



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس تحلیل و طراحی شبکه های عصبی عمیق

تمرین شماره 3

نام و نام خانوادگی : علی رشیدی مقدم | کیوان کاظمی

شماره دانشجویی : 810100357 | 810100443

بهمن ماه 1401

3 مقدمه
5 سوال اول:
7 سوال دوم:
9 سوال سوم:
11 نتیجه گیری:

در این تمرین، تابع "شاخص جداپذیری مبتنی بر مرکز" (CSI) پیاده سازی و مقدار آن در لایه های مختلف شبکه ها محاسبه می شود. سپس ارتباط این معیار با معیار هایی همچون دقت و خطا و عملکرد شبکه ها بررسی می شود.

برای محاسبه CSI، در یک مجموعه از داده با توزیع یکسان هر کلاس، ابتدا مرکز هر کلاس (میانگین بردار ویژگی اعضای هر کلاس) محاسبه شده و سپس برای هر داده فاصله اقلیدسی آن تا مرکز کلاس ها محاسبه می شود. چنانچه نزدیک ترین فاصله مربوط به کلاس خودش باشد $SI=1$ و در غیر اینصورت $SI=0$ می باشد. نهایتاً از SI تمام داده ها میانگین گرفته می شود و مقدار آن به عنوان شاخص جداپذیری مبتنی بر مرکز گزارش می شود. روابط ریاضی CSI در زیر آمده است.

$Data = \{(x_i, l_i)\}_{i=1}^m \forall i: x_i \in R^{n \times 1} \quad l_i \in \{1, 2, \dots, n_c\} \quad n_c: \text{number of classes}$

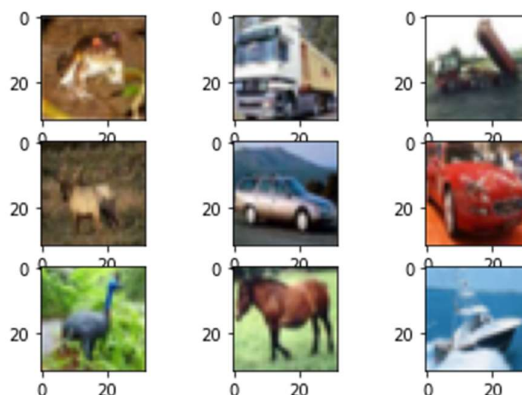
Center of each class is the mean of all input data points having the label of that class:

$$\mu_c = \frac{1}{m_c} \sum_{i=1}^m x_i \delta(l_i, c), \quad c = 1, 2, \dots, n_c \quad m_c = \sum_{i=1}^m \delta(l_i, c)$$

$$CSI(Data) = \frac{1}{m} \sum_{i=1}^m \delta(l_i, c^*)$$

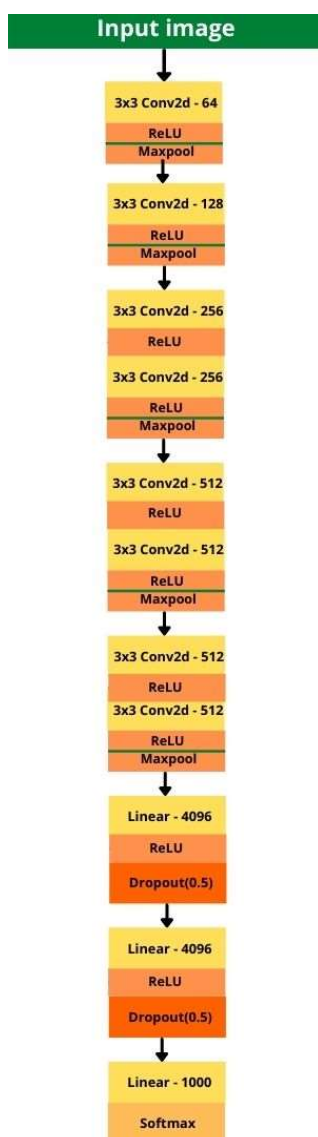
$$c^* = \arg \min_{\forall c} \|x_i - \mu_c\|$$

دادگان مورد استفاده در این تمرین CIFAR10 می باشد که تعدادی تصویر نمونه از آن در زیر مشاهده می شود.



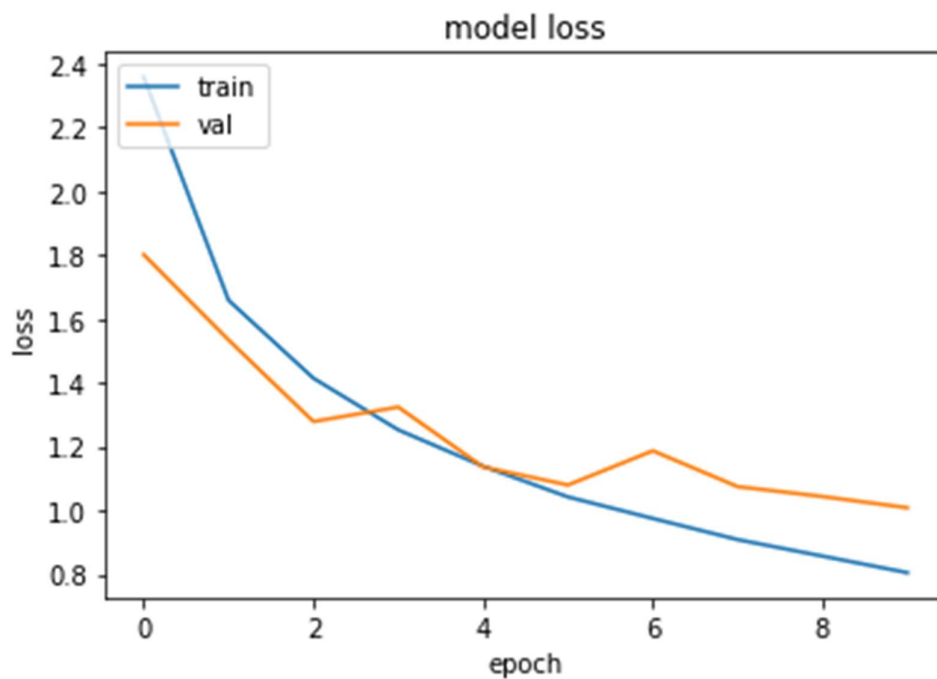
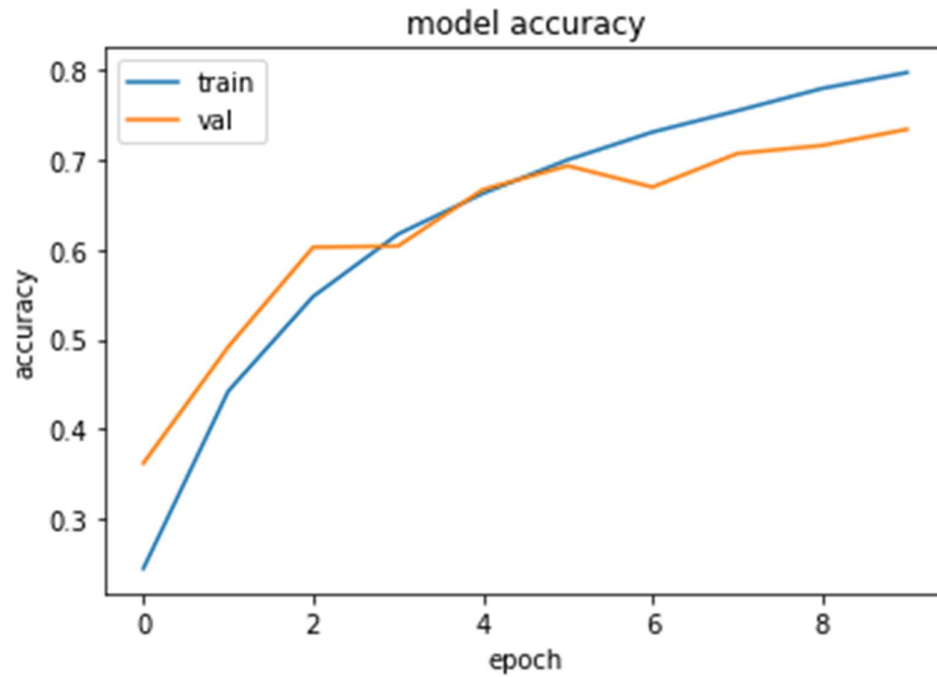
برای آموزش شبکه ها از 50000 داده استفاده می شود که 33 درصد آن به بخش ارزیابی اختصاص می یابد. همچنین برای محاسبه CSI ، 5000 داده از داده آموزش انتخاب می شود که هرکلاس 500 نماینده در آن دارد.

شبکه مورد استفاده در این تمرین شبکه vgg11 است که معماری آن در زیر آمده است.

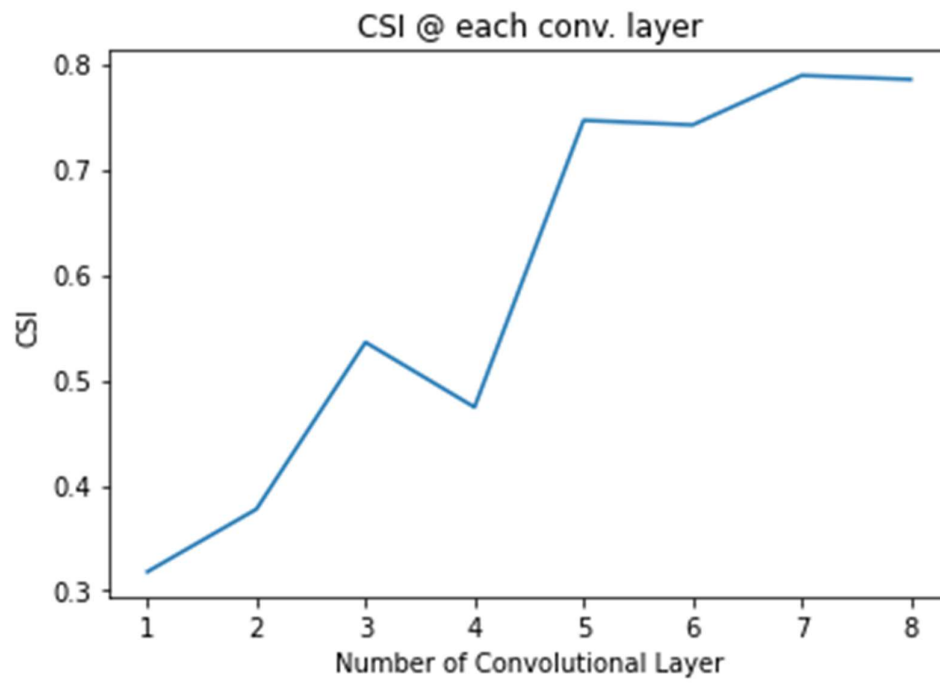


سوال اول :

ابتدا شبکه vgg11 بر روی داده‌گان ذکر شده برای 10 اپاک آموزش داده می‌شود. نمودار دقت و خطا به شکل زیر می‌باشد.



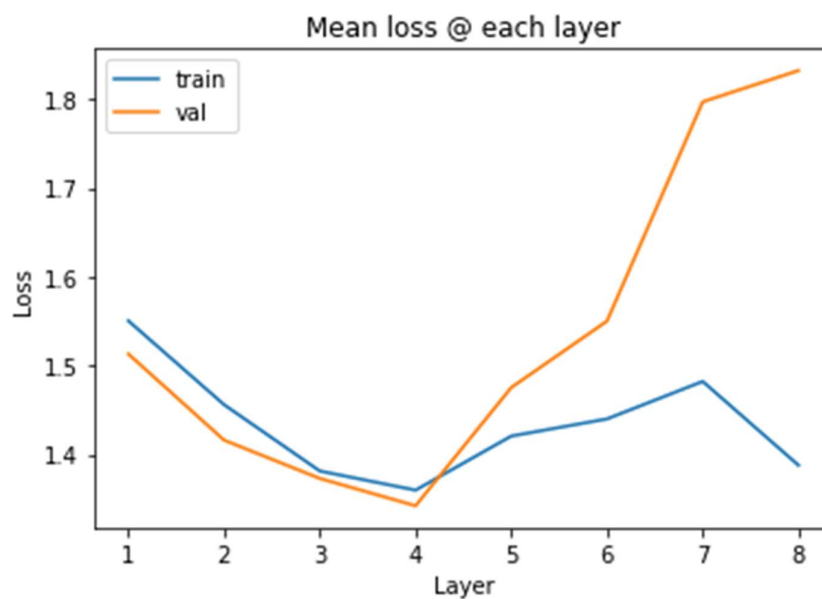
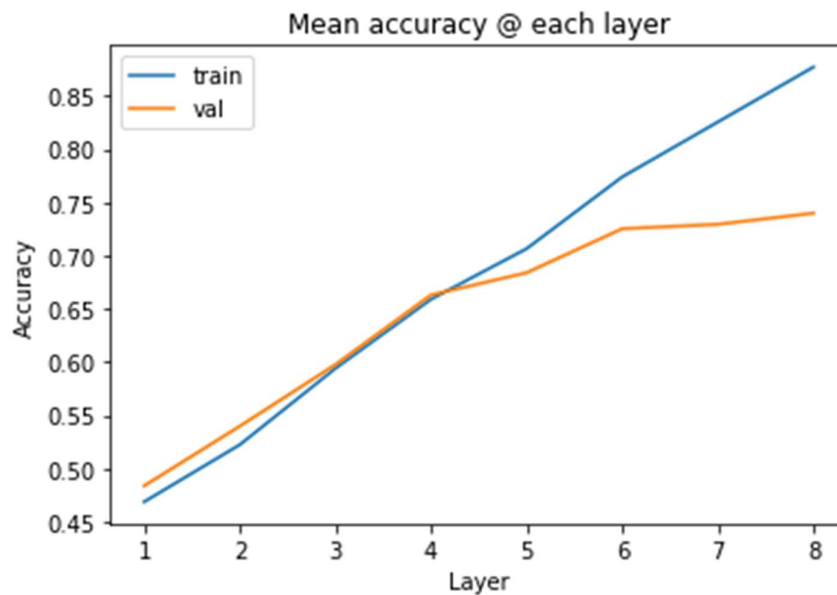
پس از آموزش شبکه، معیار CSI را برای هر لایه کانولوشنی محاسبه کرده و سپس نمودار آن رسم می شود.



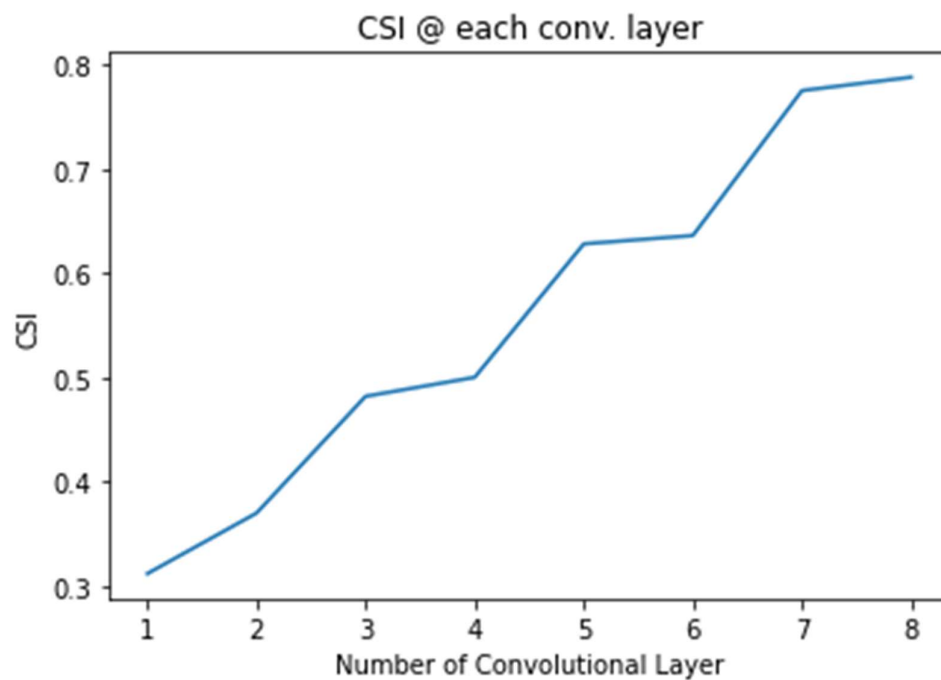
همانطور که از نمودار پیداست معیار جداپذیری در طول شبکه روند صعودی دارد.

سوال دوم :

در این قسمت شبکه vgg11 بصورت لایه به لایه آموزش می بیند. هر لایه از شبکه برای 5 اپیاک آموزش می بیند. سپس پارامترهای آن لایه منجمد شده و لایه جدید و یک FC با 10 نورون به مدل افزوده می شود و مجدداً برای 5 اپیاک عملیات آموزش صورت می گیرد. دقت و خطای هر لایه از میانگین دقت و خطای آن لایه در طول آموزش حاصل می شود. نمودار دقت و خطای هر لایه در زیر آمده است.



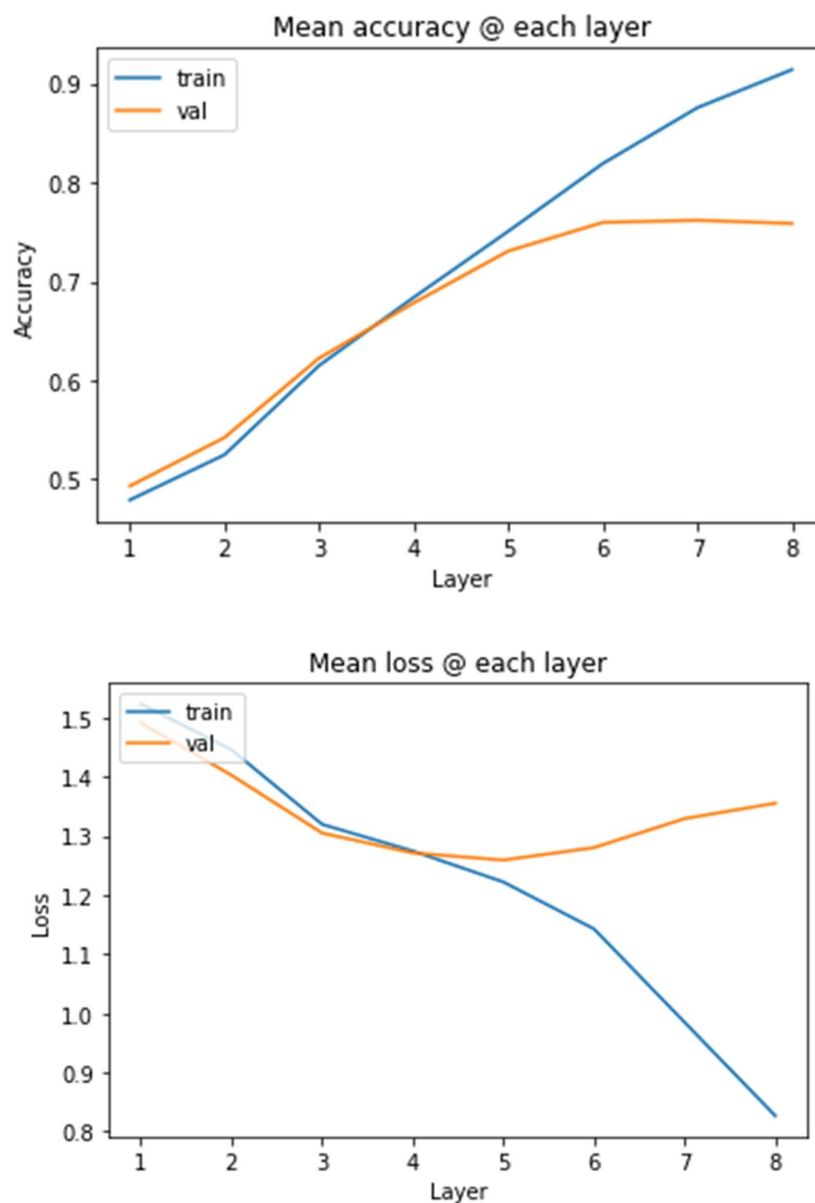
پس از پایان یافتن آموزش شبکه ، شاخص جدایی برای لایه های کانولوشنی محاسبه می شود. نتیجه به صورت زیر است:



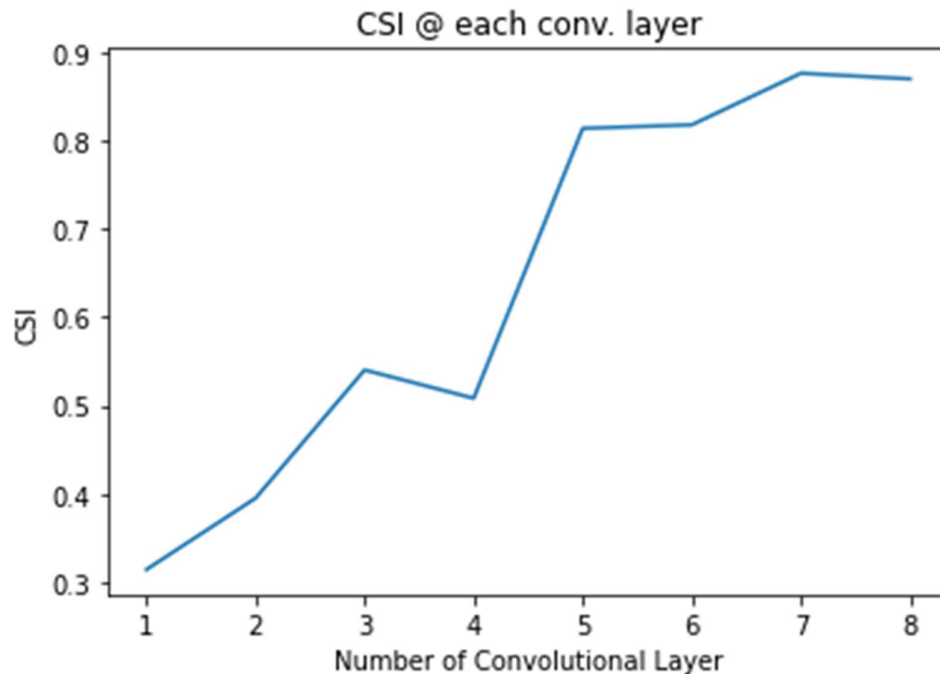
در این نمودار برخلاف نمودار قسمت الف، در تمام لایه ها شاخص جدایی در حال افزایش است. گویی آموزش لایه به لایه شبکه، با افزایش CSI هر لایه هم ارز است. اما این شبکه نسبت به شبکه الف، عمومیت پذیری و قابلیت پیش بینی ضعیف تری دارد چون دقت آن روی داده ارزیابی نسبت به قسمت الف کمتر است. در واقع در اینجا با جبر پارامتری، CSI افزایش یافته است و منجر به اورفیتینگ شده است.

سوال سوم :

در این قسمت همانند قسمت ب عمل می شود با این تفاوت که پارامترهای لایه های قدیمی منجمد نشد و آن ها نیز دچار تغییر و آموزش می شوند. در این قسمت رفته رفته حجم محاسبات و زمان آموزش شبکه ها بیشتر می شود. در اینجا نیز هر مرحله از آموزش برای 5 اپیاک انجام می شود. نمودار دقت و خطا در زیر آمده است:



پس از اتمام آموزش شبکه، شاخص جداپذیری لایه های کانولوشنی محاسبه می شود و نتیجه به صورت زیر است:



نمودار های این قسمت بسیار به قسمت الف شبیه است. در اینجا نیز CSI روند صعودی دارد اما لزوما در تمام لایه ها میزان آن نسبت به لایه قبل افزایش نیافته است. قدرت پیش بینی و عمومیت پذیری این شبکه از قسمت ب بیشتر است و شبیه به قسمت الف است. همچنین مقدار شاخص جداپذیری این شبکه در لایه آخر از تمام شبکه ها بیشتر است. در این شبکه چون آموزش لایه به لایه صورت می گیرد و همچنین می توان تمام لایه ها را آپدیت کرد، این امکان وجود دارد تا با یک دید کلی به جزیی شبکه آموزش یابد و از تعداد پارامتر کم به سوی تعداد پارامتر زیاد با پایداری بهتری حرکت کند و بهینه سازی بهتری صورت گیرد.

نتیجه گیری :

در این تمرین شاخص جداپذیری در لایه های کانولوشنی یک شبکه طبقه بند بررسی شد. در یک شبکه قدرتمند این شاخص بطور کلی روند افزایشی دارد. هرچند ممکن است در برخی لایه ها شاخص جداپذیری کاهش یابد. همچنین نشان داده شد که اصرار بیش از حد بر افزایش شاخص جداپذیری با جبر پارامتری می تواند منجر به کاهش قدرت پیش بینی شبکه گردد. برای مقایسه بهتر، نمودار تمام قسمت ها بطور یکجا در زیر آمده است.

