

شبکه عصبی تمرین سوم

سحر محمدی ۴۰۲۱۳۹۰۱۰۹

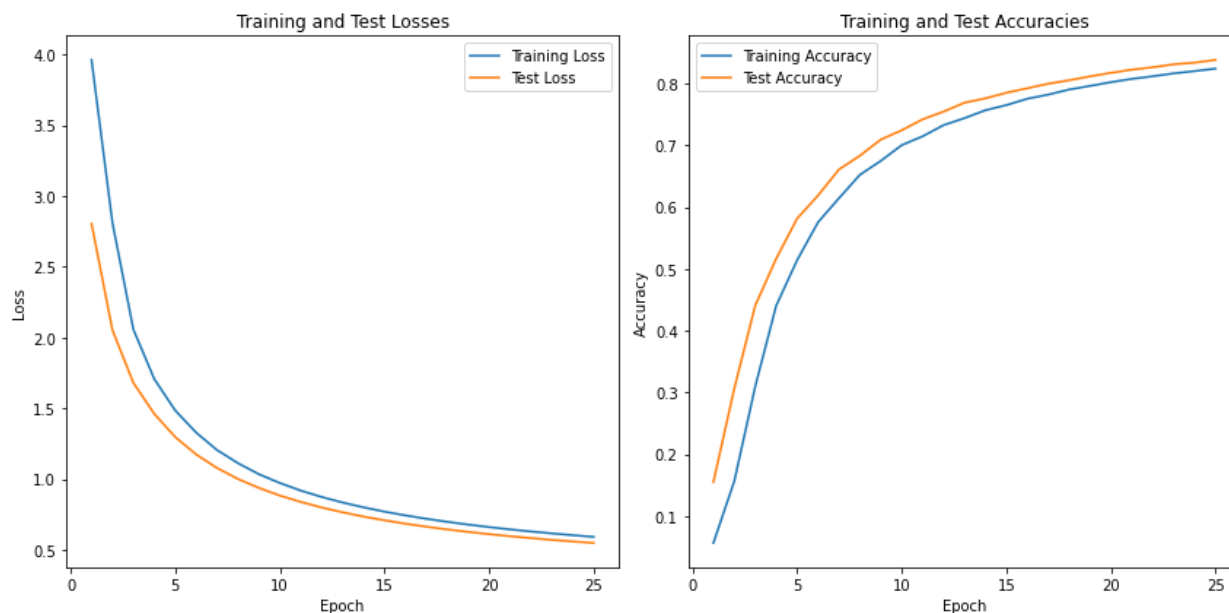
هدف این تمرین حل یک مسئله طبقه بندی چند کلاسه برای داده‌های MNIST به کمک شبکه MLP است. به همراه بررسی تاثیر پیش پردازش داده‌ها و چند روش Regularization.

شرح پیاده سازی

طبق موارد خواسته شده، پیاده سازی انجام شد طوری که در لایه آخر از تابع فعالسازی softmax استفاده شود و برچسب‌ها هم به بردارهای one-hot تبدیل شدند.

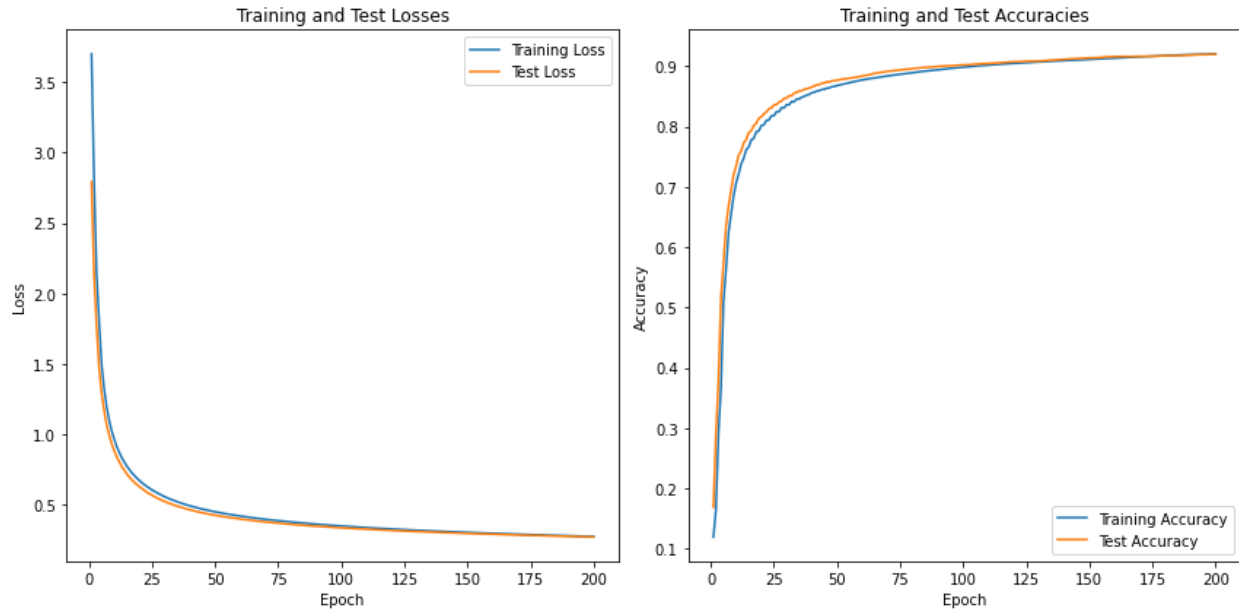
آزمایشات

۱. حل طبقه بندی چند کلاسه با شبکه که یک لایه مخفی با ۱۲۸ نرون و تابع فعال سازی Relu و در لایه خروجی هم تابع فعال سازی softmax باشد. با نرخ یادگیری ۰.۰۰۰۱ و ۲۰ اپوک.



نمودار سمت چپ میزان ضرر روی داده‌های آموزشی و تست و نمودار سمت راست دقت روی داده‌های آموزشی و تست را نشان می‌دهند. شبکه با دقت 0.8383 داده‌های تست را طبقه بندی می‌کند.

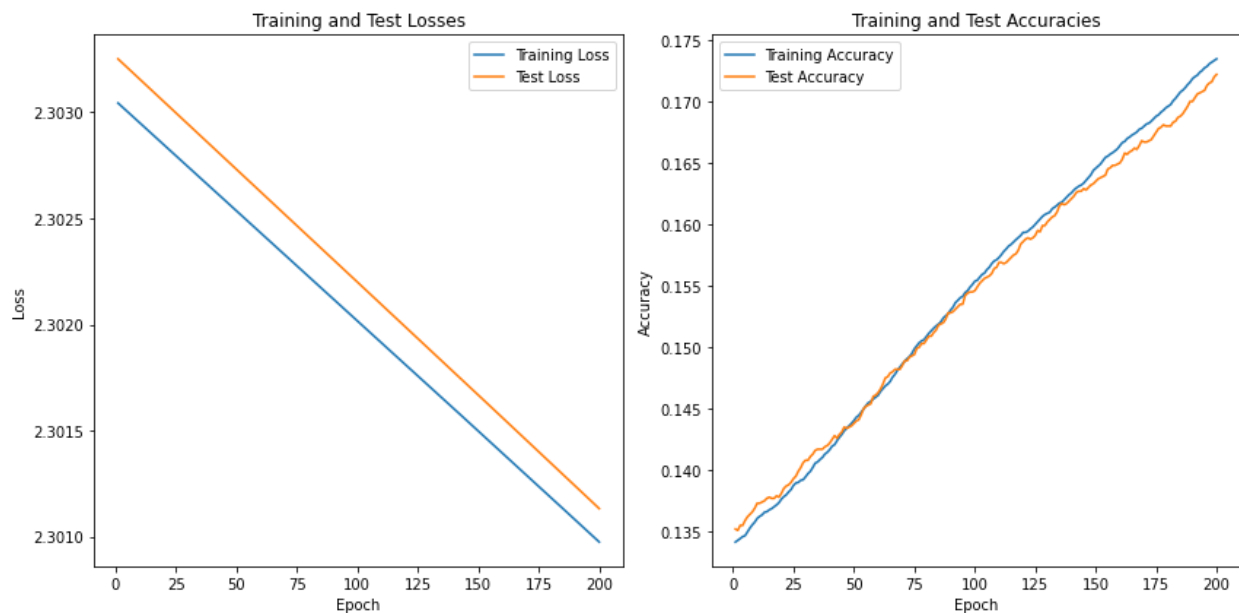
که البته با ۲۰۰ اپوک به دقت 0.9198 می‌رسد.



مشاهده می‌شود که بعد از ۲۵ اپوک دقت با شیب کمی افزایش پیدا می‌کند.

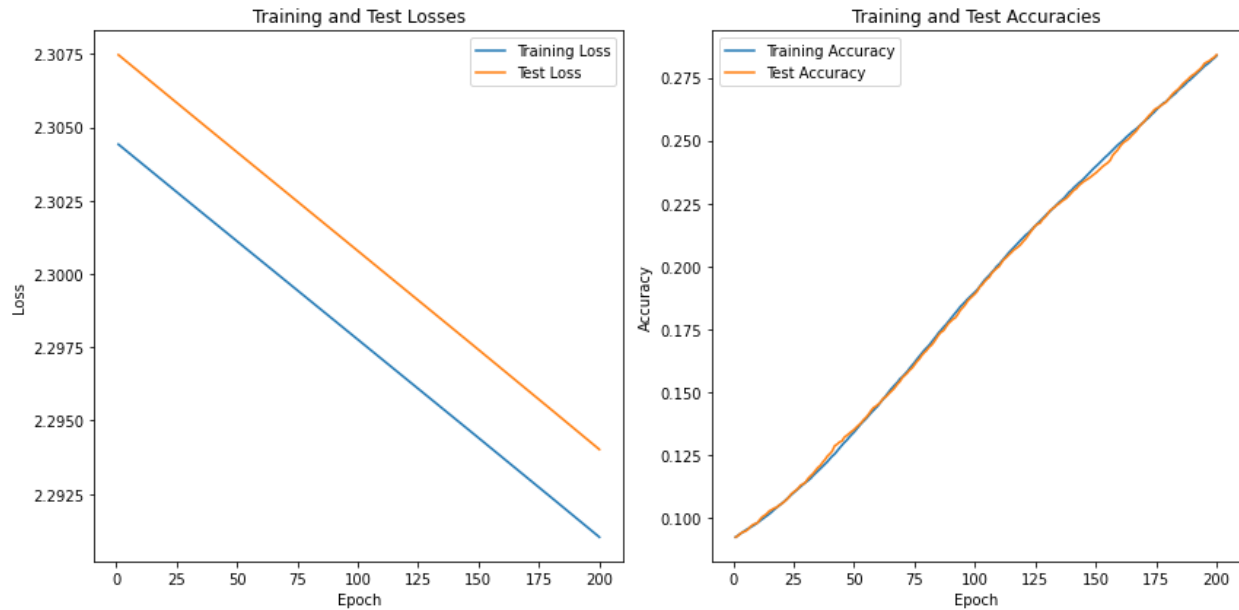
۲. بررسی تاثیر پیش پردازش

- ابتدا فقط normalization به گونه‌ای که داده تقسیم بر ابعادشون یعنی ۲۵۵ شوند. که نتیجه مثبتی نداشت. دقت با ۲۰۰ اپوک فقط به ۱۷ درصد رسید.



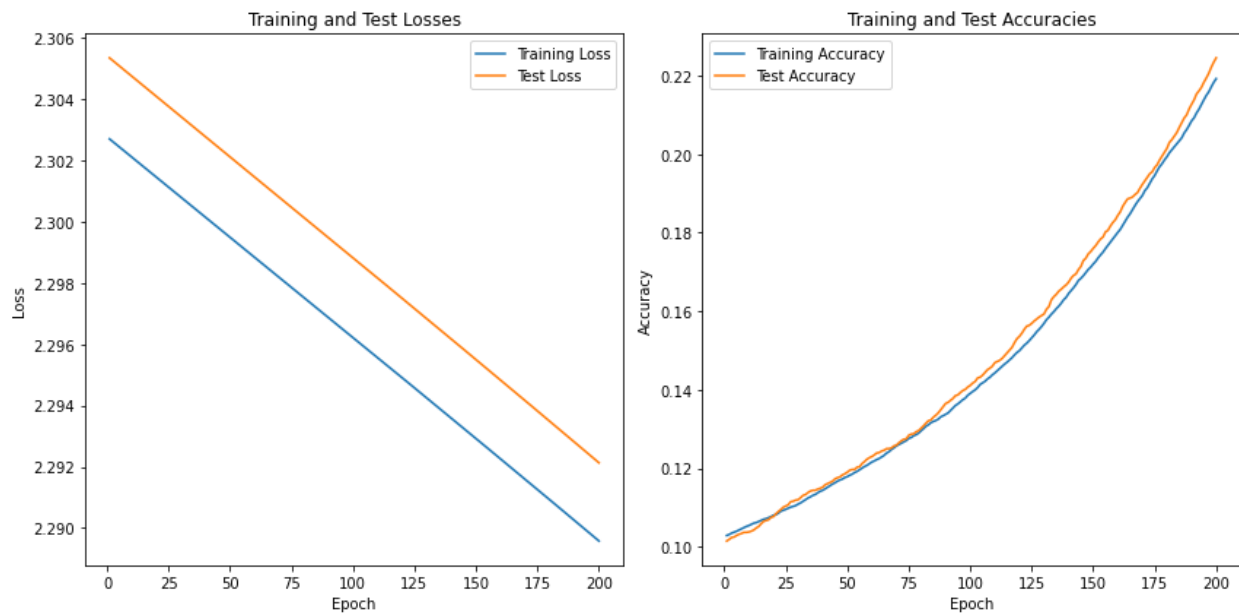
- در مرحله بعد فقط standardization به گونه‌ای که داده‌ها منهای میانگین تقسیم بر انحراف معیار شوند. برای اینکه تقسیم بر صفر رخ ندهد یک اپسیلون به محرج اضافه شد.

دقت با ۲۰۰ اپوک به ۲۸ درصد رسید.



• و در مرحله سوم استفاده از نرمال سازی و استاندارد کردن باهم.

که به دقت ۲۲ درصد رسید.



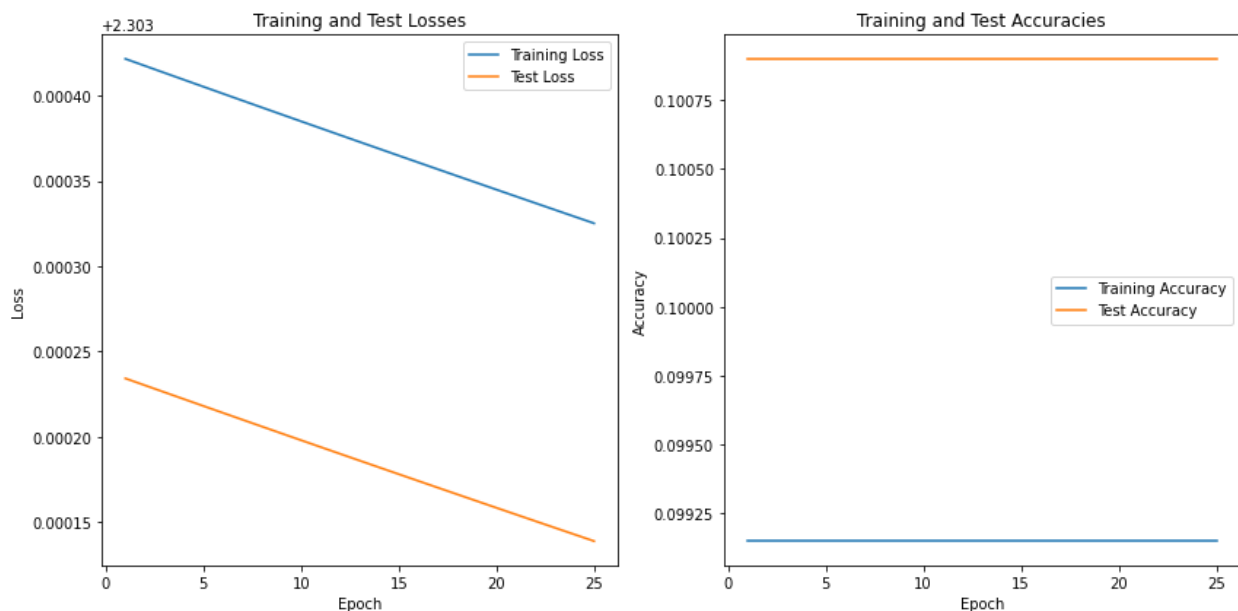
پس نتیجه می گیریم برای این داده ها و این شبکه بدون پیش پردازش به نتیجه بهتری خواهیم رسید.

۳. بررسی vanishing gradient

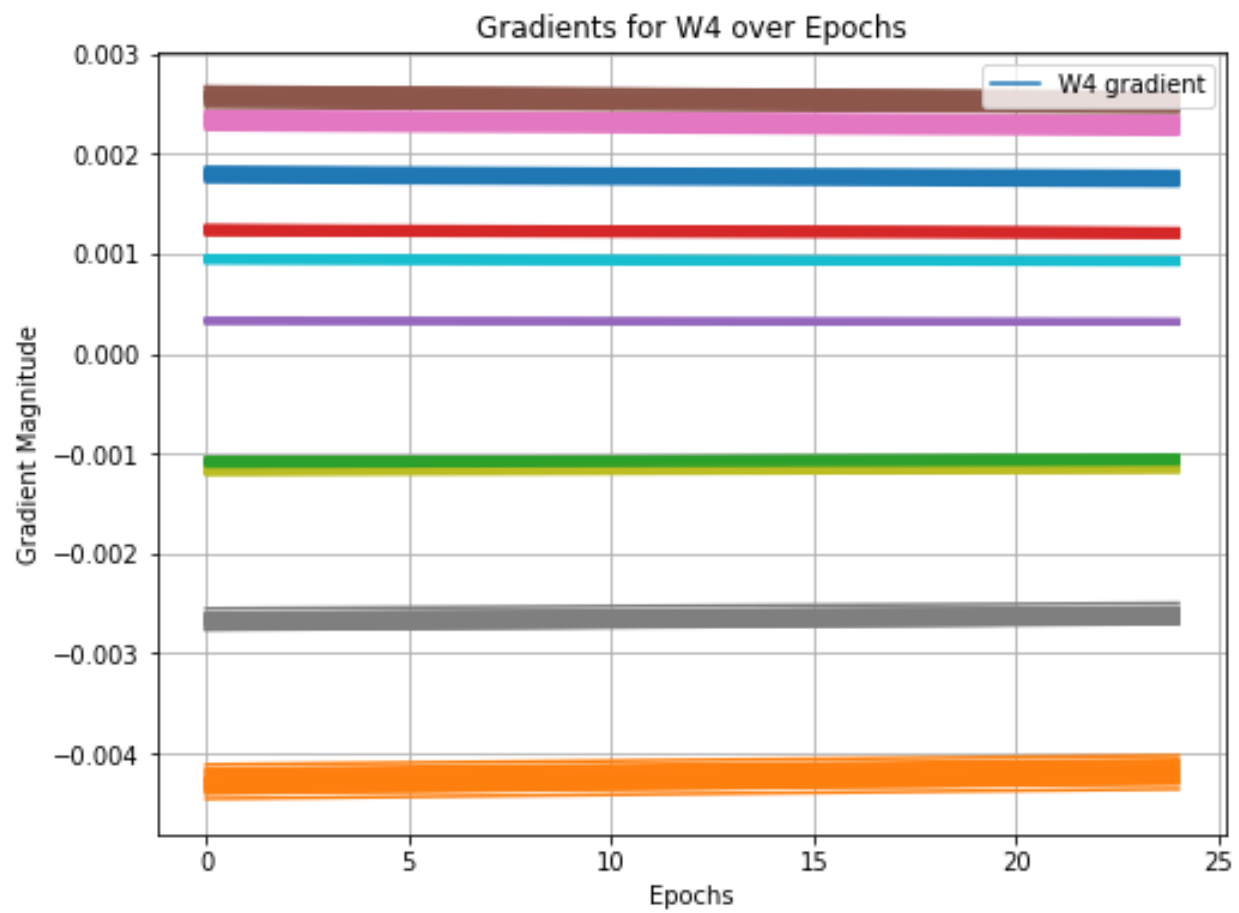
مشکل کاهش گرادیان (Vanishing Gradient Problem) هنگام آموزش شبکه‌های عصبی عمیق اتفاق می‌افتد، زمانی که گرادیان‌های تابع خطا نسبت به پارامترها (وزن‌ها و بایاس‌ها) به‌طور بسیار کوچکی در طول عملیات بازگشتی (بک‌پروپاگیشن) از طریق شبکه منتشر می‌شوند. این ممکن است باعث شود به‌روزرسانی پارامترها تقریباً ناچیز شود و به‌طور موثر فرآیند یادگیری به‌سرعت متوقف شود.

این مشکل به‌طور خاص در شبکه‌های عمیق با تعداد لایه‌های زیاد، به ویژه زمانی که از برخی توابع فعال‌سازی مانند سیگموئید یا تانژانت هایپربولیک استفاده می‌شود، مشکل است. این توابع فعال‌سازی دارای مناطق اشباع هستند که گرادیان به صفر نزدیک می‌شود، که باعث کاهش گرادیان هنگامی که فرایند بک‌پروپاگیشن به صورت معکوس از طریق شبکه حرکت می‌کند، می‌شود.

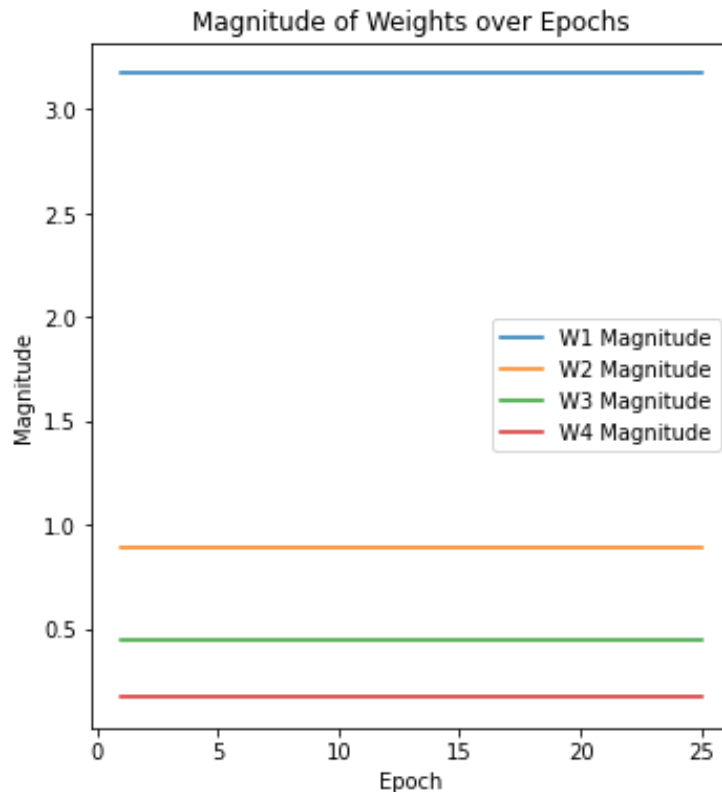
برای بررسی از یک شبکه با سه لایه مخفی با تابع فعالیت سیگموئید استفاده شد.



همان طور که مشاهده می‌شود بدلیل ناچیز بودن مقادیر گرادیان‌ها، تغییرات وزن‌ها بسیار ناچیز است پس یادگیری انجام نشده.



میزان گرادیان روی w_4 در طول اپوک‌ها.



میزان بردارهای وزن در طول اپوک‌ها.

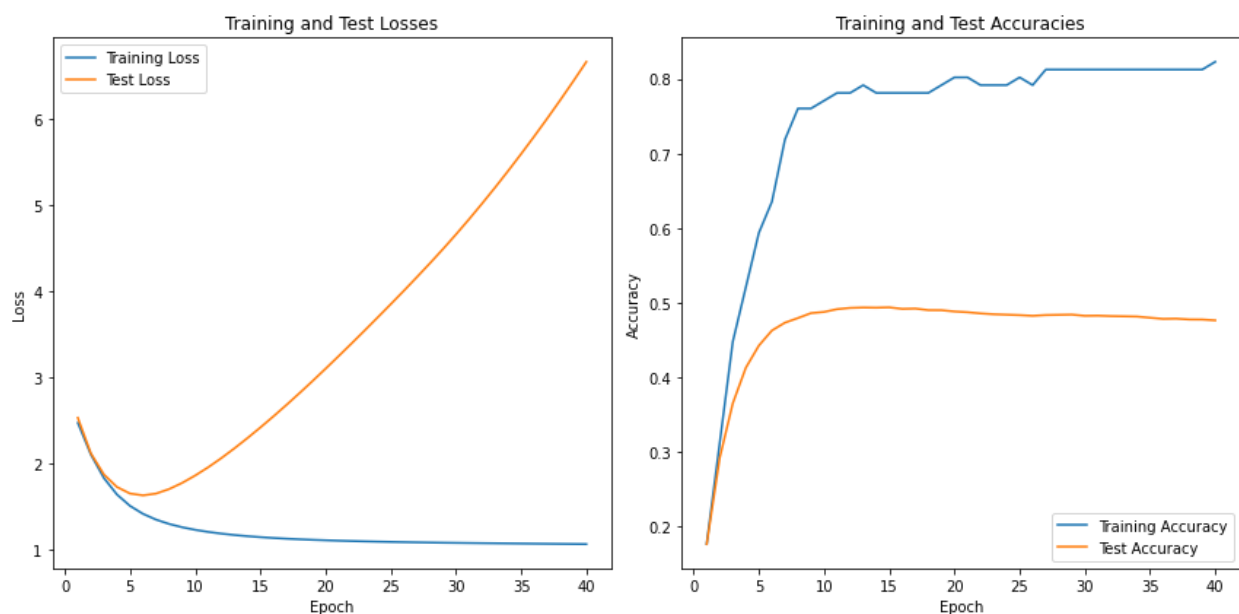
۴. بررسی تاثیر Regularization

• Batch normalization

تکنیک نرمال‌سازی دسته‌ای (Batch Normalization) یا (BN) از روش‌های استفاده شده برای بهبود سرعت آموزش، پایداری و عملکرد شبکه‌های عصبی مصنوعی است، به ویژه شبکه‌های عصبی عمیق. این تکنیک به مسئله شیف‌کواریت درونی پاسخ می‌دهد که به تغییر در توزیع فعال‌سازی‌های شبکه ناشی از به‌روزرسانی پارامترها در زمان آموزش اشاره دارد. BN ورودی هر لایه را به منظور داشتن میانگین صفر و انحراف معیار یک نرمال می‌کند که به پایداری فرآیند یادگیری کمک می‌کند.

ایده اصلی پشت BN این است که ورودی‌های هر لایه را به گونه‌ای نرمال‌سازی کند که توزیع مشابهی در اعضای دسته‌های کوچک داشته باشد. این کار با محاسبه میانگین و واریانس فعال‌سازی‌ها در هر دسته کوچک آموزش و سپس نرمال‌سازی فعال‌سازی‌ها با استفاده از این آمارها انجام می‌شود. نرمال‌سازی دسته‌ای دو پارامتر اضافی برای هر لایه معرفی می‌کند: مقیاس (گاما) و شیف (بتا)، که اجازه می‌دهد مدل مقیاس و شیف بهینه برای فعال‌سازی‌های نرمال‌سازی شده را یاد بگیرد.

حین آموزش، نرمال‌سازی دسته‌ای قبل از اعمال تابع فعال‌سازی بر روی فعال‌سازی‌های هر لایه اعمال می‌شود. این مرحله نرمال‌سازی باعث کاهش مشکلات مربوط به محوشدن یا انفجار گرادیان می‌شود به دلیل اطمینان از این که ورودی‌های لایه‌های بعدی در محدوده مشابهی قرار دارند که منجر به گرادیان‌های پایدارتر و همچنین همگرایی سریع‌تر می‌شود.



که البته طبق آزمایش، تاثیر مثبتی نداشت. و بنظر می‌رسد که بیشتر باعث overfitting شده.

• Dropout

دراپ‌آوت (Dropout) یک تکنیک رگولاریزاسیون است که در شبکه‌های عصبی برای جلوگیری از بیش‌برازش و بهبود توانایی تعمیم‌پذیری مدل استفاده می‌شود. این تکنیک با انتخاب تصادفی و به‌طور تصادفی یک بخش از نورون‌های شبکه را در زمان آموزش "حذف" می‌کند (به این معنی که خروجی هر نورون با احتمال مشخصی صفر می‌شود)، که به طور موثر ظرفیت شبکه را کاهش می‌دهد و جلوی اعتماد زیاد به یک مجموعه خاص از نورون‌ها را می‌گیرد.

در هر تکرار آموزش، Dropout به طور تصادفی یک زیرمجموعه از نورون‌ها را با احتمال مشخصی (معمولاً بین ۰.۲ تا ۰.۵) حذف می‌کند. این بدان معنی است که خروجی هر نورون در هر تکرار، با احتمال مشخصی به صفر تنظیم می‌شود و به طور موثر از شبکه حذف می‌شود. نورون‌های حذف شده به جلوی گذر قدمت یا برگرداندن گرادیان‌ها در زمان آموزش کمکی نمی‌کنند.

با حذف تصادفی نورون‌ها، Dropout به شبکه کمک می‌کند ویژگی‌های قوی‌تر و مقاوم‌تری را یاد بگیرد و از حفظ بیش‌ازحد داده‌های آموزش به زیاده روی پرهیز کند. این کمک می‌کند که بیش‌برازش کاهش یابد و توانایی مدل در تعمیم به داده‌های دیده نشده بهبود یابد. در زمان آزمون، معمولاً Dropout غیرفعال شده و شبکه کامل برای پیش‌بینی استفاده می‌شود.



که البته این مورد هم تاثیر مثبتی نداشت.

• Early stopping

ایست زودهنگام یک تکنیک برای جلوگیری از بیش‌برازش و بهبود عملکرد تعمیم استفاده می‌شود. ایده اصلی ایست زودهنگام این است که در طول فرآیند آموزش، عملکرد مدل روی یک مجموعه داده اعتبارسنجی را نظارت کند. زمانی که عملکرد مدل روی مجموعه اعتبارسنجی شروع به کاهش می‌کند و نشان دهنده آن است که مدل دارد بیش‌برازش به داده‌های آموزش می‌شود، فرآیند آموزش زودهنگام متوقف می‌شود تا از ادامه‌ی بیش‌برازش جلوگیری شود.

و اما بعد از پیاده سازی این روش، از آنجایی که تنظیمات اولیه خوب بود و نتیجه خوبی می‌داد، به ایست نرسید و در ۶۰۰ اپوک به دقت ۹۴ درصد رسید.

