

شبکه‌های عصبی و یادگیری عمیق

دکتر صفا بخش



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

رضا آدینه پور ۴۰۲۱۳۱۰۵۵

تمرین دوم
شبکه چندلایه پرسپترون

۲۰ فروردین ۱۴۰۳

سوال اول - عملی

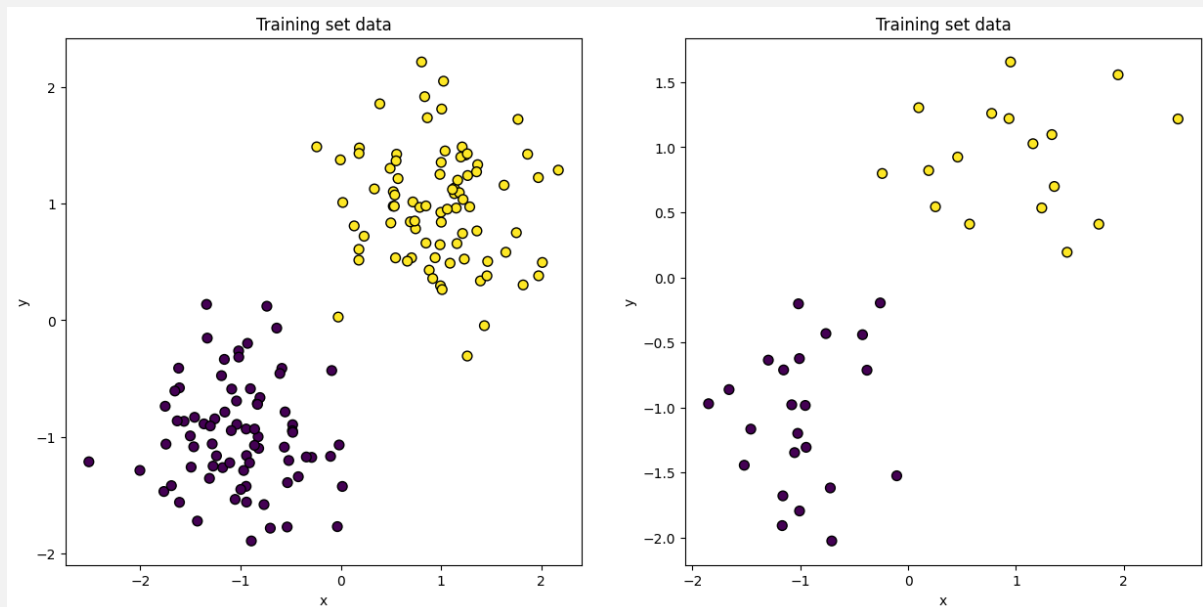
فرض کنید یک مجموعه داده دو کلاسه در اختیار دارید که کاملاً به صورت خطی کلاس‌ها از هم جداپذیر هستند. یک شبکه چند لایه پرسپترون (با طراحی دلخواه) طراحی نموده‌اید که لایه خروجی آن شامل دو نرون می‌باشد که تابع فعال‌ساز softmax بر آن اعمال می‌شود. در زمان آموزش، از تابع خطای binary cross entropy برای محاسبه خطا و بهینه‌سازی وزن‌ها استفاده می‌شود. آیا این امکان وجود دارد که خطای حاصل صفر شود؟ اگر امکان ندارد، با استدلال و اثبات ریاضی نشان دهید و اگر امکان دارد، با معرفی چهار داده (دو داده به ازای هر کلاس) و پرسپترون مد نظر نشان دهید که خطا می‌تواند دقیقاً صفر شود.

پاسخ

در این سوال برای تولید دیتا از تابع `make_blobs` از کتابخانه `sklearn` به صورت زیر استفاده شده است:

```
x, y = datasets.make_blobs(n_samples=200, centers=[(-1, -1), (1, 1)], cluster_std=0.5)
```

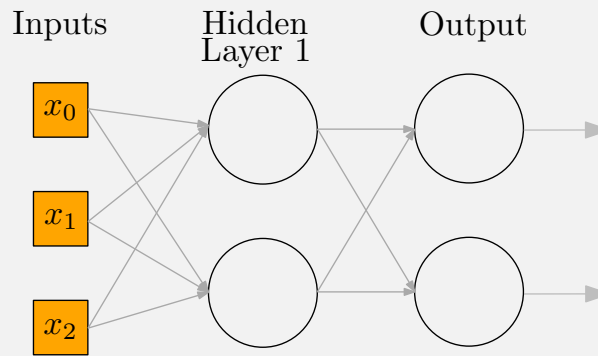
که داده‌هایی با تعداد ۲۰۰ نمونه و انحراف معیار ۰/۵ حول نقطه‌های $(-1, -1)$ و $(1, 1)$ می‌کند که کاملاً جداپذیر خطی است (شکل ۱)



شکل ۱: داده‌های تولید شده برای مسئله

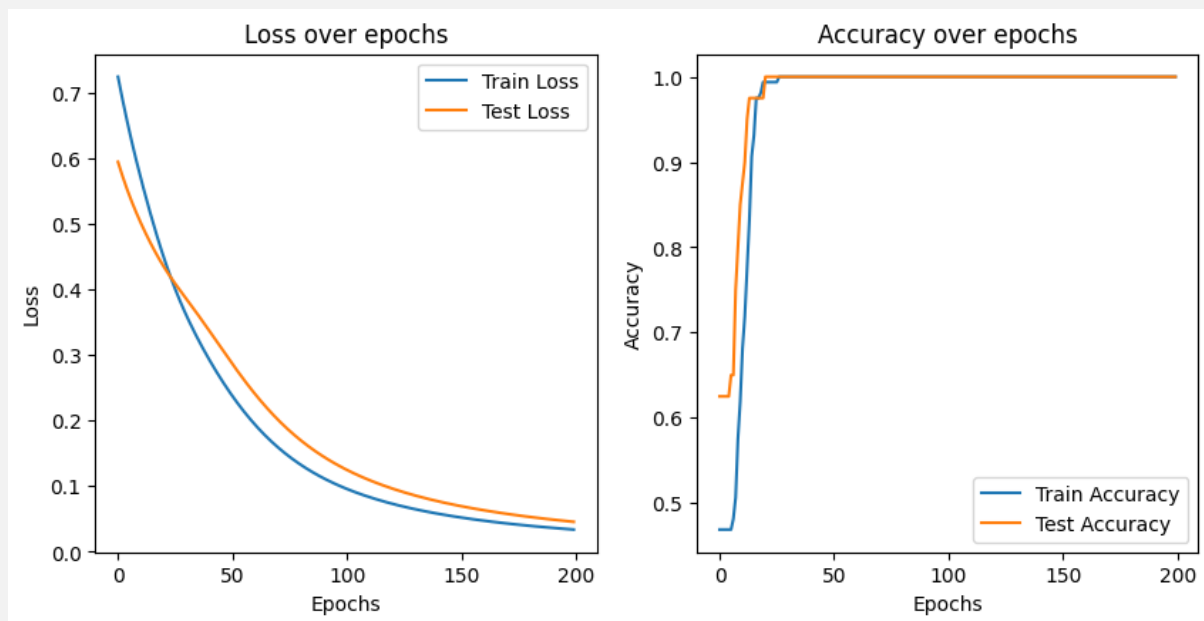
برای حل این مسئله از شبکه‌ی پرسپترون چند لایه‌ای با ساختار «شکل ۲» استفاده شده است. با توجه به اینکه داده‌ها جداپذیر خطی هستند، می‌توان این مسئله را با یک نرون پرسپترون نیز حل نمود اما با توجه به اینکه در صورت سوال گفته شده است شبکه‌ای چند لایه طراحی کنید، از شبکه چند لایه پرسپترون استفاده کردیم.

پاسخ



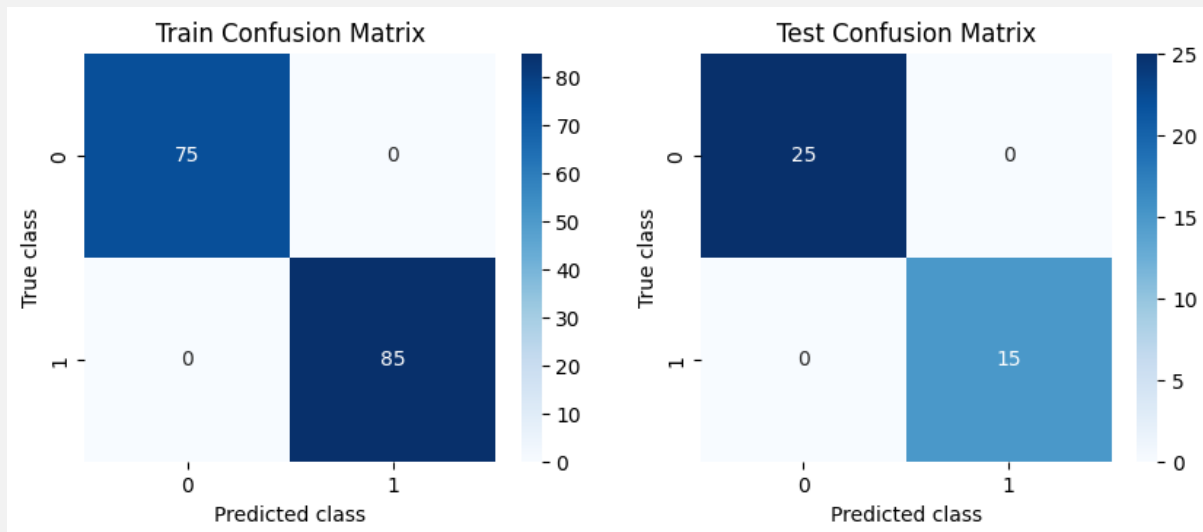
شکل ۲: ساختار شبکه طراحی شده

تابع فعال‌ساز لایه مخفی، ReLU در نظر گرفته شده است و در لایه خروجی نیز از softmax استفاده شده است. طبق خواسته مسئله، از **binary cross entropy** به عنوان تابع خطا استفاده شده است. همچنین از تابع بهینه‌ساز ADAM در این مسئله استفاده شده است. نتایج آموزش در ۲۰۰ دوره آموزشی به صورت زیر گزارش می‌شود:

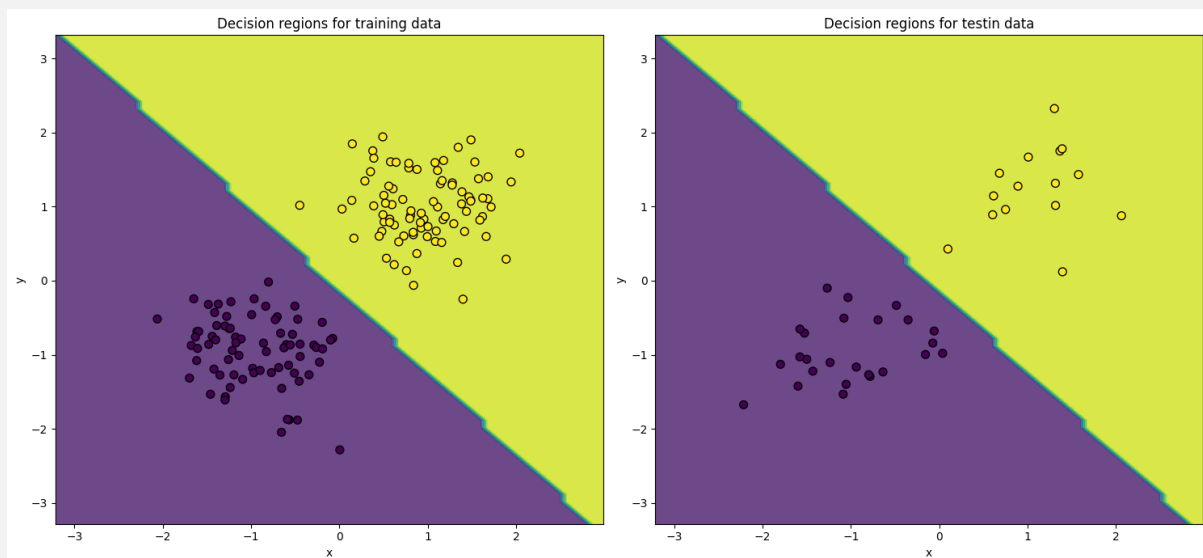


شکل ۳: منحنی خطا و دقت برای داده‌های آموزش و تست

پاسخ



شکل ۴: ماتریس پراکندگی



شکل ۵: ناحیه تصمیم

سوال دوم - نظری

تعیین اندازه دسته (Batch) به چه عواملی بستگی دارد و تاثیر آن در روند آموزش شبکه چیست؟ فرض کنید که اندازه و سائز هر نمونه از مجموعه داده بگونه ای بزرگ و حجیم است که وقتی اندازه دسته بیش از ۲ باشد، خطای حافظه دریافت می‌شود. (Out of memory) چگونه می‌تواند این مشکل و چالش را بدون ارتقای سخت‌افزار حل نمود؟ راهکار مدنظر را معرفی و با جزئیات کامل پیاده سازی کنید.

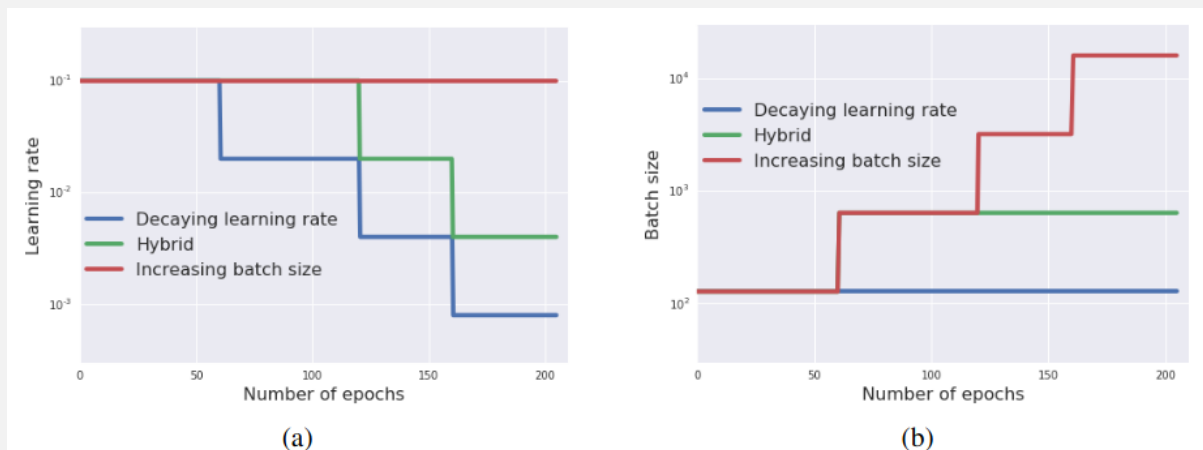
پاسخ

در سیستم‌های یادگیری عمیق، هایپرپارامترهای زیادی دخیل هستند، یکی از مهمترین آنها اندازه دسته یا batch size است. [۱] اندازه دسته از این جهت مهم است که زیرا با افزایش اندازه دسته می‌توان سرعت محاسباتی را برای آموزش مدل افزایش داد. اما مشکلی که اندازه دسته بزرگ می‌تواند ایجاد کند عدم آموزش مناسب شبکه است. زیرا اگر اندازه دسته خیلی بزرگ باشد، در فرایند آموزش، داده‌هایی که در ابتدای دسته وجود داشتند، هنگامی که فرایند آموزش به انتهای دسته می‌رسد موجب این می‌شود که داده‌های ابتدایی فراموش شود و شبکه در memorisation دچار مشکل می‌شود و شبکه به درستی آموزش نمی‌بیند. [۲]

همواره مصالحه ای بین اندازه دسته و سرعت همگرایی وجود دارد. اندازه دسته ای برابر با اندازه کل مجموعه داده ها، همگرایی را تضمین می‌کند اما به قیمت کاهش سرعت فرایند آموزش از طرفی کوچک گرفتن اندازه دسته سرعت همگرایی را بالا می‌برد اما تضمینی برای همگرایی مدل وجود ندارد. [۲]

بنابر این پیشنهاد می‌شود که از اندازه دسته کوچک شروع به آموزش دادن شود و در حین آموزش کم کم اندازه دسته را افزایش داد. [۲]

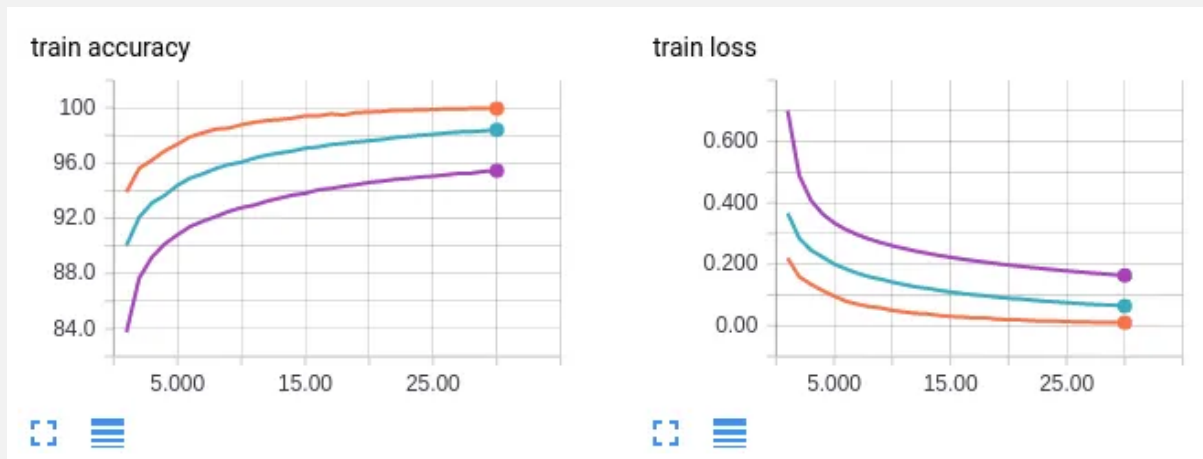
در [۲] روند تغییر هایپرپارامترهای شبکه مثل نرخ یادگیری و اندازه دسته به صورت زیر پیشنهاد شده است:



شکل ۶: روند تغییر پویا هایپرپارامترهای مهم شبکه

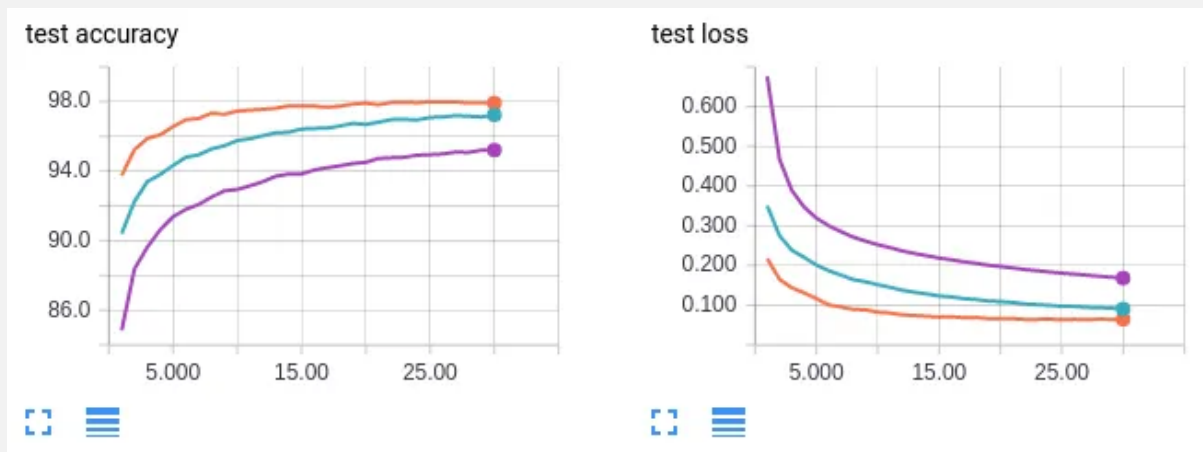
در ادامه مثالی از سرعت همگرا شدن یک شبکه MLP که با اندازه دسته های مختلف بر روی دیتاست MNIST آموزش داده شده است آورده شده است. در شکل های «۷ و ۸» نمودار نازجی رنگ مربوط به اندازه دسته ۶۴، نمودار آبی مربوط به اندازه دسته ۲۵۶ و نمودار بنفش مربوط به اندازه دسته ۱۰۲۴ است. همانطور که توضیح داده شد، سرعت همگرایی با اندازه دسته کوچکتر، بیشتر است.

پاسخ



شکل ۷: تاثیر اندازه دسته در سرعت همگرایی آموزش

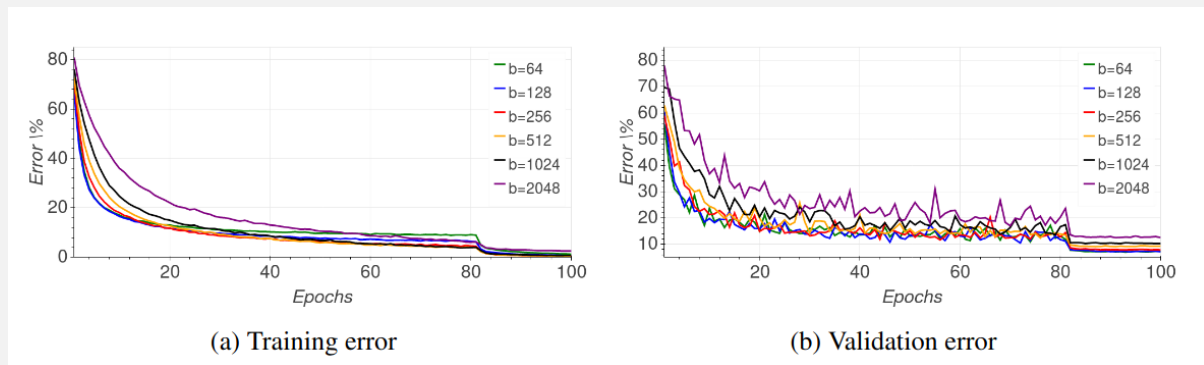
این قضیه در مورد فاز تست شبکه پس از آموزش نیز صادق است:



شکل ۸: تاثیر اندازه دسته در سرعت همگرایی تست شبکه

یکی دیگر از عواملی که در تعیین اندازه دسته مهم است، حافظه سیستمی است که از آن استفاده می‌کنیم. اگر حافظه سیستم‌مان محدود باشد نمی‌توان اندازه دسته را بزرگ گرفت چرا که به هنگام آموزش حافظه سیستم پر خواهد شد و با خطای out of memory مواجه خواهیم شد. بنابر این در پاسخ به قسمت دوم سوال می‌توان گفت در صورت مواجه شدن با این خطا می‌بایست اندازه دسته را کوچک تر کنیم. با تکرار و سعی و خطا می‌توان اندازه دسته مناسب برای آنکه حافظه پر نشود را پیدا کرد.

در شکل «۹» می‌توان مثال دیگری از این قضیه را مشاهده کرد.



شکل ۹: تاثیر اندازه دسته در سرعت همگرایی تست شبکه [۳]

*

References

- [1] Bengio Y. Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade: Second edition* 2012 May (pp. 437-478). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [2] Smith SL, Kindermans PJ, Ying C, Le QV. Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*. 2017 Nov 1.
- [3] Hoffer E, Hubara I, Soudry D. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. *Advances in neural information processing systems*. 2017;30.

سوال سوم - نظری

نشان دهید شبکه چند لایه پرسپترونی که فقط از تابع فعال‌سازی ReLU و یا pReLU استفاده می‌کند، تابع پیوسته تکه‌ای خطی می‌سازد.

سوال چهارم - نظری

تقارن در شبکه عصبی به چه معناست؟ آیا نیاز داریم این تقارن را بشکنیم؟ در صورتی که جواب شما مثبت است، کیس مورد نظر را طراحی کرده و توضیح دهید. بررسی کنید چه روش‌هایی برای شکستن تقارن وجود دارد. چند مورد نام ببرید.

سوال پنجم - عملی

ابروصوح یک کاربرد در بینایی کامپیوتر می‌باشد که در آن هدف ارتقای وضوح تصاویر می‌باشد. این امر می‌تواند در مقاصد مختلف نظیر تصویر برداری پزشکی، بهبود تصاویر نظارتی-امنیتی، بازسازی تصاویر قدیمی و ... به‌کار گرفته شود. در این سوال هدف طراحی و پیاده سازی یک شبکه عصبی چندلایه برای هدف فوق می‌باشد.

۱. ۱۰ تصویر دلخواه از اینترنت که حاوی گستره رنگی مختلفی می‌باشد را به‌عنوان مجموعه داده انتخاب کنید و آن را نمایش دهید. حال وضوح هر یک از تصاویر را نصف کنید. اکنون به ازای هر یک از پیکسل‌ها در عکس اصلی، متناظر آن و هشت همسایگی مجاور آن در عکس با وضوح پایین تر را بیابید و مجموعه داده موردنظر را بدست آورید. ابعاد ورودی برابر با ۲۷ ویژگی (پیکسل متناظر و هشت همسایگی آن به ازای سه کانال رنگی در وضوح پایین) خواهد بود و خروجی (لیبل) نیز شامل سه مقدار (مقدار سه کانال RGB در تصویر اصلی) خواهد بود. این روند را برای تمامی پیکسل‌های ۱۰ تصویر انجام دهید تا برای هر تصویر i یک مجموعه داده به صورت $W_i * H_i, 27, 3$ پدید آید. دو تصویر را برای آزمون، و یک تصویر را برای اعتبار سنجی و هفت تصویر باقی مانده را برای آموزش استفاده کنید. می‌توانید پیکسل‌های حاصل از تصاویر مختلف در گروه آموزش را باهم ترکیب کرده و درهم (shuffle) سازید که ابعاد آن مجموعه داده به صورت $\sum_{i=1}^7 W_i * H_i, 27, 3$ درآید.

۲. یک شبکه چند لایه پرسپترون طراحی و آموزش دهید که بتواند به ازای ۲۷ ویژگی ورودی در وضوح پایین، مقدار پیکسل رنگی در وضوح بالا را محاسبه کند. معماری خود را ترسیم نموده و آموزش شبکه را توضیح دهید. از چه تابع خطایی برای آموزش استفاده کرده‌اید؟ موارد ذکر شده در ابتدای پروژه را برای این سوال به‌صورت کامل گزارش دهید و نتایج را تحلیل کنید.

۳. مقدار تابع خطا را به ازای مجموعه داده آزمون محاسبه و گزارش نمایید و در گام نهایی تصاویر با وضوح بالا را تولید نموده و آن را با تصاویر اصلی مقایسه کنید. علاوه بر مقایسه‌ی بصری، معیارهای کمی SSIM و PSNR را مطالعه کرده و بر اساس آن عملکرد شبکه خود را ارزیابی کنید.

۴. اگر از شبکه‌ی قسمت ۲، دوبار متوالی استفاده شود، می‌تواند وضوح تصویر را چهار برابر کند. به نظر شما این رویکرد می‌تواند مفید واقع شود یا یک شبکه‌ی ای که به صورت مستقیم چهار برابر وضوح را افزایش می‌دهد؟ با انجام آزمایش و گزارش کمی نتیجه مورد نظر را نشان دهید.