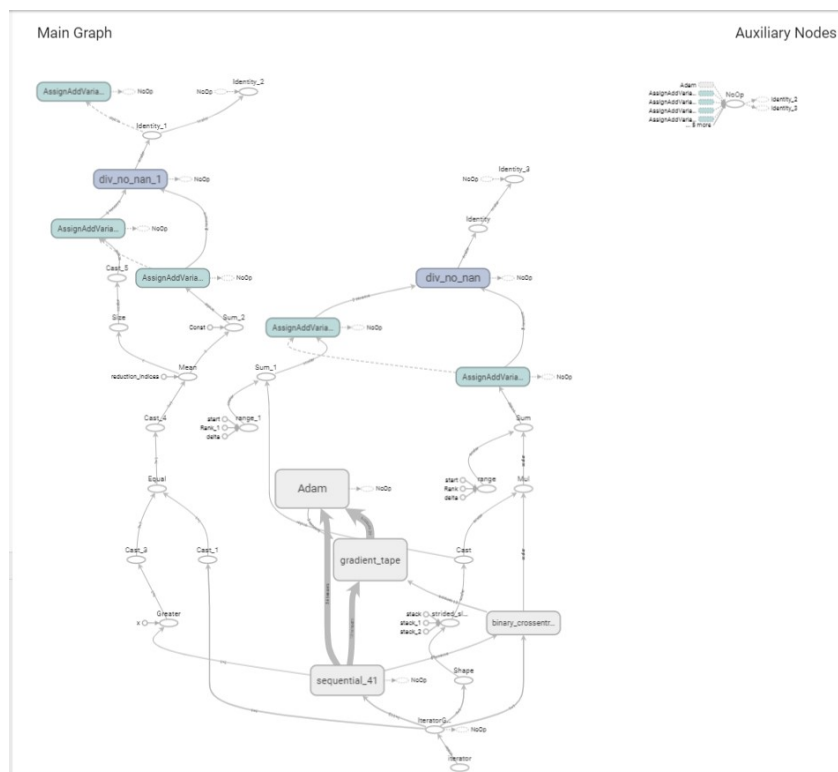


شکل ۲۸: گراف شبکه بیش‌برازش شده

گراف شبکه بیش‌برازش شده با ابزار تنسور بورد نیز تولید شد که به شرح زیر است.

۷ سوال هفتم

تعمیم‌پذیری به توانایی مدل برای انطباق مناسب با داده‌های جدید و قبلاً دیده نشده (اعتبارسنجی و تست) اشاره دارد. این مفهوم اساساً به این معنی است که مدل ما چقدر در یادگیری از داده‌های داده شده در جاهای دیگر (استفاده واقعی) خوب است. درواقع مدل اساساً برای استفاده در محیط‌های جدید و داده‌های دیده نشده آموزش می‌بیند. پس شبکه باید تعمیم‌پذیر باشد. برای رسیدن به بهترین تعمیم‌پذیری، مجموعه داده باید به سه قسمت تقسیم شود: مجموعه آموزشی برای آموزش شبکه استفاده می‌شود. خطای این مجموعه داده در طول آموزش به حداقل رسانده می‌شود. مجموعه اعتبارسنجی برای تعیین عملکرد یک شبکه عصبی بر روی الگوهایی که در طول یادگیری آموزش ندیده‌اند استفاده می‌شود. و مجموعه تست که داده‌های محیط جدید در نظر گرفته می‌شوند. یکی از استدلال‌هایی که می‌توان برای تعمیم‌پذیری مدل‌ها ارائه دید، بررسی خطای و صحت مجموعه اعتبارسنجی است. به عنوان

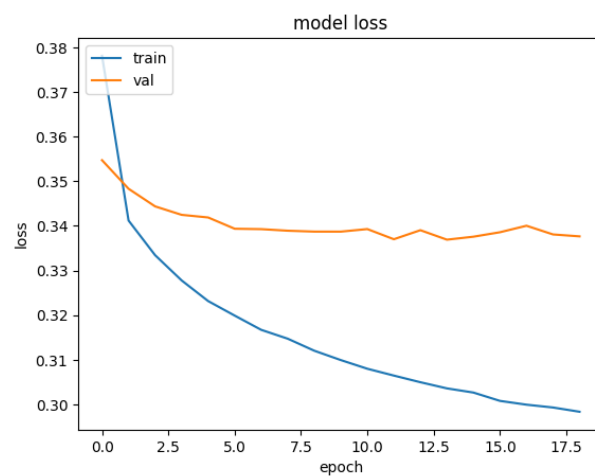


شکل ۲۹: گراف شبکه بیش‌برازش شده با ابزار تانسور بورد

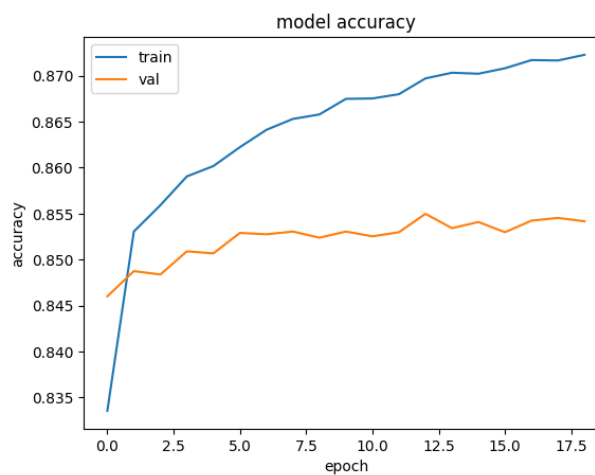
مثال صحت این مجموعه باید قابل قبول و نزدیک به مجموعه آموزشی باشد. زمانی که این ختلاف زیاد شود، تعمیم‌پذیری مدل پایین خواهد بود. با این روش می‌توان تعمیم‌پذیری مدل را نشان داد. برای بهبود تعمیم‌پذیری راه‌هایی که وجود دارند: محدود کردن تعداد وزن‌ها، تقسیم وزن، توقف زودهنگام آموزش، Regularization، کاهش وزن و اضافه کردن نویز به ورودی‌ها. برای طراحی این شبکه از همان نتایج سوال ۳ استفاده شد. درواقع بهترین نتیجه سوال ۳ که مربوط به شبکه با دو لایه پنهان ۱۶ و ۶۴ نورون بود انتخاب شد و دوباره آموزش داده شد. در این مرحله از آموزش از یک callback به نام early stopping استفاده شد تا از بیش‌برازش جلوگیری شود.

صحت مجموعه تست	صحت مجموعه آموزش	معماری شبکه	تعداد لایه پنهان	شماره
85.34	87.27	۶۴-۱۶	۲	۸

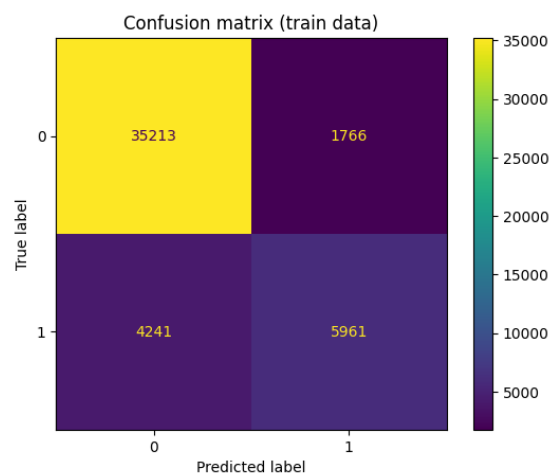
همانطور که مشاهده می‌شود، صحت تست و آموزش نزدیک به هم و همچنین از صحت مناسبی برخوردار است نمودارهای صحت و خطا برای مجموعه آموزشی و تست و همچنین ماتریس درهم‌ریختگی برای مجموعه آموزشی و تست به شرح زیر می‌باشد:



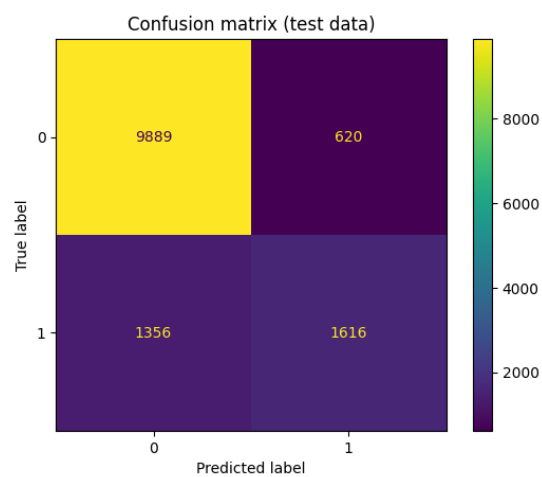
شکل ۳۰: نمودار خطا بر اساس ای پاک



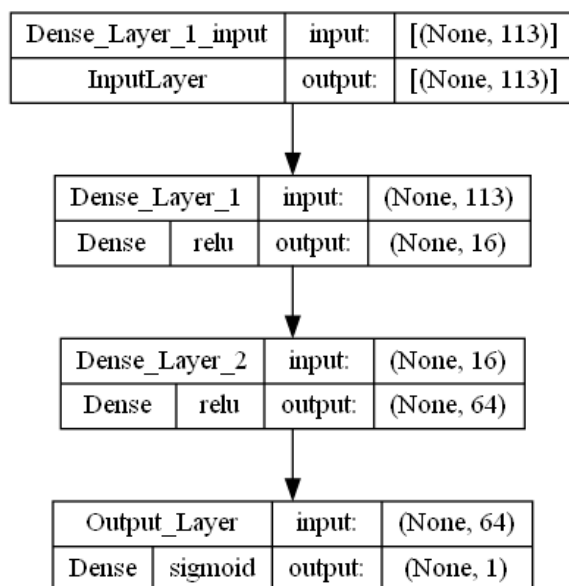
شکل ۳۱: نمودار صحت بر اساس ای پاک



شکل ۳۲: ماتریس درهم‌ریختگی مجموعه آموزش بر اساس ای‌پاک

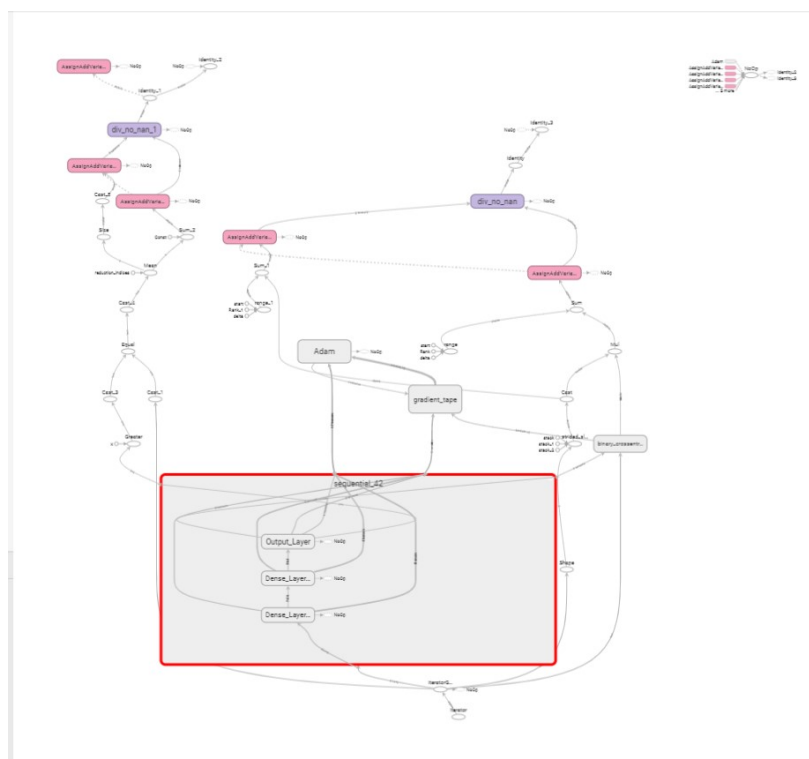


شکل ۳۳: ماتریس درهم‌ریختگی مجموعه تست بر اساس ای‌پاک



شکل ۳۴: گراف شبکه بیش‌برازش شده

گراف شبکه بیش‌برازش شده با ابزار تنسور بورد نیز تولید شد که به شرح زیر است.



شکل ۳۵: گراف شبکه بیش‌برازش شده با ابزار تنسور بورد