



## تکلیف شماره ۴

- قبل از شروع تمرین، فایل مربوط به قوانین حل و تحویل تمرین‌ها را مطالعه کنید.
- سوالات و مشکلات خود را درباره این تمرین می‌توانید در گروه تلگرامی درس یا با طراحان این تمرین مطرح کنید.
- نوشتن گزارش کامل و تفسیر نتایج اجباری است. جزئیاتی مانند روش‌های مورد استفاده، تأثیر هر روش در نتیجه نهایی و بهبود معیارهای ارزیابی خواسته شده در گزارش ضروری است. با هر تغییر و هر بهبود مقادیر معیارهای ارزیابی نیز ذکر شود. ضمن اینکه برای گزارش سوالات عملی، نیاز به فایل دیگری نیست و در صورت تمایل در همان محیط کدنویسی، در قالب بلاک‌های مارک‌داون و نوشتار، توضیحات مورد نیاز را قرار دهید.
- پاسخ سوالات تئوری را در یک فایل پی‌دی‌اف<sup>۱</sup> قرار دهید.
- مهلت ارسال پاسخ‌ها: سه‌شنبه ۱۱ دی ساعت ۲۳:۵۹
- طراحان این تمرین: [ماهان ویسی](#) – [امیر ایزدی](#)

۳۵+۱۰ نمره

## سوالات تئوری

### سؤال ۱ (۱۵ نمره)

شبکه‌های عصبی پیچشی عمیق<sup>۲</sup> توانایی یادگیری ویژگی‌های سلسله‌مراتبی را دارند، اما معماری‌های بسیار عمیق اغلب با چالش‌هایی مانند مشکل کاهش عملکرد<sup>۳</sup> مواجه می‌شوند، به‌طوری‌که افزایش عمق شبکه باعث بدتر شدن عملکرد در آموزش و تست می‌شود.

الف) توضیح دهید چرا این اتفاق رخ می‌دهد و بر روی این تمرکز کنید که چگونه ناپدید شدن گرادیان‌ها<sup>۴</sup> باعث اختلال در جریان گرادیان در لایه‌های پیچشی و کاهش کارایی یادگیری می‌شود.

ب) در شبکه‌های عصبی پیچشی، مقداردهی اولیه وزن‌ها<sup>۵</sup> مقادیر اولیه فیلترهای پیچشی را تعیین می‌کند. توضیح دهید چرا مقداردهی اولیه فیلترها با صفر منجر به تقارن در به‌روزرسانی‌ها و شکست در آموزش می‌شود. اصول اولیه مقداردهی Xavier و

<sup>۱</sup> PDF

<sup>۲</sup> Convolutional Neural Networks (CNNs)

<sup>۳</sup> Performance Degradation Problem

<sup>۴</sup> Vanishing Gradients

<sup>۵</sup> Weight Initialization

He را بررسی و مقایسه کنید و توضیح دهید چرا مقداردهی اولیه He برای توابع فعال سازی 'ReLU' مؤثرتر است چگونه باعث تثبیت جریان گرادیان در شبکه‌های عمیق می‌شود.

ج) بلوک‌های Residual، که در ResNet معرفی شده‌اند، یک نوآوری معماری برای حل مشکل کاهش عملکرد در شبکه‌های پیچشی عمیق هستند. توضیح دهید که چگونه اتصالات میانبر در ResNet باعث بهبود جریان گرادیان و دستیابی به عملکرد بهتر در شبکه‌های عمیق می‌شوند.

## سؤال ۲ (۱۰ نمره)

یک لایه کانولوشن<sup>۲</sup> با مولفه های زیر را در نظر بگیرید:

- اندازه فیلتر: ۳ در ۳
- تعداد فیلتر: ۱
- پدینگ: معتبر<sup>۳</sup>
- گام<sup>۴</sup>: ۱

تصویر ورودی به این لایه یک ماتریس ۴ در ۴ با مقادیر زیر است:

$$X = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 1 & 2 & 1 \end{bmatrix}$$

فیلتر اولیه W نیز به صورت زیر تعریف شده است:

$$W = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$$

همچنین فرض کنید تابع هزینه L نسبت به خروجی های لایه کانولوشن به صورت زیر محاسبه شده و مشتق برای خروجی لایه کانولوشن ( که یک ماتریس ۲ در ۲ است) به دست آمده:

$$\frac{\partial L}{\partial Y} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

الف) گرادیان تابع هزینه نسبت به هر وزن<sup>۵</sup> را محاسبه کنید.

<sup>1</sup> Rectified Linear Unit

<sup>2</sup> Convolution

<sup>3</sup> Valid

<sup>4</sup> Stride

<sup>5</sup> Gradient of Loss with respect to each weight

ب) برای محاسبه گرادیان تابع هزینه نسبت به ورودی<sup>۱</sup> از فیلتر  $W$  استفاده کنید و توضیح دهید چگونه این گرادیان می تواند در لایه های قبل تر شبکه به کار رود.

### سؤال ۳ (۱۰ نمره)

برای وظیفه تحلیل احساسات (دسته بندی احساسات یک جمله از منفی تا مثبت (مقادیر صحیح از ۰ تا ۴)) که ورودی ما جمله ای شامل  $n$  کلمه است؛ توضیح دهید که چگونه می توانید از یک RNN برای انجام این وظیفه استفاده کنید. به طور خاص مشخص کنید که چگونه RNN در زمان آزمایش (نه در آموزش) استفاده می شود و مشخص کنید:

الف) چند خروجی، یعنی چند بار تابع  $\text{softmax}(\hat{y}^n)$  از RNN فراخوانی می شود. اگر تعداد خروجی ها ثابت نیست، آن را بصورت متغیر بیان کنید.

ب) هر  $\hat{y}^n$  یک توزیع احتمالاتی روی چه چیزی است (مثلا توزیع احتمالاتی روی تمام گونه سگ ها).

ج) کدام ورودی ها در هر گام زمانی برای تولید هر خروجی شبکه وارد می شوند.

### سؤال ۴ (۱۰ نمره امتیازی)

در حوزه یادگیری عمیق، یکی از چالش های اساسی این است که بفهمیم مدل های پیچیده چگونه تصمیم می گیرند. شبکه های عصبی کانولوشنی در زمینه ی تشخیص و طبقه بندی تصاویر عملکرد بسیار خوبی از خود نشان داده اند. با این حال، این شبکه ها اغلب به عنوان جعبه سیاه<sup>۲</sup> شناخته می شوند، زیرا فهمیدن دلیل انتخاب یک کلاس خاص توسط شبکه دشوار است.

یکی از روش های مهم برای تفسیر و توضیح شبکه های عمیق به ویژه در حوزه بینایی کامپیوتر، Grad-CAM<sup>۳</sup> است (لینک مقاله). هنگامی که یک مدل شبکه عصبی کانولوشنی برای یک وظیفه طبقه بندی تصویری آموزش داده می شود، Grad-CAM

کمک می کند تا بخش هایی از تصویر که بیشترین نقش را در تصمیم گیری مدل برای یک کلاس خاص دارند، برجسته و قابل

مشاهده شوند. به عبارت دیگر، اگر یک شبکه عصبی تصویر را به عنوان ورودی دریافت کرده و تصمیم می گیرد که تصویر متعلق

به یک کلاس خاص (مثلاً «گربه») است، Grad-CAM به ما نشان می دهد کدام قسمت های تصویر برای این نتیجه «گربه»

از نگاه شبکه بیشترین اهمیت را داشته اند. به طور کلی، از اطلاعات گرادیان (مشتق خروجی نسبت به ویژگی های میانی شبکه)

برای ایجاد یک نقشه حرارتی<sup>۴</sup> بهره می برد. این نقشه حرارتی نشان می دهد که در لایه های کانولوشنی میانی شبکه، کدام نواحی

ویژگی ها در افزایش نمره خروجی کلاس مورد نظر سهم بیشتری داشته اند. با استفاده از این تکنیک، ما می توانیم بهتر بفهمیم که

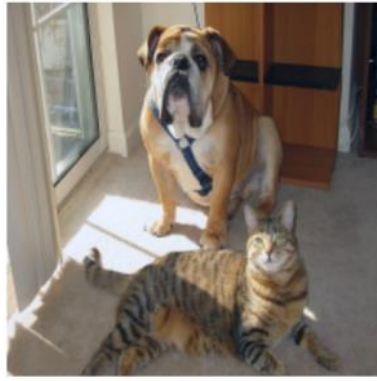
مدل به چه الگوها و بخش هایی از تصویر توجه کرده است.

<sup>1</sup> Gradient of Loss with respect to the Input

<sup>2</sup> Black Box

<sup>3</sup> Gradient-weighted Class Activation Mapping

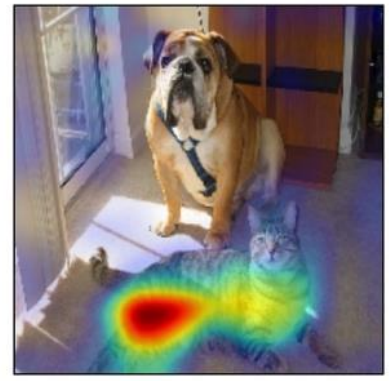
<sup>4</sup> Heatmap



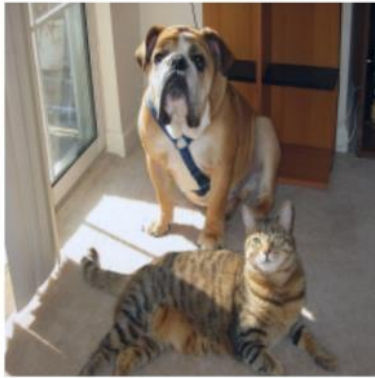
(a) Original Image



(b) Guided Backprop 'Cat'



(c) Grad-CAM 'Cat'



(g) Original Image



(h) Guided Backprop 'Dog'



(i) Grad-CAM 'Dog'

نمونه‌ای از نقشه حرارتی خروجی Grad-Cam (ستون سمت راست)

در این سوال یک شبکه عمیق پیچشی از پیش آموخته برای یک مسئله کلاس بندی در نظر بگیرید. فرض کنید  $y_c$  بیانگر نمره قبل از اعمال softmax برای کلاس  $c$  (logit for class  $c$ ) و نشانگر  $A^k \in \mathbb{R}^{H \times W}$  کانال  $k$  ام از لایه کانولوشن انتخابی است.

الف) ضریب اهمیت برای هر نقش ویژگی با استفاده از محاسبه مشتقات جزئی  $y_c^1$  نسبت به هر  $A_{ij}^k$  درایه و میانگین گیری این گرادیان هاست:

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \frac{\partial y_c}{\partial A_{ij}^k}.$$

تفسیر کنید و توضیح دهید که چرا مشتقات جزئی میانگین شده به عنوان معیاری برای اندازه گیری تاثیر نقش ویژگی  $k$ -ام در پیش‌بینی کلاس  $c$  استفاده می‌شود.

ب) با استفاده از وزن اهمیت های محاسبه شده، نقش خام Grad-CAM را به صورت ترکیب خطی نقشه های ویژگی تعریف می شود

<sup>1</sup> Partial Derivatives

$$L_{\text{Grad-CAM}}^c = \sum_k \alpha_k^c A^k.$$

نقشه نهایی را بصورت اعمال تابع فعال ساز ReLU بر نقشه خام بالا در نظر میگیریم، توضیح دهید چرا نگه داشتن بخش مثبت این ترکیب خطی ویژگی هایی را برجسته می کند که بطور مثبت بر امتیاز کلاس C تاثیر می گذارند.

## ۶۵ نمره

## سوالات عملی

### سؤال ۱ (۲۰ نمره)

از شما میخواهیم که یک بخش ساده شده از مرحله انتشار روبه جلو یک شبکه عصبی پیچشی (CNN) که شامل یک بلوک Residual است را به صورت دستی را پیاده سازی کنید. از یک ماتریس ورودی ثابت  $5 \times 5$ ، یک فیلتر پیچشی مشخص، و پارامترهای لایه های تعریف شده استفاده کنید. هدف این است که در هر مرحله، خروجی های محاسبه شده با مقادیر مورد انتظار مطابقت داشته باشند. بنابراین مقادیر ماتریس های حاصله در هر مرحله را به صورت بصری نمایش دهید.

از ماتریس ورودی ثابت زیر

$$\begin{bmatrix} 100 & 150 & 200 & 50 & 30 \\ 80 & 120 & 180 & 255 & 90 \\ 40 & 70 & 110 & 230 & 200 \\ 30 & 60 & 100 & 150 & 120 \\ 20 & 50 & 80 & 100 & 90 \end{bmatrix}$$

و فیلتر پیچشی  $3 \times 3$  با گام و پدینگ ۱ و وزن های زیر استفاده کنید.

$$\begin{bmatrix} 0.1 & -0.2 & 0.1 \\ -0.1 & 0.4 & 0.1 \\ 0.2 & -0.1 & 0.1 \end{bmatrix}$$

مراحل انجام تمرین:

- ابتدا عملیات پیچش را با استفاده از فیلتر مشخص شده انجام دهید.
- بر روی ماتریس خروجی پیچش، تابع ReLU را اعمال کنید تا مقادیر منفی به صفر تبدیل شوند.
- خروجی را با ماتریس ورودی اصلی skipconnection جمع و دوباره تابع ReLU را اعمال کنید.
- بر روی خروجی بلوک Residual یک نمونه برداری  $2 \times 2$  و گام ۲ در حالت بیشینه انجام دهید.

- در نهایت بر روی خروجی نمونه‌برداری بیشینه<sup>۱</sup>، در نهایت با استفاده از یک فیلتر `global average pooling` به یک مقدار واحد برسید.

**توجه:** در این تمرین تنها استفاده از کتابخانه‌های `numpy` و `matplotlib` مجاز است. یک نمونه از فایل اولیه نوتبوک به همراه نتایج مورد انتظار از هر مرحله به شما داده شده‌است که می‌توانید آن را کامل کنید. البته در صورت تمایل می‌توانید نوت بوک‌هایی دیگر با ساختارهای متفاوت اما نتایج یکسان تحویل دهید.

## سؤال ۲ (۴۵ نمره)

در این دفترچه، شما به بررسی و پیاده‌سازی معماری‌های *MobileNet V1* و می‌پردازید. این مدل‌ها که به صورت ویژه برای دستگاه‌های موبایل و توکار طراحی شده‌اند، معماری‌های سبک و کارآمدی هستند. وظایف شما شامل درک مفاهیمی مانند کانولوشن‌های جداشدنی عمقی<sup>۲</sup>، اتصالات معکوس باقی‌مانده<sup>۳</sup> و گلوگاه‌های خطی<sup>۴</sup> و همچنین پیاده‌سازی و آموزش این معماری‌ها با استفاده از `PyTorch` می‌باشد.

این دفترچه راهنمای مرحله‌به‌مرحله‌ای برای موارد زیر ارائه می‌دهد:

- پیاده‌سازی معماری‌های *MobileNet V1*
- آموزش و اعتبارسنجی روی دیتاست `CIFAR-10`
- مقایسه کارایی و عملکرد با شبکه‌های عصبی پیچشی ستی (CNN)
- بررسی تأثیر ابرپارامترهایی مانند ضریب عرض<sup>۵</sup> و ضریب وضوح<sup>۶</sup>

هدف شما این است که قسمت‌های ناقص دفترچه را تکمیل کنید، نتایج خود را اعتبارسنجی کنید و آن‌ها را با تحلیل‌های نظری موجود مقایسه کنید. به خروجی‌های مورد انتظار مراجعه کنید تا از صحت عملکرد اطمینان حاصل نمایید.

سالم و موفق باشید.

---

<sup>1</sup> Max Pooling

<sup>2</sup> Depth-wise Separable Convolutions

<sup>3</sup> Inverted Residual Connections

<sup>4</sup> Linear Bottlenecks

<sup>5</sup> Width Multiplier

<sup>6</sup> Resolution Multiplier