



تمرین سری چهارم

درس یادگیری عمیق

نام مدرس: دکتر محمدرضا محمدی

دستیاران آموزشی مربط: حجت شهرابی،
آیسا میاهی نیا، بهداد نادری فرد، امیر جبلی، ساره
افتخاری، دنیا اسفندیارپور

مهلت تحويل (بدون کسر نمره):
۲۴ آذر ماه

سوالات تئوری (۳۸ نمره)

سوال ۱ - ۱۴ نمره

شبکه عصبی کانولوشنی زیر را در نظر بگیرید. فرض کنید تصویر ورودی رنگی با اندازه 128×128 در اختیار داریم:

Layer1: Conv2d(in_channels=3, out_channels=32, kernel_size=(7,7), stride=1, padding=0)

Layer2: Conv2d(in_channels=32, out_channels=64, kernel_size=(5,5), stride=2, padding=0)

Layer3: MaxPool2d(kernel_size=(2,2), stride=2)

Layer4: Conv2d(in_channels=64, out_channels=128, kernel_size=(3,3), stride=1, dilation=2, padding=0)

Layer5: Conv2d(in_channels=128, out_channels=128, kernel_size=(3,3), stride=1, dilation=1, padding=0)

Layer6: MaxPool2d(kernel_size=(2,2), stride=2)

Layer7: Conv2d(in_channels=128, out_channels=256, kernel_size=(3,3), stride=1, padding=0)

Layer8: AvgPool2d(kernel_size=(2,2), stride=2)

Layer9: Linear(N, 1024) → Linear(1024, 1024)

Layer11: Dropout(0.5)

Layer12: Linear(10)

- نکته: در لایه ۹ دو لایهی خطی متوالی داریم، N همان تعداد ویژگی‌های حاصل از flatten خروجی لایه ۸ است،
و هر دو لایه ۱۰۲۴ نرون دارند.

(الف) اندازه خروجی، تعداد پارامترها (وزن‌ها و بایاس‌ها جداگانه اعلام شود) و میدان تأثیر هر لایه را با ذکر راه حل به تفکیک محاسبه کنید. (۸ نمره)

(ب) تعداد کل عملیات ضرب و جمع (FLOPs) برای هر لایه را محاسبه کنید. (۳ نمره)

(ج) اگر تغییرات زیر اعمال شود، اثر هر کدام را بنویسید: (۲ نمره)

1. اگر stride لایه دوم نصف شود

2. اگر kernel لایه چهارم به 5×5 تغییر کند

3. اگر کانال‌های لایه هفتم دو برابر شود

4. اگر لایه AvgPool حذف شود

سوال ۲ - ۱۴ نموده

برای سوالات زیر پاسخ‌های مستدل ارائه دهید:

الف) چرا شبکه‌های عصبی کاملاً متصل (Fully Connected) برای پردازش داده‌هایی مانند تصاویر یا سیگنال‌های صوتی مناسب نیستند؟ (۱/۵ نمره)

ب) چرا در شبکه‌های عصبی کانولوشنی (Convolutional Neural Network) وزن‌های یک فیلتر در مکان‌های مختلف تصویر به اشتراک گذاشته می‌شوند؟ این اشتراک وزن چه تأثیری بر تعداد پارامترهای مدل و توانایی شبکه در تشخیص الگوها دارد؟ (۱/۵ نمره)

ج) در لایه‌های کانولوشنی، هر نورون خروجی فقط به بخش کوچکی از نورون‌های لایه‌ی قبلی متصل است. این محدود بودن اتصال هر نورون به یک ناحیه‌ی کوچک از ورودی، چه نتیجه‌ای به دنبال دارد؟ مفهوم "میدان تأثیر" (receptive field) را در این رابطه توضیح دهید. (۲ نمره)

د) خاصیت هم‌تغییر بودن بازنمایی (Equivariant Representation) در شبکه‌های کانولوشنی را توضیح دهید. منظور از Equivariant بودن ویژگی‌های استخراج شده چیست و این خاصیت چه اهمیتی در پردازش تصویر دارد؟ (۲ نمره)

ه) در چه شرایطی در پردازش تصویر ممکن است به اشتراک گذاری وزن در لایه‌های کانولوشنی مطلوب نباشد؟ اگر در یک لایه‌ی کانولوشنی وزن‌ها در سراسر تصویر به صورت مشترک استفاده نشوند (یعنی هر ناحیه‌ی مکانی از تصویر فیلترهای مستقلی داشته باشد)، خاصیت Equivariant بودن شبکه چگونه تحت تأثیر قرار می‌گیرد؟ (۱/۵ نمره)

و) افزایش تعداد فیلترهای کانولوشنی در یک لایه چه تأثیری بر قابلیت آن در تحلیل تصویر دارد؟ آیا افزودن فیلترهای بیشتر همیشه به بهبود عملکرد شبکه منجر می‌شود؟ دلایل خود را شرح دهید. (۱ نمره)

ز) در شبکه‌های CNN اغلب بیشترین تعداد پارامترها که باعث overfit شدن می‌شوند در لایه‌ی FC آخر هستند، چرا؟ ایده‌ی Stride بزرگ‌تر از ۱ چگونه به کاهش overfit ناشی از این لایه کمک می‌کند؟ اعمال گام بزرگ‌تر چه اثری بر تعداد پارامترهای مدل و هزینه‌ی محاسباتی آن دارد؟ تفاوت Down Sampling و Strided Sampling را بیان کنید. در عمل کدام یک انجام می‌شود؟ (۱/۵ نمره)

ط) توضیح دهید که عملیات Pooling چگونه باعث می‌شود بازنمایی ویژگی‌های شبکه نسبت به تغییرات یا جابجایی‌های کوچک در ورودی تا حدی (invariant) باشد. این خاصیت چه کمکی به پایداری تشخیص الگوها می‌کند؟ (۰/۵ نمره)

ی) ترکیب به کارگیری گام (stride) و لایه Pooling در یک شبکه کانولوشنی چه مزایایی دارد؟ آیا بهتر است ابتدا از یک کانولوشن stride دار استفاده کنیم و سپس pooling انجام دهیم یا برعکس؟ دلیل خود را توضیح دهید. (۰/۵ نمره)

ک) اگر در پیش‌پردازش تصاویر، نسبت ابعاد (عرض به ارتفاع) آن‌ها حفظ نشود و تصاویر به شکل غیریکسان در جهات مختلف کشیده/فسرده شوند، چه تأثیری بر عملکرد یک شبکه عصبی خواهد داشت؟ آیا تغییر نسبت طول به عرض تصویر می‌تواند برای یادگیری شبکه مشکل‌ساز باشد؟ دلایل خود را شرح دهید. (۱ نمره)



سوال ۳ - ۴ نمره

فرض کنید تصاویر ورودی 128×128 پیکسل (خاکستری) دارند. شبکه کانولوشنال ساده زیر را در نظر بگیرید:

- لایه کانولوشن 3×3 با 1×1 کanal ورودی و 8 کanal خروجی.
- لایه $\text{Stride}=2$ با 2×2 MaxPooling
- لایه کانولوشن 3×3 با 1×1 کanal ورودی و 16 کanal خروجی.
- لایه $\text{Stride}=2$ با 2×2 MaxPooling
- لایه کانولوشن 3×3 با 1×1 کanal ورودی و 32 کanal خروجی.
- لایه $\text{Stride}=2$ با 2×2 MaxPooling

اگر بخواهیم receptive field نهایی دست‌کم 50×50 پیکسل باشد، پیشنهاد دهید چگونه معماری شبکه را تغییر دهیم (برای مثال اندازه فیلترها، Stride یا استفاده از لایه‌های اضافه). تصمیم خود را با استدلال توضیح دهید.

سوال ۴ - ۸ نمره

یک شبکه عصبی کانولوشنال ساده برای طبقه‌بندی تصاویر با ابعاد ورودی 256×256 (سه کanal رنگی) در نظر بگیرید. هدف این است که در نهایت نقشه ویژگی (Feature Map) خروجی ابعاد 16×16 داشته باشد (یعنی کاهش ابعاد ورودی به ضریب ۱۶). دو روش کاهش ابعاد زیر را مقایسه و تحلیل کنید:

- روش A : استفاده از چهار لایه کانولوشن 3×3 با $\text{Stride}=2$ (بدون استفاده از Pooling)، به طوری که هر لایه ابعاد نقشه ویژگی را به نصف کاهش می‌دهد.
- روش B : استفاده از سه بلوک متوالی (هر بلوک شامل یک لایه کانولوشن 3×3 با $\text{Stride}=1$ و یک لایه $\text{Stride}=2 \times 2$ MaxPooling)

- الف) ابعاد نقشه ویژگی خروجی را برای هر یک از روش‌های A و B حساب کنید. (۲ نمره)
- ب) میدان تاثیر (receptive field) را برای هر روش تحلیل کنید. (۳ نمره)
- ج) مزایا و معایب هر روش را از نظر حفظ اطلاعات مکانی (spatial information) و خصوصیت ثبات پاسخ در برابر جابجایی تصویر مقایسه کنید. (۲ نمره)
- د) پیشنهاد دهید در چه سناریوهایی استفاده از لایه‌های Pooling به جای Stride (یا بالعکس) مناسب‌تر است. استدلال خود را بیان کنید. (۱ نمره)

سوالات عملی (۷۰ نمره)

سؤال ۵ - ۳۰ نمره

مقاله [Xception](#) را دانلود و مطالعه کنید. نوآوری آن را در گزارش به طور کامل شرح دهید.
مدل را مطابق جزئیات مقاله پیاده سازی کنید.

برای انجام طبقه‌بندی باینزی به منظور تشخیص سرطان پروسات در این سوال، از ستون csPca فایل اکسل slices_info استفاده خواهیم کرد. در این تسک، هدف این است که از ویژگی‌های موجود در داده‌ها برای تشخیص وجود یا عدم وجود سرطان پروسات استفاده کنیم.

لود کردن داده‌ها و رسم هیستوگرام: ابتدا داده‌ها را بارگذاری کرده و هیستوگرام آن‌ها را رسم کنید. پس از رسم هیستوگرام، تحلیل کنید و بیان کنید که چه مشکلات یا چالش‌هایی ممکن است در داده‌ها وجود داشته باشد.

لینک داده‌ها: [دریافت فایل](#)
آموزش مدل و گزارش نتایج:

- مدلی را که در مرحله قبل پیاده‌سازی کردید، روی این دیتاست آموزش دهید. نتایج حاصل از آموزش مدل را گزارش کنید و به تحلیل آن بپردازید. در این مرحله، تحلیل‌های زیر را انجام دهید:
 - دقت کلی (accuracy) مدل را محاسبه کنید.
 - کانفیوژن ماتریس را رسم کرده و تحلیل کنید.
 - امتیاز (AUC) Area Under Curve را گزارش دهید و تحلیل کنید.
 - دقت مدل را برای هر کلاس به طور جداگانه بررسی کنید.
- در نهایت، ارزیابی کنید که آیا دقت کلی (accuracy) معیاری مناسب برای ارزیابی مدل در این مورد است یا خیر. چرا؟

آموزش مدل EfficientNet : در این مرحله، مدل EfficientNet را آموزش دهید و مراحل قسمت قبل را تکرار کنید. (نیاز به پیاده‌سازی مدل نیست)

بهبود عملکرد مدل: با توجه به بالانس نبودن کلاس‌ها و نتایج به دست آمده از دو مدل مختلف، یک مدل را انتخاب کنید و سعی کنید آن را با استفاده از روش‌های مناسب بهبود دهید. در این مرحله، دلایل انتخاب مدل را توضیح دهید و سپس تغییرات اعمال شده را شرح دهید. در نهایت، نتایج جدید را تحلیل کرده و مقایسه کنید که مدل بهبود یافته چگونه عملکرد بهتری نسبت به مدل اولیه ارائه داده است.

در صورتی که ماسک segmentation پروسات در دسترس باشد، چگونه می‌توان از آن برای بهبود مدل طبقه‌بندی باینزی در تشخیص سرطان پروسات استفاده کرد؟

در این سوال، تمرکز بیشتر بر روی تحلیل روش‌ها و مدل‌های مختلف است تا رسیدن به عملکرد مناسب نهایی. بررسی اینکه چرا یک روش نسبت به روش دیگر بهتر یا بدتر عمل کرده است و چه تغییراتی در معماری، داده‌ها یا تنظیمات های پارامترها می‌تواند به بهبود عملکرد کمک کند.

سوال ۶ - نمره ۲۰

بهینه‌سازی، فشرده‌سازی و تحلیل امنیت مدل‌های یادگیری عمیق

مدل هدف: MobileNetV2

دیتاست‌ها:

1. CIFAR-10 (تصاویر 32×32 پیکسل - رزولوشن پایین)

2. STL-10 (تصاویر 96×96 پیکسل - رزولوشن بالا)

هدف این پژوهه، پیاده‌سازی و مقایسه‌ی جامع تکنیک‌های آموزش (Training)، فشرده‌سازی (Compression) و استقرار (Deployment) مدل‌های عصبی است. دانشجویان موظف‌اند یک پایپ‌لاین کامل را روی دو دیتاست با کیفیت‌های متفاوت اجرا کرده و نتایج را تحلیل کنند.

فاز اول: استراتژی‌های آموزش (Training Strategies)

در این فاز باید مدل MobileNetV2 را با سه رویکرد زیر آموزش دهید و دقت نهایی را ثبت کنید:

۱. تنظیم جزئی (Partial Fine-Tuning)

روش: بارگذاری وزن‌های پیش‌آموزش‌دیده (ImageNet)، فریز کردن (Freezing) تمام لایه‌های استخراج ویژگی و آموزش تنها لایه طبقه‌بند (Classifier).

هدف: ایجاد خط مبنا (Baseline) برای سنجش سرعت همگرایی و قدرت ویژگی‌های اولیه.

۲. تنظیم دقیق کامل (Full Fine-Tuning)

روش: آنفریز کردن کل شبکه پس از آموزش اولیه و آموزش مجدد تمام لایه‌ها با نرخ یادگیری بسیار پایین ($1e-5$).

هدف: دستیابی به بالاترین دقت ممکن (Standard Baseline).

۳. تقطیر دانش (Knowledge Distillation)

توضیح: انتقال دانش از یک مدل بزرگ (Teacher) به مدل کوچک (Student).

روش: استفاده از روش هینتون (Hinton's Response-Based Knowledge Distillation).

معلم:

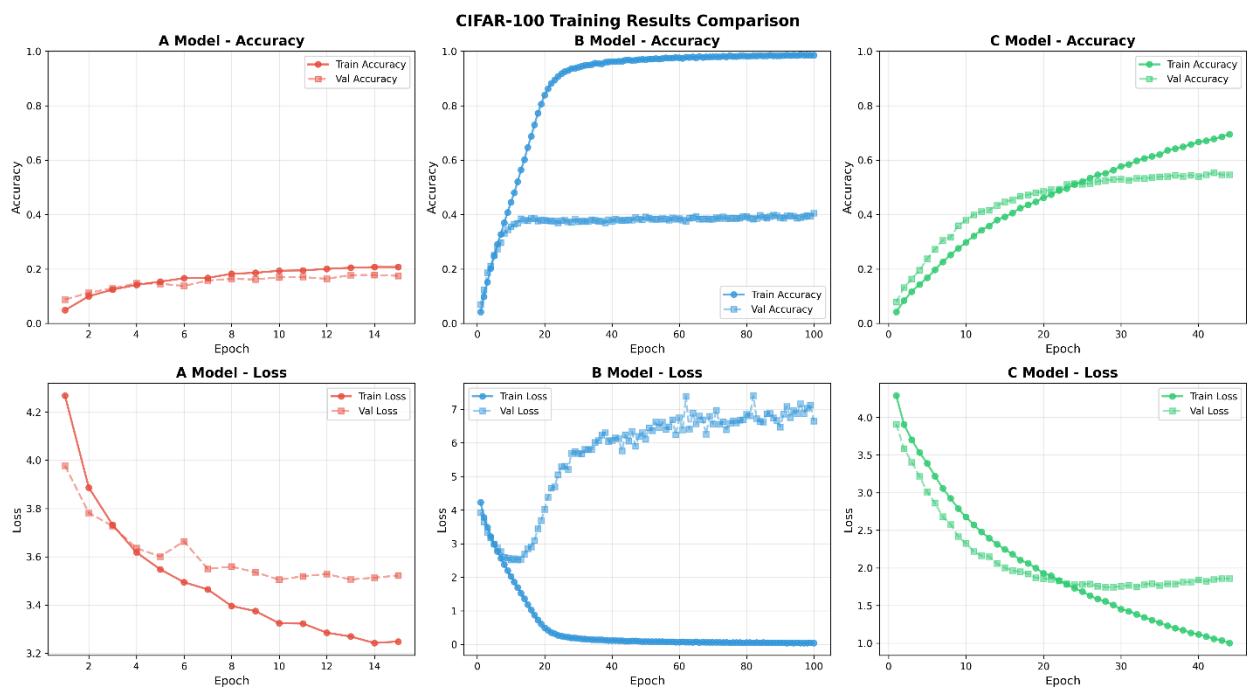
- ResNet-50
- دانش‌آموز: MobileNetV2 (با مقداردهی تصادفی)
- تابع هزینه (Loss): ترکیبی از CrossEntropy (برای لیل‌های واقعی) و KL-Divergence (برای نزدیک کردن توزیع احتمال خروجی دانش‌آموز به معلم با پارامتر دما T).

جدول نهایی گزارش (Deliverable Table)

در پایان پژوهه، در یک جدول نتایج به دست آمده را با هم مقایسه کنید.

سوال ۷ - نموده

در این تمرین، سه شبکه عصبی کانولوشنی با رفتارهای یادگیری متفاوت که آنها را با نام‌های Model A، Model B و Model C مشخص می‌کنیم، بر روی دیتاست CIFAR-100 آموزش داده شده‌اند. این دیتاست شامل صد کلاس مختلف با 600 تصویر در هر کلاس است و اندازهٔ کوچک تصاویر (32×32) موجب می‌شود استخراج الگوهای سطح بالا و میان‌سطحی چالش‌برانگیز باشد. برای هر مدل، نمودارهای دقت و خطای آموزش و اعتبارسنجی در اختیار شما قرار داده شده است. یکی از این مدل‌ها توان استخراج ویژگی‌های کافی را از این دیتاست پیچیده ندارد و تنها ساختارهای کلی و کم‌عمق را می‌آموزد (A)؛ مدل دیگر بیش از حد روی جزئیات خاص داده‌های آموزشی متمرکز شده و بازنمایی‌های نامعمول و غیرقابل‌تعیین تولید می‌کند (B)؛ و مدل سوم توانسته تعادلی میان این دو وضعیت برقرار کند (C).



مقاله "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization"

یکی از مهم‌ترین روش‌های تبیینی در شبکه‌های عصبی کانولوشنی را معرفی می‌کند. Grad-CAM، با استفاده از گرادیان‌های مربوط به کلاس هدف، ناحیه‌های مهم تصویر را که برای تصمیم مدل بیشترین نقش را داشته‌اند، به صورت یک نقشهٔ حرارتی مشخص می‌کند. این روش بدون تغییر در معماری شبکه قابل استفاده است و به خوبی نشان می‌دهد مدل در هنگام پیش‌بینی واقعاً (به کجا نگاه کرده است). در این تمرین که هدف آن تحلیل رفتار سه مدل A، B و C است، Grad-CAM ابزار اصلی ما برای بررسی نواحی مرتبط و نامرتبط است که هر مدل در تصمیم‌گیری به آنها تکیه کرده است.

مقاله "Visualizing and Understanding Convolutional Networks" رويکردي مشهور برای بازسازی ورودی

از روی فعال‌سازی‌های شبکه به نام DeconvNet ارائه می‌دهد. این روش با بازگرداندن feature map لایه‌های مختلف به فضای تصویر، نشان می‌دهد هر فیلتر دقیقاً چه الگوهایی را در تصویر تشخیص می‌دهد: از لبه‌ها در لایه‌های

ابتدايی گرفته تا بافت‌ها، الگوهای ميان‌سطحی و ساختارهای کاملاً وابسته به کلاس در لایه‌های عمیق‌تر. در این تمرین، DeconvNet به ما اجازه می‌دهد کیفیت بازنمایی‌های سه مدل را مقایسه کنیم و بفهمیم کدام مدل تنها الگوهای ساده را آموزد، کدام مدل بیش از حد روی جزئیات خاص داده‌های آموزش متمرکز می‌شود، و کدام مدل بازنمایی‌های پایدار و قابل تعمیم ایجاد می‌کند؛ به‌ویژه با توجه به رزولوشن پایین تصاویر 100-CIFAR که ضعف یا قوت مدل‌ها را برجسته‌تر می‌کند.

اکنون شما باید به صورت مستقل مراحل زیر را با استفاده از مدل‌ها انجام دهید.

- استخراج Feature Map های لایه‌های مختلف
- تولید نقشه‌های فعال‌سازی **Grad-CAM**
- برگرداندن بازنمایی‌ها به فضای تصویر با روش **DeconvNet**
- جمع‌آوری نمونه‌هایی از تصاویر درست و غلط طبقه‌بندی‌شده
- محاسبه ماتریس سردرگمی برای هر مدل

سپس بر اساس مشاهدات و نتایج بدست‌آمده، به پرسش‌های زیر پاسخ دهید.

- با بررسی Feature Map ها و بازسازی‌های DeconvNet توضیح دهید هر یک از مدل‌های A، B و C چه نوع ساختارهای تصویری (لبه‌ها، بافت‌ها، الگوهای ميان‌سطحی یا الگوهای وابسته به کلاس) را به درستی یاد نگرفته‌اند یا بیش از حد برجسته کرده‌اند؟
- مشخص کنید کدام مدل بازسازی‌های مبهم، کدام مدل بازسازی‌های بیش‌جزئی‌نگر، و کدام مدل بازسازی‌های پایدار و قابل تعمیم تولید می‌کند؟
- توضیح دهید رزولوشن پایین دیتابست 100-CIFAR چگونه باعث شدت یافتن تفاوت کیفیت بازسازی میان مدل‌ها می‌شود.
- با تحلیل نمودارهای خط‌آ و دقت توضیح دهید کدام مدل تنها الگوهای ساده را آموخته است، کدام مدل وابستگی افراطی به داده‌های آموزشی نشان می‌دهد، و کدام مدل الگوی یادگیری متعادلی دارد؟
- نشان دهید چگونه کیفیت بازسازی‌ها و استخراج ویژگی‌ها این رفتارها را تأیید می‌کنند؟ (همچنین به نقش تنوع 100 کلاسی دیتابست در برجسته‌تر شدن تفاوت مدل‌ها اشاره کنید).

با بررسی نقشه‌های Grad-CAM و مقایسه آن با بازنمایی‌های DeconvNet نشان دهید کدام مدل بر نواحی نامرتبط یا نویزی تصویر متمرکز شده است و چرا؟

نکات تکمیلی:

تمام کدها از پایه و بر اساس روند حل سوال و توسط خود دانشجو نوشته شود (استفاده از چت‌بات‌ها مجاز است، به شرط مرجع دادن و تسلط به پاسخ در ارائه).

همچنین در پوشه Practical می‌توانید نسخه کامل و اولیه سوال ۶ را مشاهده کنید که در نهایت تصمیم به کوچکتر کردن آن گرفته شد که در سوال شش قابل ملاحظه است. درصورتی که علاقه‌مند به حل نسخه کامل این تمرین بودید، لطفاً اطلاع دهید و به تناسب پاسخ شما تا نیم نمره  می‌توانید نمره اضافی بگیرید.

دانشجویان محترم حتماً فایل قوانین را مطالعه کرده و در انجام و ارسال تمارین رعایت بفرمایید.

موفق و سربلند باشید.