



یادگیری ژرف

نیم سال اول ۱۴۰۲-۰۳

مدرس: دکتر حمید بیگی

زمان تحویل: ۲۹ آبان

شبکه‌های عصبی پیچشی

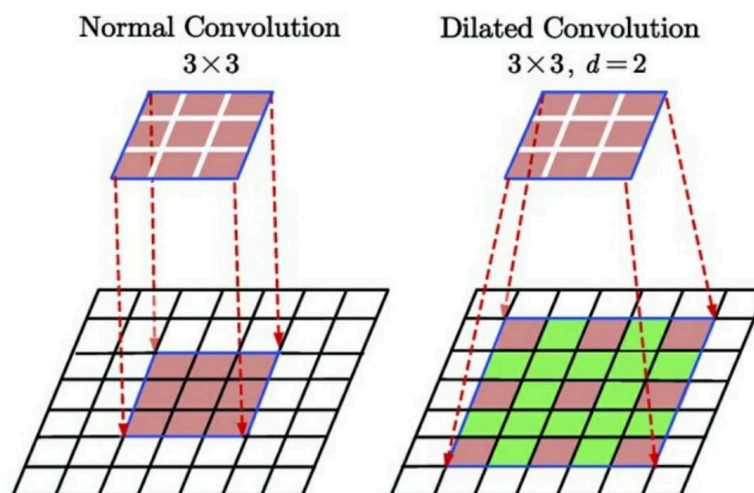
تمرین سری دوم (۱۰۰ نمره)

لطفا نکات زیر را رعایت کنید:

- سوالات خود را از طریق پست مربوط به تمرین در Quera مطرح کنید.
- در هر کدام از سوالات، اگر از منابع خارجی استفاده کرده‌اید باید آن را ذکر کنید. در صورت همفکری با افراد دیگر هم باید نام ایشان را در سوال مورد نظر ذکر نمایید.
- پاسخ ارسالی واضح و خوانا باشد. در غیر این صورت ممکن است منجر به از دست دادن نمره شود.
- پاسخ ارسالی باید توسط خود شما نوشته شده باشد. به اسکرین‌شات از منابع یا پاسخ افراد دیگر نمره‌ای تعلق نمی‌گیرد.
- در صورتی که بخشی از سوال‌ها را جای دیگری آپلود کرده و لینک آن را قرار داده باشید، حتما باید تاریخ آپلود مشخص و قابل اعتنا باشد.
- تمام پاسخ‌های خود را در یک فایل با فرمت `HW#_[SID]_[Fullname].zip` روی کوئرا قرار دهید.

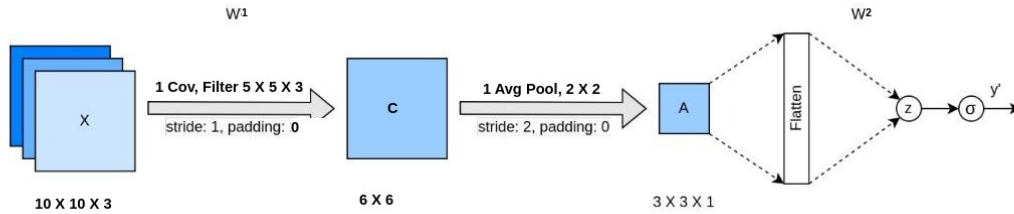
سوال ۱: (نظری) Dilated Convolution (۱۵ نمره)

یکی از فیلترهای استفاده‌شده در شبکه‌های عصبی، فیلترهای dilated می‌باشد. تفاوت این فیلترها با فیلترهای معمولی این است که پیکسل‌ها را با فاصله ورودی میگیرند، نحوه عملکرد یکی از این فیلترها در شکل ۱ دیده می‌شود. حال به سوالات زیر پاسخ دهید.



شکل ۱: نحوه عملکرد فیلتر dilated

- تصویر با ابعاد $i \times i \times 3$ داریم که ابتدا یک لایه‌ی کانولوشن معمولی با فیلتر $d_1 \times k_1 \times k_1 \times 3$ بر روی آن اعمال می‌شود و در لایه‌ی بعد d_2 لایه‌ی معمولی $d_1 \times k_2 \times k_2 \times 1$ بر روی خروجی لایه‌ی قبل اعمال می‌شود و سپس d_3 dilated فیلتر $k_3 \times k_3 \times d_2$ با پارامتر گسترش n بر روی خروجی لایه‌ی دوم اعمال می‌شود، بر حسب پارامترهای داده‌شده در صورت سوال ابعاد خروجی لایه‌ی آخر را حساب نمایید.
- وقتی از کانولوشن dilated استفاده می‌شود، مبحث محدوده‌ی دید پررنگ‌تر مطرح می‌شود، اگر در لایه‌ی اول یک فیلتر k در k بر روی ورودی اعمال شود، که پارامتر گسترش آن d_1 است، و در لایه‌ی دوم یک فیلتر $k \times k$ با پارامتر گسترش d_2 اعمال شود در لایه‌ی سوم یک فیلتر $k \times k$ با پارامتر گسترش d_3 اعمال شود برای عنصر i و j از خروجی محدوده‌ی دید را بر حسب پارامترهای داده‌شده در سوال محاسبه نمایید.



شکل ۲: شبکه CNN

سوال ۲: (نظری) انتشار به عقب (۱۵ نمره)

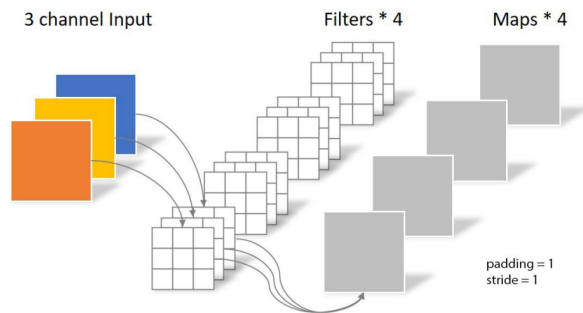
یک شبکه‌ی کانولوشنی در شکل ۲ نشان داده شده است، در این شکل W^1 وزن‌های فیلتر کانولوشن لایه‌ی اول است. $C_{i,j}$ معرفی‌کننده‌ی خروجی سطر i و ستون j از لایه‌ی کانولوشن است و $W^1_{i,j,k}$ وزن لایه‌ی کانولوشن در کانال k ام سطر i ام و ستون j ام در کانولوشن لایه‌ی اول است و W^2 وزن‌های لایه‌ی تمام متصل نهایی را نشان می‌دهد. به سوالات زیر پاسخ دهید:

- با استفاده از قاعده مشتق زنجیره‌ای عبارت $\frac{\partial loss}{\partial C_{i,j}}$ را برحسب $\frac{\partial loss}{\partial y}$ محاسبه کنید.
- حال با استفاده از روابط به دست آمده در قسمت اول و رابطه‌ی کانولوشن، عبارت $\frac{\partial loss}{\partial W^1_{i,j,k}}$ را محاسبه کنید.

سوال ۳: (نظری) کانولوشن عمقی جدایی‌پذیر (۱۰ نمره)

کانولوشن عمقی جدایی‌پذیر^۱ برای کاربردهای پردازش تصویر به کار می‌رود. در این نوع کانولوشن، ابتدا کانولوشن عمقی اجرا می‌شود و سپس کانولوشن نقطه‌ای^۲. فرض کنید ورودی، یک تصویر سه کاناله با رزولوشن 224×224 باشد. اندازه کرنل لایه‌ی کانولوشنی 3×3 و تعداد کانال‌های خروجی، ۴ است. (برای سادگی، از بایاس صرف نظر کنید.)

- تعداد پارامترهای قابل یادگیری در یک لایه کانولوشن عادی (شکل ۳) را به دست آورید.



شکل ۳: لایه کانولوشن عادی

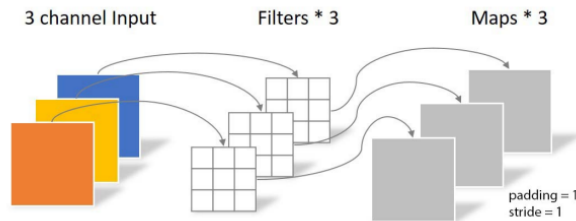
- در یک لایه کانولوشن عمقی جدایی‌پذیر، ابتدا کانولوشن عمقی (شکل ۴) بر روی ورودی اجرا می‌شود که تعداد لایه‌های خروجی برابر با تعداد لایه‌های ورودی است. سپس، بر روی خروجی این لایه، کانولوشن نقطه‌ای (شکل ۵) اجرا می‌شود که همان کانولوشن عادی با اندازه فیلتر ۱ در ۱ است. فرض کنید ۴ کانال خروجی داریم. تعداد پارامترهای قابل یادگیری برای یک لایه کانولوشن عمقی جدایی‌پذیر را محاسبه کنید.

سوال ۴: (نظری) Pooling Layers (۱۰ نمره)

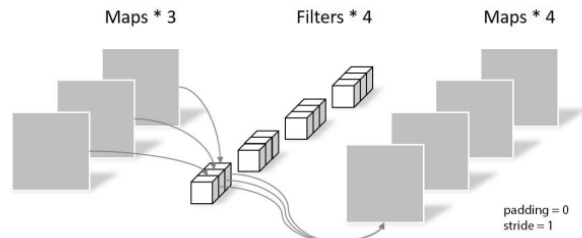
به سوالات زیر پاسخ دهید. (برای پاسخ به سوالات می‌توانید از شبه‌کد استفاده کنید.)

- پیچیدگی محاسباتی عملیات Pooling را محاسبه کنید. فرض کنید اندازه ورودی $c \times h \times w$ ، اندازه‌ی پنجره pooling $p_h \times p_w$ ، (p_h, p_w) padding و (s_h, s_w) stride است.
- دو عملیات Max Pooling و Average Pooling را با یکدیگر مقایسه کنید.

^۱ Depth-wise Separable Convolution
^۲ Point-wise Convolution



شکل ۴: لایه کانولوشن عمقی



شکل ۵: لایه کانولوشن نقطه‌ای

- آیا می‌توانیم Max Pooling و Average Pooling را به کمک کانولوشن پیاده‌سازی کنیم؟
- می‌خواهیم Max Pooling را به کمک عملیات ReLU انجام دهیم. $ReLU(x) = \max(0, x)$
- $\max(a, b)$ را به کمک ReLU پیاده‌سازی کنید.
- به کمک قسمت بالا، سعی کنید Max Pooling را به کمک لایه‌های کانولوشنی و ReLU پیاده‌سازی کنید.
- آیا می‌توانیم از Softmax برای اجرای عملیات Pooling استفاده کنیم؟ بررسی کنید.

سوال ۵: دسته‌بندی تصاویر (عملی) (۲۵ نمره)

در این تمرین می‌خواهیم یک شبکه پیچشی از قبل آموزش دیده (ResNet۵۰) را برای وظیفه دسته‌بندی روی یک دیتاست جدید (۰-Cifar) بکار بگیریم. برای حل این تمرین به نوتبوک مربوطه مراجعه کنید و قسمت های خالی و پاسخ سوال داده شده را تکمیل کنید.

سوال ۶: (عملی) پیاده سازی شبکه عصبی پیچشی (۲۵ نمره)

در این سوال شما یک شبکه عصبی پیچشی را برای وظیفه دسته‌بندی تصاویر با دادگان Fashion-MNIST پیاده‌سازی می‌کنید. نوتبوی که در اختیار شما قرار گرفته شامل بخش‌های:

- بارگذاری دادگان و تقسیم‌بندی آن به سه بخش آموزش، اعتبارسنجی و آزمون
- پیاده‌سازی ساختار شبکه پیچشی با استفاده از PyTorch و همچنین پیاده‌سازی تابع Fit
- انتخاب و تعریف نرخ یادگیری، تابع زیان و بهینه‌ساز مناسب
- آموزش مدل و رسم نمودار زیان و دقت برای دادگان آموزش و اعتبارسنجی
- پیشبینی خروجی مدل روی دادگان آزمون

است. با مراجعه به نوتبوک مورد نظر بخش‌های مشخص شده از کد را تکمیل نمایید.