



یادگیری عمیق

نیم سال دوم ۰۴-۰۳
مدرس: مهدیه سلیمانی

ددلاین تمرین : ۱۵ فروردین

تمرین دوم

- برای ارسال هر تمرین تا ساعت ۲۳:۵۹ روز ددلاین فرصت دارید. مهلت تاخیر (مجاز و غیر مجاز) برای این تمرین، ۷ روز است (یعنی حداکثر تاریخ ارسال تمرین ۲۲ فروردین است)
- مجموع نمرات تمرین نظری برابر ۱۱۰ می باشد که ۱۰ نمره آن جنبه اختیاری دارد و گرفتن نمره ۱۰۰ دریافت نمره کامل بخش نظری کفایت میکند و البته اگر نمره ای بالاتر از ۱۰۰ کسب کنید همان نمره ۱۰۰ برای شما در نظر گرفته میشود
- در هر کدام از سوالات، اگر از منابع خارجی استفاده کرده اید باید آن را ذکر کنید. در صورت همفکری با افراد دیگر هم باید نام ایشان را در سوال مورد نظر ذکر نمایید.
- پاسخ تمرین باید ماحصل دانسته های خود شما باشد. در صورت رعایت این موضوع، استفاده از ابزارهای هوش مصنوعی با ذکر نحوه و مصداق استفاده بلامانع است.
- 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 اسکرین شات از منابع یا پاسخ افراد دیگر نمره ای تعلق نمی گیرد.
- در صورتی که بخشی از سوال ها را جای دیگری آپلود کرده و لینک آن را قرار داده باشید، حتما باید تاریخ آپلود مشخص و قابل اتکا باشد.
- محل بارگذاری سوالات نظری و عملی در هر تمرین مجزا خواهد بود. به منظور بارگذاری بایستی تمارین تئوری در یک فایل pdf با نام `HW2_[First-Name]_[Last-Name]_[Student-Id].pdf` و تمارین عملی نیز در یک فایل مجزای زیپ با نام `HW2_[First-Name]_[Last-Name]_[Student-Id].zip` بارگذاری شوند.
- در صورت وجود هرگونه ابهام یا مشکل، در کوثرای درس آن مشکل را بیان کنید و از پیغام دادن مستقیم به دستیاران آموزشی خودداری کنید.
- طراحان این تمرین : علی رحیمی اکبر- امیرحسین ایزدی- امیرحسین علمدار- محمد مهدی واحدی- مهرداد مهابادی

بخش نظری (۱۰+۱۰۰ نمره)

پرسش ۱. Batch Normalization (۲۰ نمره)

۱. نحوه انجام نرمال سازی بچ در شبکه های تماماً متصل و شبکه های پیچشی را با یکدیگر مقایسه کنید. همچنین نحوه اعمال نرمال سازی بچ در مرحله آموزش و آزمایش را نیز با یکدیگر مقایسه کنید.
۲. به صورت خلاصه Covariate shift را توضیح دهید و توضیح دهید چرا در نرمال سازی بچ، Covariate shift مابین داده های آموزش و آزمایش منجر به ناپایدار شدن در نتایج مدل برای دادگان آزمایش می شود؟

۳. شبکه CNN ای را در نظر بگیرید که از بلاک‌هایی به فرم زیر استفاده می‌کند:

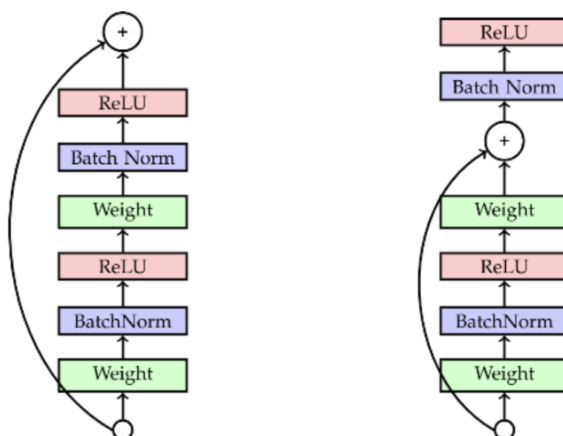
(ConvLayer) → (BatchNorm) → (Activation)

آیا حذف بایاس (b) از لایه کانولوشن در کارکرد این شبکه اختلال ایجاد می‌کند؟ چرا؟ همچنین فرض کنید شبکه را آموزش داده‌ایم؛ آیا ضرب کردن وزن‌ها در یک عدد مانند α در زمان آزمایش (Inference)، عملکرد شبکه را تغییر می‌دهد؟ ضرب کردن این ضریب در تمام درایه‌های ورودی شبکه چگونه؟

۴. نشانی دهید که نرمال‌سازی بچ، باعث ایجاد نویزی در برآورد مقادیر گرادیان‌ها در مرحله آموزش می‌شود که به‌طور ضمنی یک منظور ساز است. این اثر را با اثر منظم سازی dropout مقایسه کنید.

۵. بررسی کنید که آیا استفاده پشت سر هم بلوک نرمال‌سازی بچ و dropout عموماً می‌تواند مفید باشد؟ چرا؟

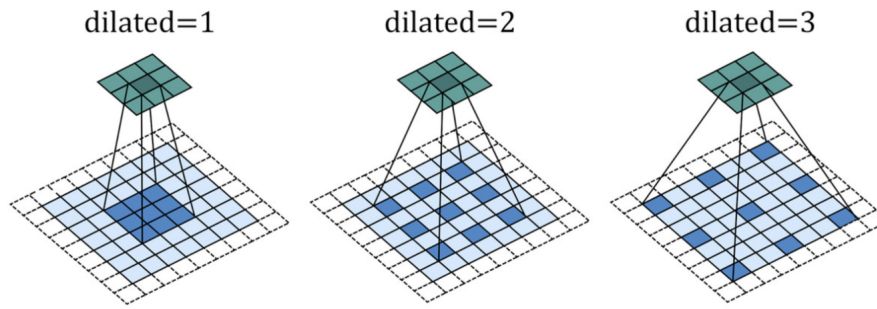
۶. شبکه‌های باقی‌مانده (ResNet) نقش مهمی در بهبود یادگیری عمیق داشته‌اند و امکان آموزش مدل‌های بسیار عمیق را فراهم کرده‌اند. با این حال، مکان قرارگیری نرمال‌سازی بچ (BN) نسبت به اتصالات میان‌بر تأثیر قابل توجهی بر پایداری آموزش، تعمیم‌پذیری و رفتار مدل در مرحله تست دارد. دو طراحی متفاوت برای بلوک باقی‌مانده را در نظر بگیرید که به بلوک باقی‌مانده با پیش‌فعال‌سازی و پس‌فعال‌سازی معروف است:



با در نظر گرفتن تأثیر نرمال‌سازی دسته‌ای بر انتشار گرادیان، پایداری بهینه‌سازی، سازگاری بین آموزش و تست، تغییر واریانس، یادگیری نمایش‌های عمیق، تحلیل کنید که چگونه مکان BN می‌تواند دینامیک آموزش شبکه را تحت تأثیر قرار دهد. (توجه کنید که منظور از Weight می‌تواند لایه تماماً متصل و یا پیچشی باشد).

پرسش ۲. Dilated Convolution (۱۵ نمره)

در شبکه‌های پیچشی به صورت متداول از لایه‌های کانولوشن ساده استفاده می‌شود که با آن آشنا هستید. نوع دیگری از لایه‌ها که می‌توان از آنان در شبکه‌های پیچشی استفاده نمود، لایه کانولوشن گسترش یافته یا متسع است. در شکل ۵ تصویر شهودی از فیلتر کانولوشن گسترش یافته ارائه شده است، این فیلترها میان‌خانه‌هایی که فیلتر با استفاده از اطلاعات آن‌ها لایه بعد را محاسبه می‌کند فاصله می‌اندازند یا به بیانی دیگر در زمان اعمال فیلتر و انجام عملیات ضرب کانولوشن، بر روی ورودی با گام (dilated) بزرگتری حرکت می‌کنیم، توجه کنید طول گام مفهومی متفاوت نسبت به طول گام (stride) در لایه‌های شبکه کانولوشن دارد.



شهودی از کانولوشن گسترش یافته با گام های متفاوت

همانطور که در شکل ۱ نیز مشخص است این روش، یک روش کم هزینه برای افزایش محدوده دید شبکه های پیچشی است. کانولوشن گسترش یافته بصورت فرم بسته ریاضی زیر تعریف می شود.

$$(K \star_D I)(i, j) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} K(m, n) I(i + Dm, j + Dn)$$

فرض کنید یک شبکه عصبی کانولوشنال با L لایه طراحی شده است که هر لایه شامل فیلترهای کانولوشن با پارامترهای زیر است:

- اندازه فیلتر (Kernel Size): $k_\ell \times k_\ell$ (مربعی)،

- نرخ اتساع (Dilation Rate): d_ℓ ،

- گام (Stride): s_ℓ ،

- بدون پدینگ (No Padding).

هدف ما بررسی تأثیر پارامترها بر **گستره دید نسبی** (Relative Receptive Field) است. گستره دید نسبی به صورت نسبت گستره دید در خروجی لایه L -ام به اندازه ورودی اصلی ($M \times N$) تعریف می شود.

۱. فرمول کلی گستره دید نسبی $R_{\text{relative}}^{(L)}$ را به صورت تابعی از k_ℓ ، d_ℓ و s_ℓ استخراج کنید.

۲. فرض کنید هدف ما این است که گستره دید نسبی بیشتر از یک حد آستانه (T) باشد، در حالی که هزینه های محاسباتی (FLOPs) کمترین مقدار ممکن باشد:

- معادله ای برای تعیین شرایط بهینه d_ℓ و s_ℓ بنویسید.

- آیا این شرایط بهینه به تعداد لایه ها (L) وابسته است؟ چرا؟

پرسش ۳. ROI Alignment (۱۰ نمره)

۱. یکی از مسائل در برخی روش های تشخیص شی، ROI Alignment است. نحوه ی کار کرد درونیابی خطی و نزدیک ترین همسایه را توضیح دهید. برای درونیابی خطی روابط مربوطه را بنویسید.

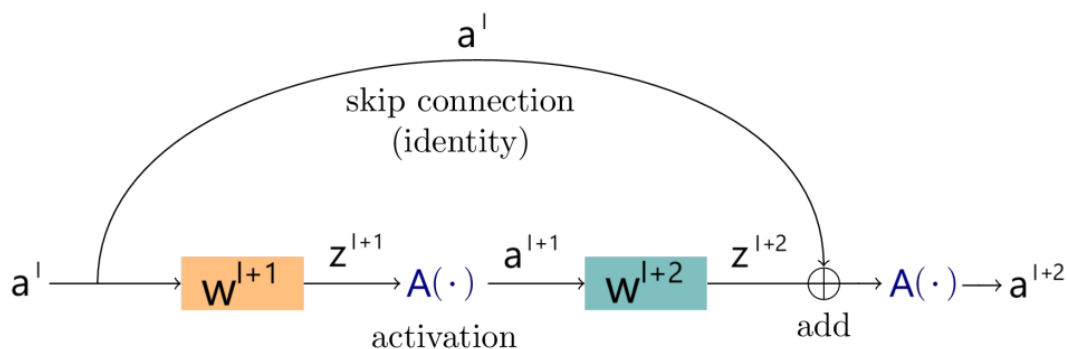
۲. یک عکس ۳۲ در ۳۲ را در نظر بگیرید، فرض کنید به یک activation map ۱۰ در ۱۰ تبدیل شده باشد. مقدار متناظر با نقطه ی $x=4$ و $y=8$ در عکس اولیه را در نقشه ی نهایی برحسب مقادیر پیکسل های نقشه محاسبه کنید.

پرسش ۴. Convolution Gradient (۱۵ نمره)

۱. بردار یک بعدی \vec{x} با چهار درایه را در نظر بگیرید. فرض کنید روی این بردار یک کانولوشن یک بعدی با ساین کرنل ۳ و Padding ۱ اعمال کنیم. عملیات انجام شده را به فرم ماتریسی بنویسید و خروجی را محاسبه کنید.
۲. حال فرض کنید یک تابع زیان روی خروجی این لایه اعمال شده و به زیان L رسیده‌ایم. مشتق L را نسبت به \vec{x} با کمک قاعده زنجیره‌ای محاسبه کنید.
۳. به طور دقیق مشخص کنید باید چه عملیاتی روی مشتق جزئی تابع زیان نسبت به خروجی این لایه انجام دهیم تا مشتق جزئی زیان نسبت به بردار \vec{x} بدست آید؟
۴. حال فرض کنید بردار \vec{x} به طول چهار به شما داده شده است و می‌خواهید با کمک upsampling آن را به فضای \mathbb{R}^6 ببرید. برای اینکار از Transpose Convolution با padding صفر و stride یک استفاده می‌کنیم. اگر کرنل ما \vec{w} به طول سه باشد عملیات را به فرم ماتریسی بنویسید. ماتریس حاصل را با ماتریس بخش ۱ مقایسه کنید. اگر padding یک باشد چه اتفاقی می‌افتد؟

پرسش ۵. محو شدن گرادیان (۱۵ نمره)

یکی از مشکلاتی که در الگوریتم‌های انتشار به عقب (back propagation) وجود دارد بحث محو شدن گرادیان (Gradient Vanishing) است. این موضوع مهم و قبل از اهمیت زیاد باعث عدم آموزش درست و کامل مدل در روند آموزش می‌شود. در این مسئله قصد داریم به بررسی این اتفاق بپردازیم.



۱. ابتدا مقدار $\frac{\partial a^{l+2}}{\partial a^l}$ را بدون در نظر گرفتن skip connection محاسبه کنید.
۲. حال یکی از راه حل‌ها که در بسیاری از مدل‌ها استفاده می‌شود در نظر گرفتن skip connection می‌باشد. این حالت چه کمکی به مدل می‌کند؟ $\frac{\partial a^{l+2}}{\partial a^l}$ با کمک قاعده زنجیره‌ای محاسبه کنید.
۳. با توجه به نتایج دو قسمت قبل بگویید که این الگوریتم به چه صورت می‌تواند مشکل محو شدن گرادیان را حل کند. (فرض کنید که داریم $W^i < 1 - \epsilon$)

پرسش ۶. MobileNet (۳۵ نمره)

معماری‌های MobileNet (شامل نسخه‌های V1، V2، و V3) از جمله شبکه‌های عصبی پیچشی سبک‌وزن هستند که به طور خاص برای اجرا بر روی دستگاه‌های کم‌مصرف مانند گوشی‌های هوشمند و سخت‌افزارهای لبه طراحی شده‌اند.

این مدل‌ها با کاهش تعداد محاسبات و پارامترها، بدون افت چشمگیر در دقت، توانسته‌اند به تعادلی میان کارایی و عملکرد دست یابند. از مهم‌ترین نوآوری‌های به‌کاررفته در این معماری‌ها می‌توان به کانولوشن‌های عمقی قابل تفکیک (Depthwise Separable Convolutions)، بلوک‌های باقیمانده معکوس (Inverted Residual Blocks) دارای گلوگاه‌های خطی (Linear Bottlenecks)، مکانیزم‌های فشرده‌سازی و تحریک (SE Blocks) و استفاده از جستجوی معماری عصبی (NAS) اشاره کرد. در این سوال به بررسی برخی از این موارد می‌پردازیم. (لازم به ذکر است برای حل این سوال پیشنهاد اکید میشود از سرچ در منابع مختلف برای mobilenet بهره ببرید)

۱) کانولوشن‌های عمقی قابل تفکیک و تئوری تقریب
کانولوشن‌های عمقی قابل تفکیک سنگ بنای شبکه‌های پیچشی سبک‌وزن هستند که باعث کاهش قابل توجه هزینه‌های محاسباتی می‌شوند و در عین حال ظرفیت نمایشی منطقی را حفظ می‌کنند.

۱-۱. کانولوشن عمقی قابل تفکیک را بطور کامل حین مقایسه با کانولوشن عادی، توضیح دهید (به صورت ریاضی).

۱-۲. عمل ریاضی کانولوشن استاندارد را در نظر بگیرید و آن را بصورت مجموع تنسورهای رتبه یک با استفاده از تجزیه مقادیر تکین بیان کنید. توضیح دهید این تجزیه چگونه به ساختار کانولوشن عمقی قابل تفکیک مرتبط است.

۱-۳. نسبت کاهش FLOPs در کانولوشن‌های عمقی قابل تفکیک نسبت به کانولوشن‌های استاندارد را برای اندازه ورودی $H \times W$ ، تعداد کانال‌های ورودی C_{in} ، تعداد کانال‌های خروجی C_{out} و اندازه فیلتر K را بطور پارامتری محاسبه کنید. آیا بین ظرفیت نمایشی و هزینه‌های محاسباتی ناشی از این تقریب تضادی وجود دارد؟

۲) بلوک‌های باقیمانده معکوس و تحلیل جریان گرادیان
بلوک‌های باقیمانده معکوس، که در MobileNetV2 معرفی شدند، با ترکیب باقی‌مانده معکوس و گلوگاه‌های خطی پیشرفتی در شبکه‌های پیچشی سبک‌وزن ایجاد کردند و باعث بهبود ظرفیت نمایشی در حالی که هزینه‌های محاسباتی کاهش می‌یابد، شدند.

۲-۱. بلوک‌های باقیمانده معکوس را در حین مقایسه با بلوک‌های باقیمانده عادی (ResNet)، توضیح دهید.

۲-۲. آیا این ادعا که کانولوشن‌های عمقی قابل تفکیک اطلاعات مکانی را حفظ می‌کنند، در حالی که گلوگاه‌های خطی اطلاعات کانالی را حفظ می‌کنند، صحیح است؟

۲-۳. تحلیل کنید که چگونه ضریب انبساط t بر جریان گرادیان و پایداری بهینه‌سازی تأثیر می‌گذارد. ثابت کنید که افزایش t خطر ناپدید شدن گرادیان‌ها را کاهش می‌دهد، اما هزینه‌های محاسباتی را افزایش می‌دهد. مقدار بهینه t را که تعادلی بین جریان گرادیان و کارایی ایجاد می‌کند، بر چه اساسی می‌توان انتخاب کرد؟

۳) مکانیزم‌های فشرده‌سازی و تحریک و بازنگری ویژگی‌های کانالی
بلوک‌های فشرده‌سازی و تحریک، که در MobileNetV3 استفاده شدند، تنظیم تطبیقی پاسخ‌های ویژگی در سطح کانال افزایش می‌دهند. این بلوک‌ها با استفاده از یک مکانیزم توجه (attention) وزن‌های کانال‌ها را مطابق با اطلاعات جهانی موجود در نقشه‌های ویژگی ورودی خود تنظیم می‌کنند.

۳-۱. اعمالی که در یک بلوک فشرده‌سازی و تحریک در یک لایه رخ می‌دهد را به صورت ریاضی فرمول بندی کنید. فرض کنید نقش ویژگی ورودی این بلوک با ابعاد $C \times H \times W$ است.

۳-۲. ثابت کنید که بلوک‌های SE ظرفیت نمایشی را با بازنگری تطبیقی پاسخ‌های ویژگی‌های کانالی بهبود می‌دهند. از تئوری اطلاعات برای اندازه‌گیری اطلاعات متقابل بین ورودی و خروجی بلوک SE استفاده کنید. توضیح دهید که ضریب کاهش r چگونه بر تضاد بین هزینه‌های محاسباتی و عملکرد تأثیر می‌گذارد.

۳-۳. FLOPs اضافه شده توسط یک بلوک SE با $r = 16$ برای $C = 64$ را محاسبه کنید. این مقدار را با کل FLOPs یک کانولوشن عمقی قابل تفکیک با پارامترهای مشابه مقایسه کنید.

۴) جستجوی معماری عصبی (NAS) و تکنیک کوچک‌سازی پیش‌رونده جستجوی معماری عصبی (NAS) برای بهینه‌سازی معماری MobileNetV3 استفاده شد و منجر به بهترین سطح عملکرد در دستگاه‌های موبایل شد.

۴-۱. این فرآیند را بطور کامل تحلیل کنید. فضای جستجو برای NAS در MobileNetV3 چه بوده است؟ از نظریه گراف برای مدل‌سازی فضای جستجو به عنوان یک گراف جهت‌دار بدون دور (DAG) چگونه می‌توان استفاده کرد؟

۴-۲. تکنیک کوچک‌سازی پیش‌رونده استفاده‌شده در NAS را توضیح دهید.

بخش عملی (۱۰۰ نمره)

در این مجموعه سه تمرین ارائه شده است. از شما خواسته می‌شود هر تمرین را به صورت جداگانه در سامانه Quera بارگذاری نمایید. هر سوال دارای توضیحات لازمه داخل نوت‌بوک خود می‌باشد.