



## گزارش تمرین اول درس شبکه‌های عصبی



فاطمه غلامزاده

۹۹۱۳۱۰۰۳

### سوال اول

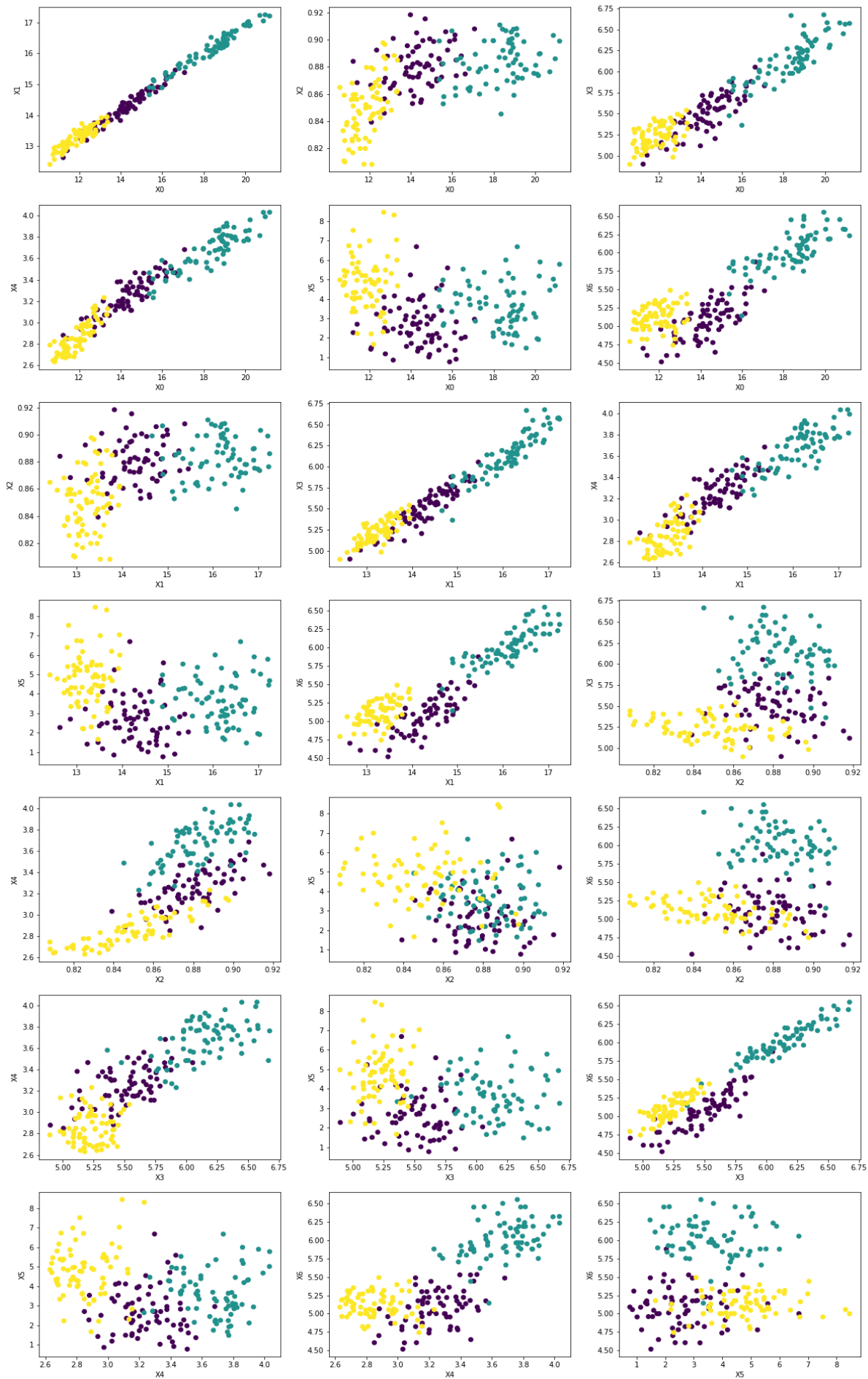
نمودار داده‌ها در شکل صفحه بعد رسم شده است. همانطور که مشاهده می‌شود ۲۱ نمودار رسم شده که در هر نمودار ۲ ویژگی از ۷ ویژگی انتخاب شده‌اند و برای آن‌ها نمودار رسم شده است.

با توجه به این شکل‌ها، داده‌ها **جداپذیر خطی نیستند** زیرا همانطور که مشاهده می‌شود در هر شکل قسمتی از داده‌های یک کلاس در محدوده‌ی داده‌های کلاس دیگری قرار گرفته و نمی‌توان با خط این داده‌ها را از هم تفکیک کرد.

### فیت کردن مدل خطی به داده‌ها:

برای اطمینان از جداپذیر خطی نبودن داده‌ها، یک مدل خطی به داده‌ها فیت شد. به این منظور از SVM با کرنل خطی با هایپرپارامتر C خیلی بزرگ (۲ به توان ۳۲) استفاده شد. زیرا وقتی هایپرپارامتر C در SVM را بزرگ در نظر می‌گیریم بهینه‌ساز را مجبور می‌کنیم که در دسته‌بندی خطای صفر داشته باشد تا تابع خطا را بهینه کند. در نتیجه مدل به داده‌ها بیش‌برازش می‌شود. اگر بتوانیم داده‌ها را به یک مدل خطی بیش‌برازش کنیم به این معناست که داده‌ها جداپذیر خطی هستند.

کد انجام این عمل در فایل Part1\_chek\_for\_linear\_Separability.ipynb آورده شده است. میزان دقت به دست آمده ۹۵ درصد بود و ۱۰۰ درصد نشد و خطا به صفر نرسید. بنابراین می‌توان نتیجه گرفت که داده‌ها جداپذیر خطی نیستند.



## سوال دوم و سوم

- **لایه ورودی:**

در این لایه از ۷ نورون استفاده شده است زیرا تعداد نورون‌ها در لایه ورودی برابر است با تعداد ویژگی‌ها یا به عبارت دیگر ابعاد داده‌های ورودی. با توجه به اینکه داده‌های مورد استفاده در این سوال، ۷ ویژگی دارند در لایه ورودی به ۷ نورون نیاز داریم. این نورون‌ها هیچ پردازشی روی داده ورودی انجام نمی‌دهند و فقط داده را به شبکه وارد می‌کنند. به همین دلیل در لایه ورودی نیازی به تعریف تابع فعالیت نیست.

- **لایه خروجی:**

در این لایه از ۳ نورون استفاده شده است. زیرا یک مساله دسته بندی سه کلاسه داریم و برای تفکیک سه کلاس، می‌توان از سه نورون یا  $\log 3$  نورون استفاده کرد که ترجیحا از ۳ نورون استفاده می‌شود. در این لایه از تابع فعالیت Relu استفاده شده است.

- **لایه پنهان:**

در این لایه از تابع فعالیت Relu استفاده شده است. از آن جا که هیچ قاعده مشخصی برای تعیین قطعی تعداد بهینه لایه‌های مخفی و نورون‌ها وجود ندارد، از فرآیند آزمون و خطا استفاده شده است. در این فرآیند، شبکه‌ای با تعداد نورون‌های متفاوت ایجاد می‌شود و با تعداد تکرارهای متفاوت آموزش می‌بیند و در نهایت نتایج حاصل ذخیره و مقایسه می‌شوند. آزمایش‌ها برای شبکه‌ای با تعداد ۱ و ۲ و ۳ لایه مخفی انجام شده‌اند و از آن جایی که افزایش تعداد لایه‌های مخفی دقت روی داده‌های تست را تغییری نمی‌داد (یا حتی کاهش می‌داد) از انجام آزمایش برای تعداد لایه‌های مخفی بیشتر خودداری شد.

برای انجام این آزمایشات یک تابع به صورت زیر تعریف شده که تعداد نورونها در لایه مخفی را به عنوان ورودی دریافت میکند و یک شبکه بر همین اساس می‌سازد:

```
#create model
def create_five_Layer_model(number_of_neurons1,number_of_neurons2,number_of_neurons3):

    model = keras.models.Sequential()

    #input layer
    model.add(keras.layers.Input(shape=(7,)))

    #hidden layer 1
    model.add(keras.layers.Dense(units=number_of_neurons1,activation='relu'))
    #hidden layer 2
    model.add(keras.layers.Dense(units=number_of_neurons2,activation='relu'))
    #hidden layer 3
    model.add(keras.layers.Dense(units=number_of_neurons3,activation='relu'))

    #output layer
    model.add(keras.layers.Dense(3,activation='softmax'))

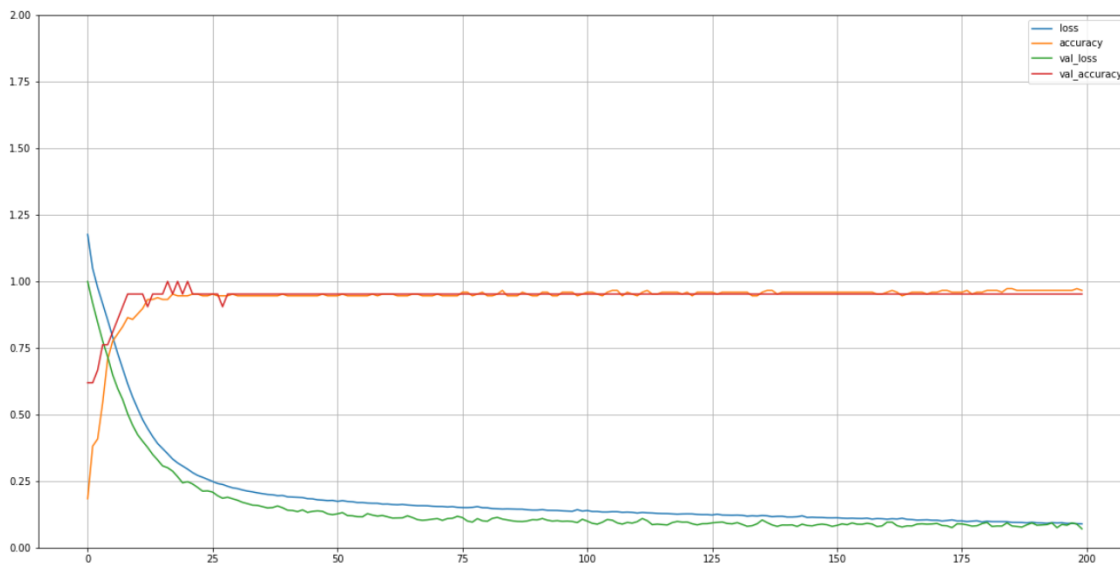
    return model
```

نتایج آزمایش‌ها در جدول زیر قابل مشاهده است:

| تعداد لایه های مخفی | Epochs | تعداد نورون های لایه مخفی اول | تعداد نورون های لایه مخفی دوم | تعداد نورون های لایه مخفی سوم | Test Loss | Test Accuracy |
|---------------------|--------|-------------------------------|-------------------------------|-------------------------------|-----------|---------------|
| 1                   | 50     | 16                            | -                             | -                             | 0.21      | 0.88          |
| 1                   | 100    | 16                            | -                             | -                             | 0.15      | 0.90          |
| 1                   | 200    | 16                            | -                             | -                             | 0.04      | 1             |
| 1                   | 500    | 16                            | -                             | -                             | 0.06      | 0.95          |
| 1                   | 50     | 64                            | -                             | -                             | 0.17      | 0.90          |
| 1                   | 100    | 64                            | -                             | -                             | 0.08      | 0.95          |
| 1                   | 200    | 64                            | -                             | -                             | 0.08      | 0.95          |
| 1                   | 500    | 64                            | -                             | -                             | 0.11      | 0.95          |
| 1                   | 50     | 128                           | -                             | -                             | 0.13      | 0.90          |
| 1                   | 100    | 128                           | -                             | -                             | 0.12      | 0.95          |
| 1                   | 200    | 128                           | -                             | -                             | 0.09      | 0.95          |
| 1                   | 500    | 128                           | -                             | -                             | 0.12      | 0.95          |
| 1                   | 50     | 256                           | -                             | -                             | 0.13      | 0.95          |
| 1                   | 100    | 256                           | -                             | -                             | 0.07      | 0.95          |
| 1                   | 200    | 256                           | -                             | -                             | 0.10      | 0.95          |
| 1                   | 500    | 256                           | -                             | -                             | 0.13      | 0.95          |
| 2                   | 50     | 64                            | 128                           | -                             | 0.21      | 0.88          |
| 2                   | 100    | 64                            | 128                           | -                             | 0.13      | 0.95          |
| 2                   | 200    | 64                            | 128                           | -                             | 0.08      | 0.95          |
| 2                   | 500    | 64                            | 128                           | -                             | 0.14      | 0.95          |
| 2                   | 50     | 64                            | 256                           | -                             | 0.18      | 0.88          |
| 2                   | 100    | 64                            | 256                           | -                             | 0.11      | 0.95          |
| 2                   | 200    | 64                            | 256                           | -                             | 0.11      | 0.95          |
| 2                   | 500    | 64                            | 256                           | -                             | 0.17      | 0.95          |
| 2                   | 50     | 128                           | 256                           | -                             | 0.19      | 0.88          |
| 2                   | 100    | 128                           | 256                           | -                             | 0.08      | 0.95          |
| 2                   | 200    | 128                           | 256                           | -                             | 0.11      | 0.95          |
| 2                   | 500    | 128                           | 256                           | -                             | 0.18      | 0.95          |
| 3                   | 50     | 16                            | 64                            | 128                           | 0.09      | 0.95          |
| 3                   | 100    | 16                            | 64                            | 128                           | 0.24      | 0.92          |
| 3                   | 200    | 16                            | 64                            | 128                           | 0.42      | 0.92          |
| 3                   | 500    | 16                            | 64                            | 128                           | 0.60      | 0.92          |
| 3                   | 50     | 16                            | 64                            | 256                           | 0.33      | 0.90          |
| 3                   | 100    | 16                            | 64                            | 256                           | 0.62      | 0.90          |
| 3                   | 200    | 16                            | 64                            | 256                           | 0.78      | 0.92          |
| 3                   | 500    | 16                            | 64                            | 256                           | 0.93      | 0.92          |
| 3                   | 50     | 16                            | 128                           | 256                           | 0.14      | 0.92          |
| 3                   | 100    | 16                            | 128                           | 256                           | 0.34      | 0.92          |
| 3                   | 200    | 16                            | 128                           | 256                           | 0.47      | 0.92          |
| 3                   | 500    | 16                            | 128                           | 256                           | 0.56      | 0.92          |
| 3                   | 50     | 64                            | 128                           | 256                           | 0.15      | 0.90          |
| 3                   | 100    | 64                            | 128                           | 256                           | 0.16      | 0.95          |
| 3                   | 200    | 64                            | 128                           | 256                           | 0.38      | 0.95          |
| 3                   | 500    | 64                            | 128                           | 256                           | 0.46      | 0.95          |

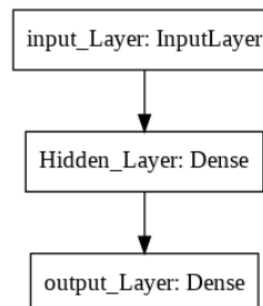
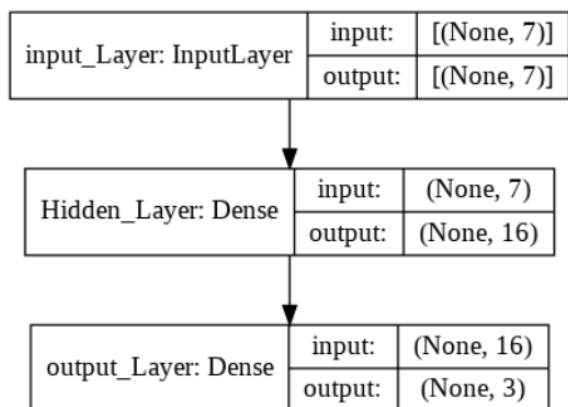
همانطور که از نتایج مشخص است، ساختار شبکه عصبی با یک لایه مخفی که دارای ۱۶ نورون است و با ۲۰۰ بار تکرار آموزش، بهترین دقت را روی داده‌های تست داشته است. همچنین دارای کمترین میزان loss روی داده‌های تست است. بنابراین این ساختار را برای شبکه انتخاب می‌کنیم.

برای اطمینان از اینکه این شبکه دچار بیش برازش نشده باشد نمودار دقت و خطا برای داده‌های آموزشی و validation را رسم می‌کنیم. اگر در این نمودار مشاهده شود که از جایی به بعد با وجود کاهش در خطای روی داده‌های آموزشی، خطا روی داده‌های validation در حال افزایش است، می‌توان نتیجه گرفت که مدل دچار بیش برازش شده است اما با توجه به نمودار زیر اینطور نیست بنابراین این ساختار شبکه را انتخاب می‌کنیم.

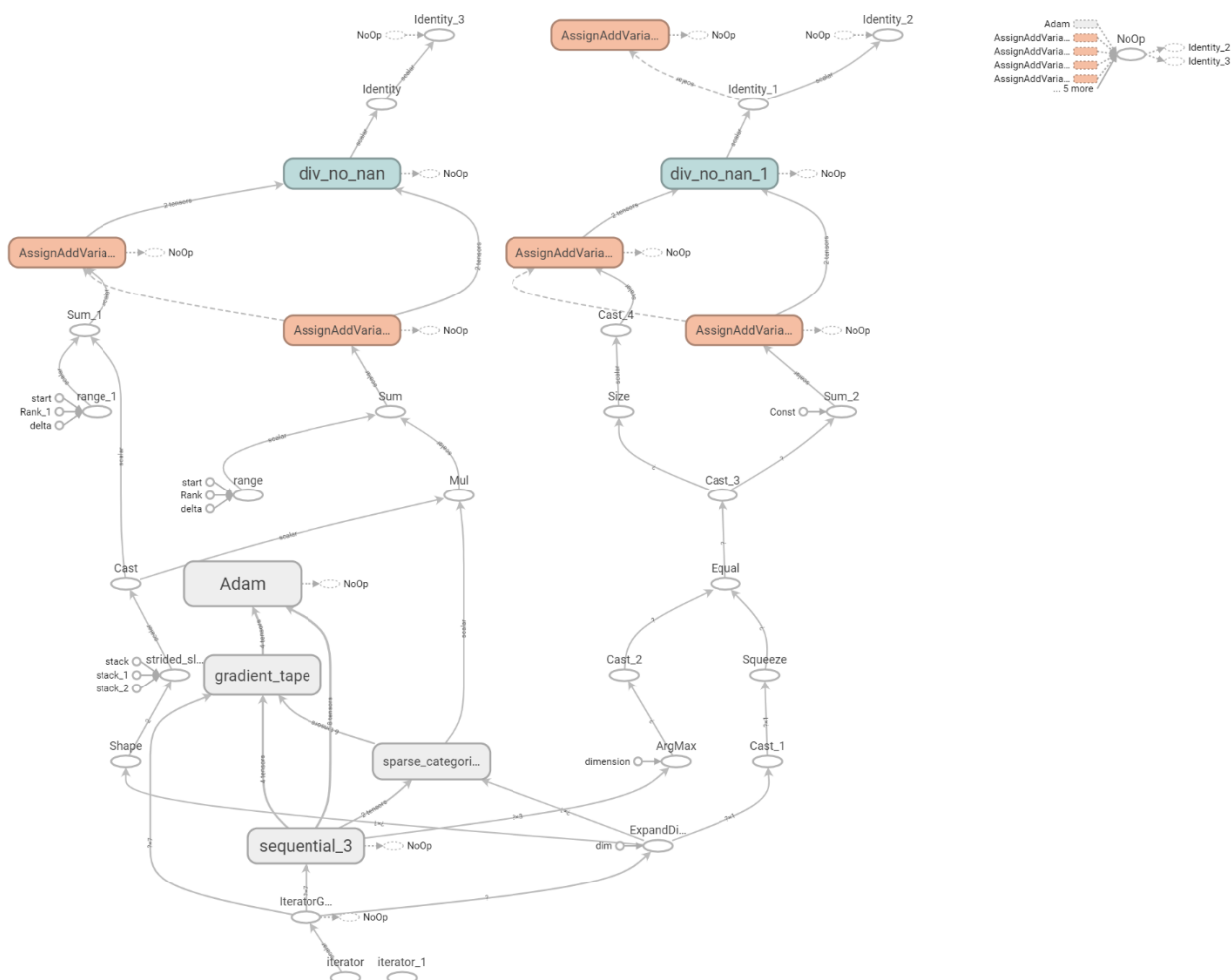


بنابراین ویژگی‌های مدل انتخاب شده به صورت خلاصه این موارد هستند:

- ۷ نورون در لایه ورودی
- یک لایه مخفی با ۱۶ نورون
- لایه خروجی با ۳ نورون
- تابع خطای SparseCategoricalCrossentropy
- بهینه‌ساز Adam با نرخ یادگیری 0.01

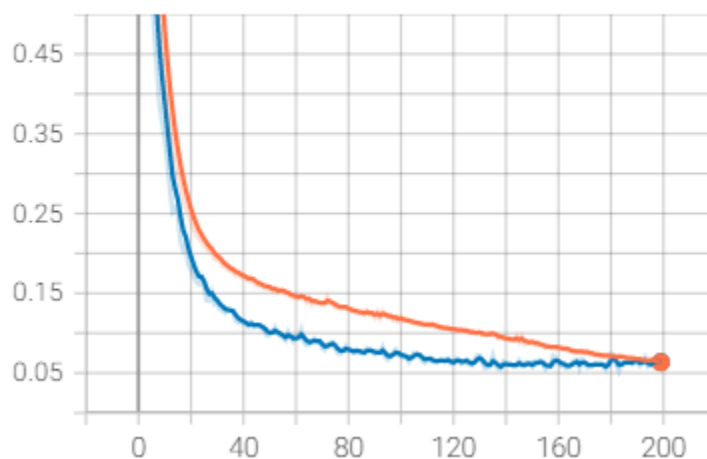


## گراف شبکه:



نمودار epoch-loss رسم شده در تانسوربرد:

epoch\_loss  
tag: epoch\_loss



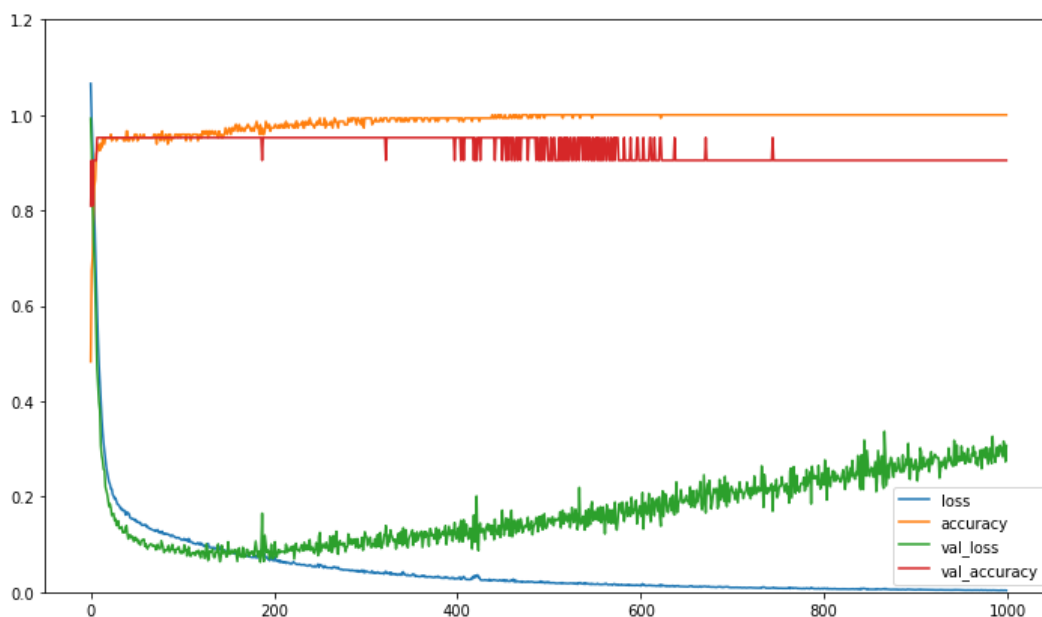
Model: "sequential\_3"

| Layer (type)            | Output Shape | Param # |
|-------------------------|--------------|---------|
| Hidden_Layer (Dense)    | (None, 16)   | 128     |
| output_Layer (Dense)    | (None, 3)    | 51      |
| Total params: 179       |              |         |
| Trainable params: 179   |              |         |
| Non-trainable params: 0 |              |         |

## سوال چهارم

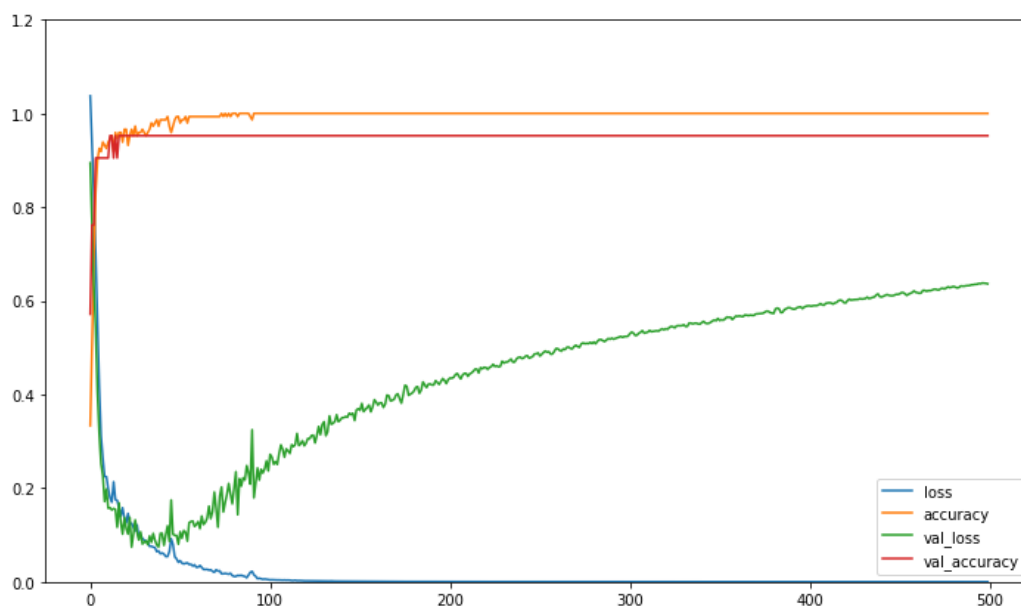
دو راه حل برای بیش برآزش کردن مدل وجود دارد:

- ۱- یک راه این است که تعداد مراحل آموزش (epoch) را زیاد کنیم. برای این منظور، برای همان شبکه‌ای که در قسمت ۳ و ۲ استفاده کردیم، تعداد epochها را از ۲۰۰ به ۱۰۰۰ افزایش می‌دهیم. نمودار آن به این صورت است:



همانطور که مشاهده می‌شود در فرآیند آموزش تا تعداد مشخصی epoch که جلو می‌رویم loss هم برای دیتای آموزش و هم برای دیتای validation کاهش پیدا می‌کند اما از یک جایی به بعد با افزایش epoch ها loss در دیتای validation شروع به افزایش می‌کند در حالی که loss در دیتای آموزش همچنان در حال کاهش است. این مورد نشانه‌ی واضحی برای بیش برآزش است. نشانه‌ی دیگر این است که دقت روی داده‌های تست نسبت به مدلی که تعداد epoch کمتری آموزش را تکرار می‌کند، کاهش پیدا می‌کند.

۲- راه دیگری که برای بیش برآزش کردن مدل می‌تواند انجام بگیرد افزایش پیچیدگی مدل است. این کار می‌تواند با افزایش تعداد لایه‌های مخفی و افزایش تعداد نورون‌ها در هر لایه مخفی صورت بگیرد. به این منظور تعداد لایه‌های مخفی را از ۱ به ۴ افزایش داده و در هر کدام از لایه‌های مخفی به جای ۱۶ نورون ۱۲۸ نورون قرار می‌دهیم. نمودار آن به صورت زیر است:



در اینجا هم مانند حالت قبل از یک جایی به بعد با افزایش epoch ها loss در دیتای validation شروع به افزایش می‌کند در حالی که loss در دیتای آموزش همچنان در حال کاهش است، هم چنین دقت روی داده‌های تست نسبت به مدل ساده‌تر، کاهش پیدا می‌کند بنابراین بیش برآزش داریم.

## سوال پنجم

روش‌های متعددی برای انجام منظم‌سازی در تنسورفلو وجود دارد که در اینجا به معرفی ۵ تکنیک می‌پردازیم:

1. L1 Regularization

$$L_1 = (wx + b - y)^2 + \lambda|w|$$

2. L2 Regularization

$$L_2 = (wx + b - y)^2 + \lambda w^2$$

3. L1\_L2 Regularization

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}} (\|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1)$$



#### 4. Dropout

تعدادی از اتصالات بین نورون ها را به صورت رندوم حذف می کند.

#### 5. Batch Normalization

روشی است که برای سریع تر و پایدارتر کردن شبکه های عصبی مصنوعی از طریق عادی سازی ورودی های لایه ها با مرکزیت مجدد و مقیاس گذاری مجدد استفاده می شود.

در جدول زیر مزایا و معایب هر یک از این روش ها بررسی شده است:

### منظم سازی L1

#### مزایا:

- ۱- وقتی تعداد فیچرهای داده زیاد باشد خوب عمل می کند.
- ۱- در مواجهه با نقاط outlier در داده های ورودی، عملکرد بهتر و مقاوم تری از L2 دارد
- ۲- مدلی تولید می کند که ساده و قابل تفسیر است و تنها زیرمجموعه ای از ویژگی های داده های ورودی را استفاده می کند.
- ۲- با استفاده از آن، خاصیت feature selection حاصل می شود زیرا می تواند فیچرهای غیرمهم را در نظر نگیرد.

#### معایب:

- ۱- سبب می شود تا الگوریتم قادر نباشد الگوهای پیچیده را یاد بگیرد.
- ۲- در مورد دیتاست هایی که ابعاد زیادی دارند (تعداد ابعاد از تعداد داده ها بیشتر است) خوب عمل نمی کند.

### منظم سازی L2

#### مزایا:

- ۱- در دیتاست هایی که تمام فیچرها روی متغیر هدف تاثیرگذارند خوب جواب می دهد.
- ۲- زمانی که وزن های داخل مدل به طور مساوی مقداردهی اولیه شده باشند خوب جواب می دهد.
- ۳- زمانی که مقدار هدف پیش بینی تابعی از همه ویژگی ها داده ورودی باشد، نرم L2 یادگیری بهتری را سبب می شود.
- ۳- با نرم L2 بر خلاف نرم L1 می توان الگوهای پیچیده را در داده ورودی یاد گرفت.
- ۴- دقت های بالاتری را ارائه می دهد و بهتر تنظیم می شود.

#### معایب:

- ۴- در قبال وجود داده های outlier در دیتاست مورد استفاده خوب عمل نمی کند؛ چرا که در نقاط outlier خطای پیش بینی مدل بسیار زیاد می شود و با داشتن جمله پنالتی نرم L2 در آن تابع، وزن های مدل کوچکتر خواهد شد.
- ۱- یک مدل خلوت ایجاد نمی کند و پیچیدگی و ابعاد مدل بالاست.
- ۲- مدل حاصل از آن، تفسیرپذیری کمی دارد.

## منظم سازی L1\_L2

### مزایا:

- ۱- مزایای منظم سازی های L1 و L2 را با هم دارد.
- ۲- تاثیر فیچرهای مختلف را کم می کند در حالی که همه آن را حذف نمی کند.
- ۳- معمولاً به L1 و L2 ترجیح داده می شود چون معایب آن ها را حذف می کند.

### معایب:

- ۱- هزینه محاسباتی بیشتری نسبت به L1 و L2 دارد.
- ۲- انعطاف پذیری کمتری دارد.

## منظم سازی Dropout

### مزایا:

- ۱- از بهینه سازی همزمان وزن تمام نورون های یک لایه جلوگیری می کند. این انطباق، که در گروه های تصادفی انجام می شود، از همگرایی همه نورون ها به یک هدف جلوگیری می کند و در نتیجه وزن ها را به هم مرتبط می کند.
- ۲- افزایش استحکام مدل و همچنین حذف هر گونه وابستگی ساده بین نورون ها.
- ۳- فعال شدن واحدهای پنهان پراکنده می شود که این نیز یک ویژگی مطلوب است.
- ۴- وزن ها در میان تمام ویژگی ها تقسیم می شود و مقادیر کمتری برای آن ها در نظر گرفته می شود و این امر تأثیر بسزایی در نظم دهی و یک دست شدن مدل دارد.

### معایب:

- ۵- برای شبکه های کانولوشنی خوب عمل نمی کند.
- ۶- هنگامی که شبکه نسبت به مجموعه داده کوچک است عملکرد مناسبی ندارد.
- ۷- اگر آموزش کوتاه باشد و تا زمان همگرایی ادامه پیدا نکند روش dropout خوب عمل نمی کند.

## منظم سازی Batch Normalization

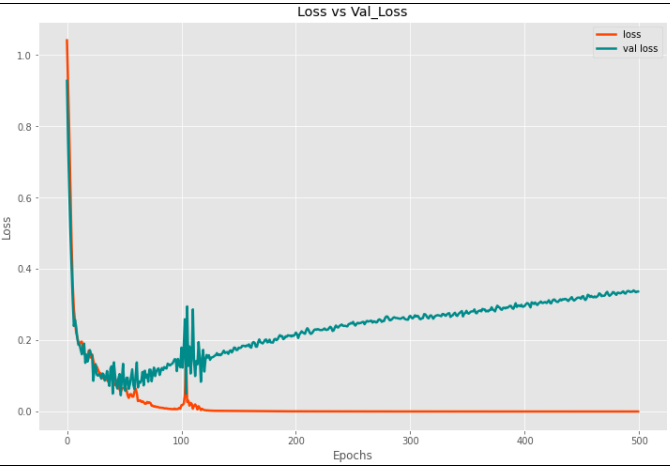
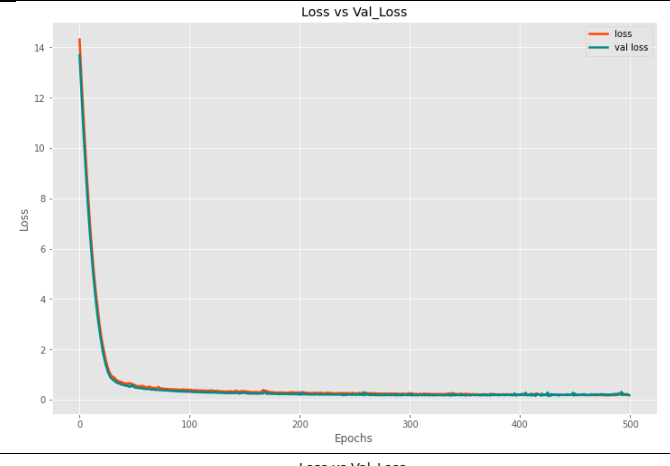
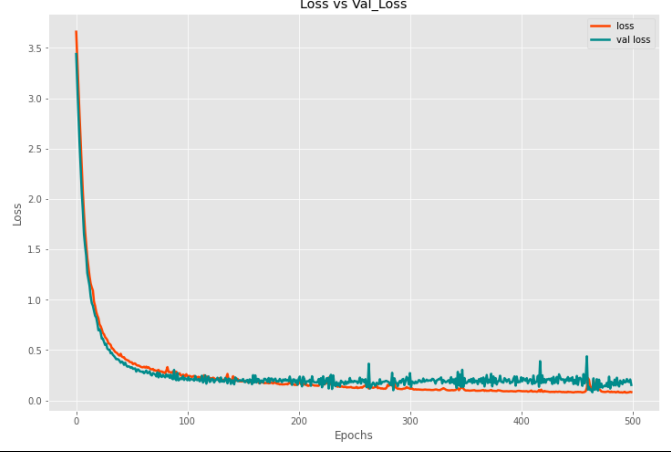
### مزایا:

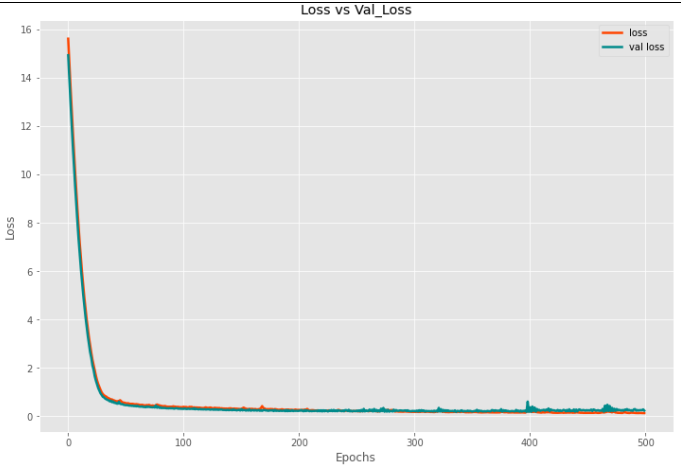
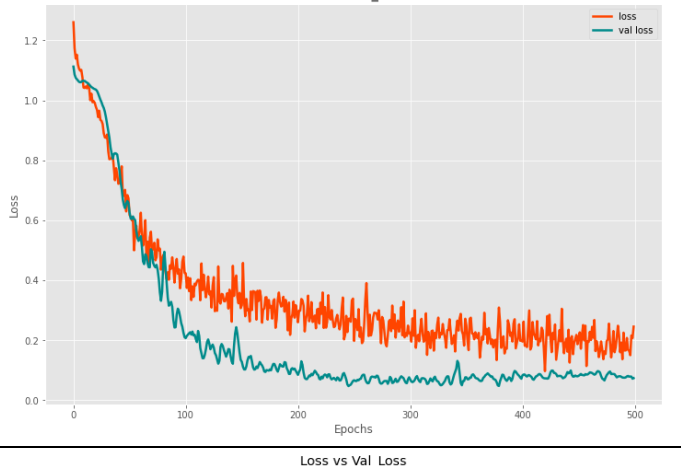
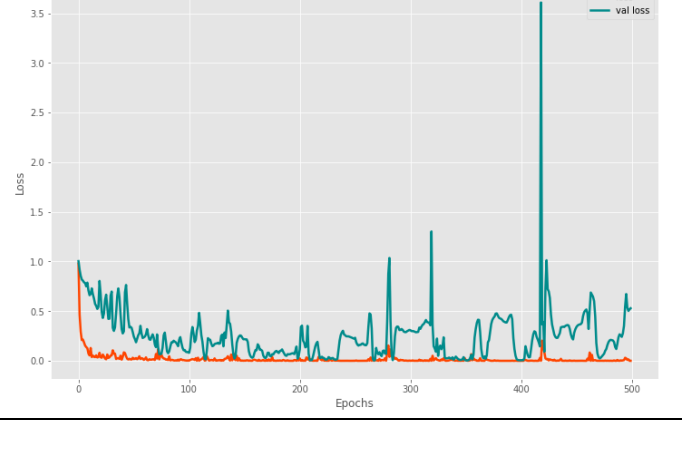
- ۱- اگر سایز batch زیاد باشد پایدار است.
- ۲- سرعت آموزش را افزایش می دهد.
- ۳- نسبت به scale و شیفت در دیتای ورودی مقاوم است.
- ۴- نسبت به تغییرات بردار وزن مقاوم است.

### معایب:

- ۱- محاسبات سنگینی دارد و حافظه زیادی اشغال می کند زیرا تمام اطلاعات آماری مربوط به دسته ها باید ذخیره شوند.
- ۲- عملکرد آن به اندازه batch (دسته) حساس است.
- ۳- برای یادگیری بر خط مناسب نیست.

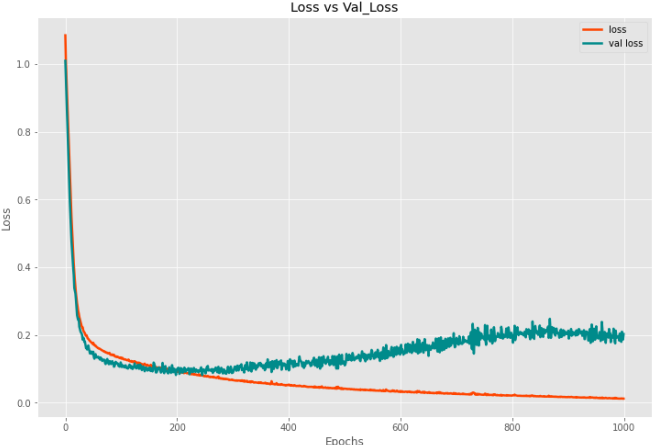
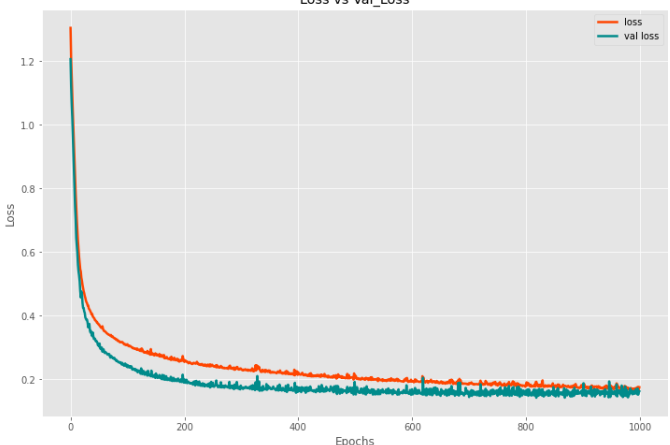
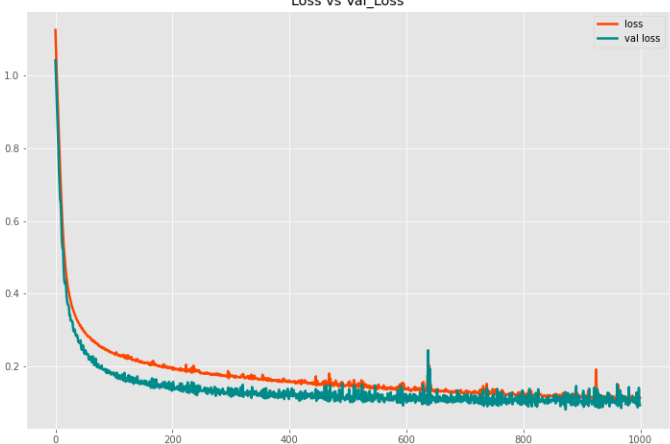
در جدول زیر مقدار دقت و خطا روی مجموعه داده تست به همراه نمودار خطای آموزش و validation به ازای اعمال هر کدام از این منظم سازی ها بر روی مدل بیش برازش شده حاصل از پیچیده سازی مدل، گزارش شده است:

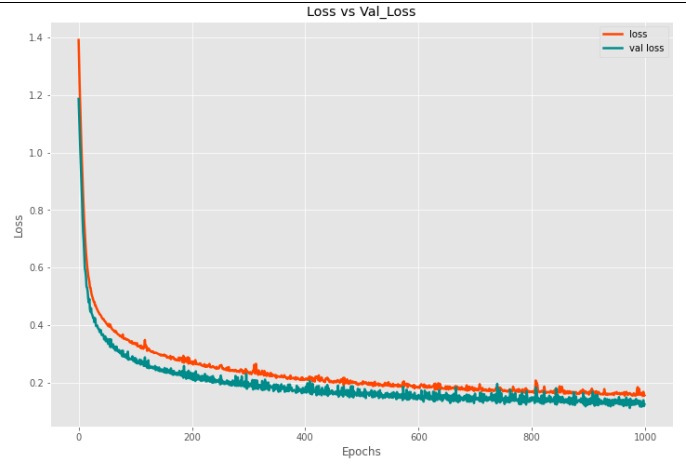
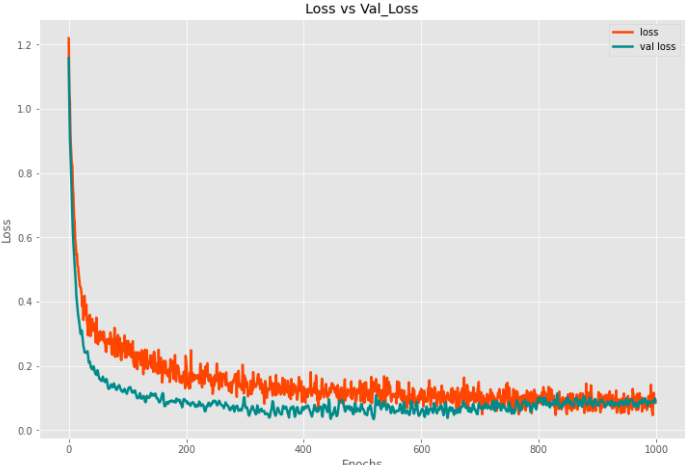
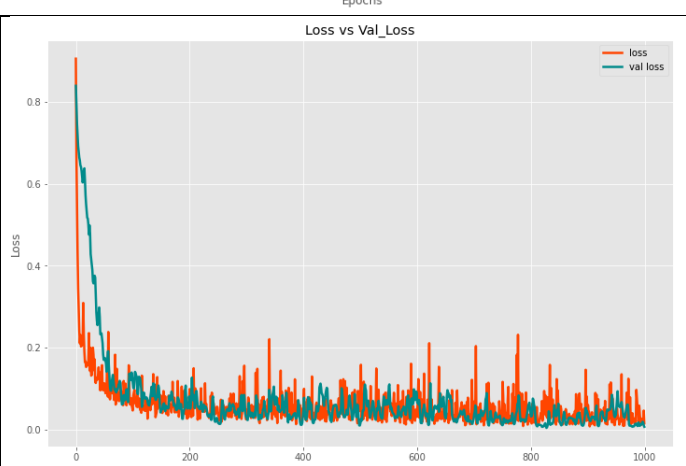
| نمودار loss و val-loss   | خطا روی داده تست | دقت روی داده تست | منظم سازی  |
|--|------------------|------------------|--|
|    | 0.29             | 0.95             | بدون اعمال منظم سازی<br><br>(مدل بیش برازش شده با استفاده از افزایش پیچیدگی مدل) |
|   | 0.22             | 0.95             | L1 Regularization  |
|  | 0.16             | 0.97             | L2 Regularization  |

|                                 |             |             |  |
|---------------------------------|-------------|-------------|--|
| <b>L1_L2<br/>Regularization</b> | <b>0.97</b> | <b>0.22</b> |     |
| <b>Dropout</b>                  | <b>0.92</b> | <b>0.12</b> |   |
| <b>Batch<br/>Normalization</b>  | <b>0.90</b> | <b>0.36</b> |  |

همانطور که مشاهده می‌شود، منظم‌سازی های **L1\_L2** و **L2** عملکرد بهتری داشته اند زیرا هم دقت را افزایش داده‌اند و هم اینکه خطا روی داده‌های تست را کاهش داده‌اند. نمودار مربوط به **loss** و **val\_loss** نیز برای این دو مورد مناسب تر از سایر منظم‌سازی هاست و نویز کمتری دارد. تغییرات **val\_loss** در نمودارهای مربوط به این دو به صورت کاهشی است و با شیب خوبی کاهش پیدا کرده است. هم‌چنین تغییرات ناگهانی ندارد.

در جدول زیر مقدار دقت و خطا روی مجموعه داده تست به همراه نمودار خطای آموزش و validation به ازای اعمال هر کدام از این منظم سازی ها بر روی مدل بیش برآزش شده حاصل از افزایش تعداد مراحل آموزش مدل، گزارش شده است:

| نمودار loss و val-loss   | خطا روی داده تست | دقت روی داده تست | منظم سازی  |
|--|------------------|------------------|--|
|    | 0.06             | 0.95             | بدون اعمال منظم سازی<br><br>(مدل بیش برآزش شده با استفاده از افزایش مراحل آموزش) |
|   | 0.19             | 0.97             | L1 Regularization  |
|  | 0.13             | 0.95             | L2 Regularization  |

|                                 |             |             |  |
|---------------------------------|-------------|-------------|--|
| <b>L1_L2<br/>Regularization</b> | <b>0.97</b> | <b>0.16</b> |     |
| <b>Dropout</b>                  | <b>0.95</b> | <b>0.12</b> |   |
| <b>Batch<br/>Normalization</b>  | <b>0.95</b> | <b>0.14</b> |  |

همانطور که مشاهده می‌شود، منظم‌سازی های  $L1$  و  $L1\_L2$  عملکرد بهتری داشته اند زیرا دقت روی داده‌های تست را افزایش. نمودار مربوط به  $loss$  و  $val\_loss$  نیز برای این دو مورد مناسب تر از سایر منظم‌سازی هاست و نویز کمتری دارد. تغییرات  $val\_loss$  در نمودارهای مربوط به این دو به صورت کاهشی است و با شیب خوبی کاهش پیدا کرده است. همچنین تغییرات ناگهانی ندارد.

به طور کلی منظم‌سازی  $L1\_L2$  چون مزایای منظم‌سازی  $L1$  و  $L2$  را با هم ترکیب می‌کند روش مناسبی است و به خوبی عمل می‌کند.