



## یادگیری عمیق

نیم‌سال اول ۱۴۰۰-۱۳۹۹  
مدرس: عمادالدین فاطمی‌زاده

تمرین سری سوم

یادگیری عمیق

زمان تحویل: ۲۲ آذر

به نکات زیر توجه کنید:

فایل گزارش را به همراه تمامی کدها در یک فایل فشرده و با عنوان HW3\_STD# در بارگذاری نمایید.  
برای هر یک از سوالات عملی، پوشه ای مجزا در نظر بگیرید و کدها را درون آن قرار دهید و از شماره سوال برای نام پوشه استفاده کنید، برای مثال نام Q5\_a را برای پوشه مربوط به قسمت اول سوال عملی ۵ قرار دهید.  
نمره کل تمرین ۱۸۰ است. ۵۰ نمره از کل تمرین امتیازی می باشد.

## مسئله ۱. Network Design (۵+۲۰ نمره)

الف) یک شبکه عصبی کانولوشنی دارای ۴ لایه کانولوشن  $3 \times 3$  با اندازه گام ۱ بدون لایه Pooling فرض کنید. هریک از نورون‌های لایه آخر این شبکه چه تعداد پیکسل از تصویر ورودی را پوشش می‌دهند؟ اجتماع آن‌ها چه تعداد پیکسل را پوشش می‌دهد؟

ب) لایه‌ی Pooling چگونه و با اضافه کردن چه خواصی به شبکه‌های کانولوشنی قدرت و تعمیم‌پذیری آن را افزایش می‌دهد.

ج) فرض کنید در قسمتی از شبکه سائز تصویر ورودی  $35 \times 35$  و تعداد کانال‌های آن برابر ۱۶ است، سائز فیلتر مورد استفاده در لایه کانولوشن  $3 \times 3$  با طول گام ۱ و تعداد فیلترها ۳۲ است. ابعاد خروجی و همچنین تعداد پارامترهای این لایه از شبکه را مشخص کنید. (پدینگ را Same در نظر بگیرید.)

د) در قسمت قبل فرض کنید قصد داریم با یک لایه تمام متصل از ورودی با ابعاد مشابه به تصویر خروجی با همان ابعاد برسیم، تعداد پارامترها در این حالت چقدر خواهد بود؟ از مقایسه با قسمت قبل چه نتیجه‌ای می‌گیرید؟

ه) شبکه زیر با پدینگ Same در نظر بگیرید:

INPUT  $[3 \times 227 \times 227]$   
Conv1  $(5 \times 5 \times 128)$   
Conv2  $(5 \times 5 \times 64)$   
Max-Pooling  $(5 \times 5, \text{stride } 2 \times 2)$   
FC<sub>512</sub>  
FC<sub>100</sub>

سائز خروجی در هریک از لایه‌های کانولوشن و Pooling و تعداد پارامترها برای تمام لایه‌ها مشخص کنید. سپس مشکلات این شبکه را شرح دهید و تغییرات لازم را برای بهبود آن انجام دهید و سپس شبکه خود را از لحاظ تعداد پارامتر با شبکه ارائه شده مقایسه کنید.

ز) ساختار Separable Convolution را توضیح دهید و تعداد پارامترها و همچنین تعداد عملیات لازم برای محاسبه خروجی آن را با کانولوشن عادی مقایسه کنید.

- ح) ساختار Capsule network را توضیح دهید و مزایا و معایب آنرا در مقایسه با شبکه کانولوشنی بیان کنید.
- ط) (+۳) ساختار Densely Connected Convolutional Networks را توضیح دهید و مزایا و معایب آنرا در مقایسه با شبکه کانولوشنی بیان کنید و همچنین پارامترهای آن از جمله growth rate را توضیح دهید.
- ی) (+۲) تابع تلف Triplet Loss را توضیح دهید و کاربرد آنرا بیان کنید.

## مسئله ۲. Back-propagation in CNN (۱۵+۵ نمره)

در حالت کلی کانولوشن ورودی  $X$  با سایز  $(N, C, i_h, i_w)$  که در آن  $N$  تعداد نمونه در ورودی  $W$  با سایز  $(F, C, f_h, f_w)$  را می‌دهد و برای یک نمونه  $b$  و فیلتر  $f$  به صورت زیر محاسبه می‌شود:

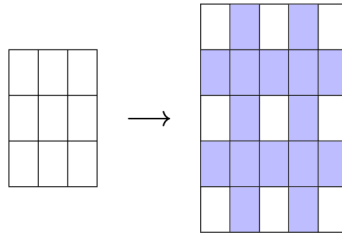
$$O(b, f, i, j) = \sum_{r=0}^{C-1} \sum_{k=0}^{f_h-1} \sum_{l=0}^{f_w-1} W(f, r, k, l) X(b, r, s * i + k, s * j + l)$$

که در آن

$$O_w = \frac{i_w - f_w}{s} + 1$$

$$O_h = \frac{i_h - f_h}{s} + 1$$

dilated convolution نوعی از کانولوشن است که در آن اندازه‌ی موثر فیلتر به وسیله پارامتر  $d$  تعیین می‌شود و در این حالت مقدار  $d-1$  صفر بین مقادیر فیلتر در هر بعد اضافه می‌شود. برای مثال فیلتر با سایز  $3 \times 3$  و فاکتور ۲ سایز موثر آن برابر است با  $5 \times 5$  و به صورت زیر نمایش داده می‌شود:



الف) مقدار موثر اندازه فیلتر را در یک dilated convolution با فاکتور  $d$  و در حالتی که فیلتر با سایز  $k \times k$  است بدست آورید.

ب) نشان دهید گرادیان پس انتشار تابع هزینه  $L$  نسبت به فیلتر  $W$  از رابطه زیر بدست می‌آید که در آن مشتق تنها نسبت به یک متغیر اسکالر  $W(f', c', k', l')$  محاسبه شده است.

$$\frac{\partial L}{\partial W(f', c', k', l')} = \sum_{b=0}^{N-1} \left( \sum_{i=0}^{O_w-1} \right) \sum_{j=0}^{O_h-1} \frac{\partial L}{\partial O(b, f', i, j)} X(b, c', s * i + k', s * j + l')$$

ج) (+۲) با استفاده از رابطه بالا برای حالتی که تنها یک نمونه داریم نشان دهید گرادیان نسبت به یک فیلتر دوبعدی

$W(f', c', :, :)$  برابر با کانولوشن dilated با فاکتور  $s$  گرادیان تابع هزینه نسبت به خروجی در ورودی  $X(b, c', :, :)$  است.

د) (+۳) ثابت کنید گرادیان تابع اسکالر  $L$  نسبت به ورودی  $X$  به صورت زیر است:

$$\frac{\partial L}{\partial X(b', c', k', l')} = \sum_{f=0}^{F-1} \left( \sum_{i=0}^{O_h-1} \sum_{j=0}^{O_w-1} \frac{\partial L}{\partial O(b', f, i, j)} W(f, c', k' - s_h * i, l' - s_w * j) \right)$$

## مسئله ۳. Exponential Decay (۵ نمره)

یکی از روش‌های پرکاربرد برنامه‌ریزی نرخ یادگیری (learning rate scheduling) در مسائل یادگیری عمیق، استفاده از تابع کاهش نمایی (Exponential Decay) است.

الف) شیوه کار این تابع را توضیح دهید.

ب) این تابع چه مزیتی برای یادگیری شبکه عمیق دارد؟

پ) در مورد دو متغیر اصلی این تابع یعنی Decay rate و Decay steps توضیح دهید.

#### مسئله ۴. CNN (۲۰+۴۰ نمره)

در این بخش می‌خواهیم یک شبکه کانولوشن را بر روی دیتاست CIFAR10 آموزش دهیم. ابتدا از فایل CNN numpy دفترچه CNN را که پیاده سازی شبکه کانولوشنی با numpy است را کامل کنید. سپس دفترچه CNN on CIFAR برای این منظور کامل کنید. دیتاست CIFAR را از اینجا دریافت کنید.

#### مسئله ۵. CNN (۲۰+ نمره)

در این بخش می‌خواهیم یک شبکه کانولوشن را بر روی دیتاست SVHN که دیتاستی کاربردی تر از MNIST است، آموزش دهیم. در این بخش دفترچه‌ای در اختیار شما قرار میگیرد و باید بخش‌هایی از آن را تکمیل کنید که این بخش با PyTorch انجام شود. برای دریافت نمره این بخش باید بر اساس دفترچه‌ای که در اختیار شما قرار میگیرد دقت بالای ۹۳ درصد کسب کنید (فقط قسمت TODO را کامل کنید). و برای رسیدن به این هدف بهتر است از Batch Normalization و Dropout استفاده کنید.

#### مسئله ۶. Autoencoder (۲۰ نمره)

در این بخش می‌خواهیم با خودکدگذار و کاربردهای آن آشنا شویم.  
الف) یک خودکدگذار تنها با شبکه تمام‌متصل بر روی دیتاست MNIST آموزش دهید و خروجی و ورودی را رسم کنید. توجه: بعد لایه نهان و تعداد لایه‌های این شبکه را دلخواه در نظر بگیرید.

ب) حالا یک خودکدگذار با استفاده از معماری کانولوشن بر روی همین دیتاست آموزش دهید و خروجی و ورودی را رسم کنید و همچنین میزان تلف را با شبکه الف مقایسه کنید.

ج) حالا می‌خواهیم یک denoising autoencoder آموزش دهیم به این صورت که ضریبی از نویز گوسی به داده‌ها اضافه کرده و خودکدگذار به منظور رفع نویز عمل کند. میزان تلف و همچنین خروجی و ورودی را برای مقادیر مختلف این ضریب نویز رسم کنید (حداقل ۴ مقدار مختلف).

#### مسئله ۷. Sign Language Classifier CNN (۳۰ نمره)

در این بخش تلاش میکنیم تا شبکه‌ای برای آموزش داده‌های Sign Language طراحی کنیم و آن را به طور مناسب آموزش داده و نتایج را مشاهده کنیم. لازم به ذکر است که میتوانید از تمامی framework های موجود برای پیاده‌سازی شبکه استفاده کنید.

الف) دیتاست را از این آدرس دانلود کنید.

ب) این دیتاست شامل ۲۷۴۵۵ تصویر برای آموزش و ۷۱۷۲ تصویر برای تست میباشد. سائز تمامی تصاویر ۲۸ در ۲۸ بوده و حروف انگلیسی را به جز z و z به نمایش میگذارند. دو فایل با پسوند csv که مربوط به تصاویر آموزش و تست میباشد را خوانده و ابعاد را نمایش دهید.

پ) تعداد تصاویر هر حرف را نمایش دهید. آیا تصاویر به طور یکنواخت بین دسته‌ها تقسیم شده‌اند؟

ت) تصاویر و لیبل آنها را به numpy array تبدیل کرده و به صورت دلخواه ۵ تصویر آموزش و ۵ تصویر تست را همراه با ذکر لیبل آنها نمایش دهید.

ث) به دلخواه ۲۰ تا ۳۰ درصد از داده‌های آموزش را به عنوان داده validation در نظر بگیرید. سپس داده‌ها را به ۲۵۵ تقسیم کنید. بدین ترتیب تمامی داده‌ها مقداری از ۰ تا ۱ میگیرند و نرمالایز میشوند.

ج) لیبل داده‌های آموزش را به صورت one-hot کد کنید.

چ) شبکه دلخواه خود را پیاده کنید. شبکه پیشنهادی:

```
[۲۸ × ۲۸ × ۱] INPUT
Conv۲D (۱۶ × ۳ × ۳) ReLU
Max-Pooling (۲ × ۲)
Conv۲D (۳۲ × ۳ × ۳) ReLU
Max-Pooling (۲ × ۲)
Conv۲D (۶۴ × ۳ × ۳) ReLU
Max-Pooling (۲ × ۲)
Flatten
FC, ۱۲۸ ReLU
Dropout, rate = ۰.۲
FC, ۲۴ Softmax
```

ح) از تابع تلف categorical crossentropy و بهینه‌ساز Adam استفاده کنید. ابعاد داده‌های ورودی را برای ورود به شبکه اصلاح کنید. مقادیر نرخ یادگیری، تعداد epochها و batch size را تنظیم کنید.

خ) پس از آموزش شبکه، نمودارهای loss function و accuracy را برای هر دو دسته داده‌ی آموزش و validation رسم کنید.

د) دقت را بر روی داده‌های آموزش، validation و تست نمایش دهید و confusion matrix را برای هر سه دسته داده رسم کنید. دقت داشته باشید که accuracy قابل قبول بر روی داده‌های تست حداقل ۷۰ درصد میباشد. همچنین نباید اختلاف زیادی بین accuracy داده‌های آموزش و validation وجود داشته باشد به نحوی که شبکه overfit شود.

ذ) وزن‌های یادگرفته شده را برای فیلترهای کانولوشنی لایه اول به صورت عکس نمایش دهید و در گزارش بیاورید. آیا الگوی مشخصی در این فیلترها دیده میشود؟

---

موفق باشید.