

University of Tehran

دانشکده‌گان علوم و فناوری های میان رشته ای



Neural Network

HW3

Fall 2024

توضیحات مهم

سلام بر دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره‌دهی نیست، در حد نیاز توضیح دهید.
۲. نکته‌ی مهم در گزارش‌نویسی روشن بودن پاسخ‌ها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به‌صورت واضح آن را بیان کنید.
۳. برای سوالات شبیه‌سازی، فقط از دیتاست داده‌شده استفاده کنید.
۴. فایل نهایی خود را در یک فایل زیپ شامل گزارش با فرمت PDF، آپلود کنید. نام فایل زیپ ارسالی باید الگوی زیر را داشته باشد :

NN_HW3_StudentNumber

۵. کد سوالات شبیه‌سازی بصورت فایلی تحت عنوان سوال آن با پسوند py یا ipynb به همراه گزارش در فایل زیپ تحت الگوی بند ۴ پیوست شود.
۶. هرگونه شباهت در گزارش و کد مربوط به شبیه‌سازی، به منزله‌ی تقلب می‌باشد و کل تمرین برای طرفین صرف‌نظر خواهد شد.
۷. سوالات خود را به آدرس ایمیل دستیار مربوطه به آدرس زیر ارسال نمایید:

amirhosseinsafdari4@gmail.com

amir.safdari@ut.ac.ir

سوال اول:

هر ساله شبکه هایی با ساختار جدید روی دادگان imagenet معرفی میشود که بهترین نتایج روز را بهبود می دهد. از جمله آنها شبکه های Vision Transformers (ViT), Swin Transformers, EfficientNet هستند. ساختار این شبکه ها و ویژگی های آن ها را توضیح دهید.

سوال دوم:

یکی از روش های تخمین عدم قطعیت تصمیم شبکه های عمیق استفاده از dropout در فاز تست است. روش ارائه شده در مقاله زیر را برای تخمین نایقینی تصمیم شبکه به طور کلی بیان کنید.

Dropout Injection at Test Time for Post Hoc Uncertainty Quantification in Neural Networks

سوال سوم:

نشان دهید اگر ورودی شبکه عصبی پیچشی (CNN)، یک تصویر I با ابعاد $H \times W$ باشد، آنگاه با اعمال فیلتر پیچشی (Convolution) $k \times k$ ، ابعاد تصویر خروجی $H' \times W'$ به صورت زیر بدست می آیند:

$$\begin{cases} 1 + \frac{W - k + 2p}{s} = W' \\ 1 + \frac{H - k + 2p}{s} = H' \end{cases}$$

که در آن، p نماد padding (لایه گذاری) و s نماد stride (گام) است.

سوال چهارم:

به سوالات مفهومی زیر پاسخ دهید:

الف) در یک شبکه عصبی پیچشی (CNN) چه تفاوت هایی بین لایه های پیچش (Convolutional Layers) و لایه های تماماً متصل (Fully Connected Layers) وجود دارد؟

ب) در شبکه عصبی پیچشی (CNN) با استفاده از روش پس انتشار (Back-Propagation) فرمول اصلاح وزن لایه ی کانولوشنی را بدست آورید.

ج) چگونه فیلترها یا هسته ها در لایه های پیچش (Convolutional Layers) به شناسایی ویژگی های مختلف تصویر کمک می کنند؟

د) چه معایب و محدودیت‌هایی در شبکه‌های عصبی پیچشی (CNN) وجود داشت که باعث شد مدل‌هایی مانند R-CNN و Fast R-CNN توسعه یابند؟

ه) در شبکه‌های عصبی باقی‌مانده (ResNet)، مفهوم اتصالات باقی‌مانده (Residual Connections) چیست؟ چگونه باعث بهبود آموزش شبکه‌های عمیق می‌شود؟

و) فرض کنید مجبور هستیم در یک مساله با تعداد اندک داده از شبکه عمیق استفاده کنیم. چه روش‌های regularization را برای حوزه ورودی، شبکه و خروجی پیشنهاد می‌کنید.

سوال پنجم:

استفاده از شبکه‌های عصبی باقی‌مانده ResNet، به عنوان یک مدل از پیش‌آموزش‌داده‌شده (Pre-Trained)، برای طبقه‌بندی تصاویر در دیتاست COVID-19 یک روش کارآمد است. هدف این است که از قابلیت انتقال یادگیری (Transfer Learning) استفاده کرده و لایه‌های مدل را منجمد (Freeze) کنید تا تنها لایه‌های انتهایی آموزش داده شوند.

دیتاست:

از دیتاست موجود در [COVID-19 X-Ray and CT-Scan Image Dataset](#) استفاده کنید. این دیتاست شامل تصاویر X-Ray و CT-Scan از بیماران با شرایط مختلف است که به کلاس‌های COVID-19 و Normal (یا سایر دسته‌ها) تقسیم می‌شود.

مراحل:

آماده‌سازی داده‌ها:

- دیتاست را بارگذاری کرده و تصاویر را به فرمت مورد نیاز تغییر اندازه دهید.
- ۵۰۰ نمونه کووید و ۱۰۰۰ نمونه سالم را انتخاب کنید.
- داده‌ها را به سه بخش آموزش، اعتبارسنجی، و تست تقسیم کنید (مثلاً ۷۰٪ آموزش، ۱۵٪ اعتبارسنجی، و ۱۵٪ تست).

انتخاب مدل:

- از یکی از نسخه‌های مدل ResNet (مانند ResNet-50 یا ResNet-101) استفاده کنید که روی دیتاست ImageNet از پیش آموزش دیده است.

- لایه‌های اصلی مدل را منجمد (Freeze) کنید تا تنها لایه‌های Fully Connected انتهایی برای این طبقه‌بندی تنظیم شوند. همچنین توضیح دهید که منجمد کردن مدل چه تاثیری بر فرآیند آموزش و نتایج دارد.

اضافه کردن لایه‌های جدید:

- یک یا چند لایه Fully Connected به انتهای مدل اضافه کنید. مشخصات این لایه‌ها (مانند تعداد نورون‌ها، تابع فعال‌ساز، و ...) را توضیح دهید.
- از یک تابع فعال‌ساز مناسب (مانند Softmax برای طبقه‌بندی چندکلاسه یا Sigmoid برای طبقه‌بندی دودویی) در لایه‌ی خروجی استفاده کنید.

پیکربندی مدل:

- تابع هزینه (Loss function)، بهینه‌ساز (Optimizer)، و نرخ یادگیری (Learning rate) مناسب را انتخاب و دلیل انتخاب هر یک را توضیح دهید.

آموزش مدل:

- مدل را با استفاده از داده‌های آموزش و اعتبارسنجی به مدت ۴۰ تا ۵۰ دوره (Epoch) آموزش دهید.
- نمودار Loss و Accuracy را برای داده‌های آموزش و اعتبارسنجی رسم کنید.

ارزیابی عملکرد مدل:

- مدل آموزش‌داده‌شده را با داده‌های تست ارزیابی کنید و مقادیر Accuracy، Precision، Recall، و F1-Score را گزارش کنید.
- ماتریس آشفتگی (Confusion Matrix) را برای نتایج تست رسم و تحلیل کنید.

تحلیل نتایج:

- تأثیر استفاده از یک مدل از پیش‌آموزش‌داده‌شده (ResNet) و منجمد کردن لایه‌ها را بر دقت و عملکرد مدل بررسی کنید.
- پیشنهادهایی برای بهبود عملکرد مدل ارائه دهید (مانند افزایش دوره‌های آموزش، تنظیم نرخ یادگیری، یا استفاده از داده‌های بیشتر).

- مسئله عدم توازن کلاس ها و سوگیری شبکه بررسی شود. (چگونه در بخش مجموعه داده و در بخش تابع خطای شبکه یا ... می توان مشکل سوگیری شبکه به سمت کلاس حداکثر را حل کرد؟)

گزارش نهایی باید شامل موارد زیر باشد:

- مشخصات معماری مدل (ResNet و لایه های افزوده).
- توضیحات نحوه عملکرد منجمد کردن مدل و علت انجام آن.
- تنظیمات هایپرپارامترها (مانند تابع هزینه، بهینه ساز، و نرخ یادگیری).
- نمودارهای Loss و Accuracy برای داده های آموزش و اعتبارسنجی.
- ارزیابی نهایی شامل معیارهای Accuracy، Precision، Recall، F1-Score و ماتریس آشفتگی.
- تحلیل کامل از عملکرد و نتیجه ی نهایی مدل.

سوال ششم: (امتیازی)

تشخیص اشیاء، یک روش مبتنی بر بینایی کامپیوتر در حوزه هوش مصنوعی است که برای تعیین نوع و مکان اشیا در تصاویر یا فیلم ها مورد استفاده قرار می گیرد. از جمله چالش های مهم در این حوزه، حضور چند شیء مختلف در تصویر و ضرورت تشخیص مکان دقیق آن ها است.

در این سوال بایستی به بررسی تشخیص عیوب جوش مبتنی بر شبکه های پیچشی (CNNs) بپردازید و از دو مدل معروف در این حوزه به نام های YOLOv3 و Faster-RCNN استفاده کنید.

دیتاست:

از دیتاست موجود در پیوست استفاده کنید. این دیتاست شامل تصاویر X-Ray از جوشکاری با شرایط مختلف است که بایستی عیوب جوشکاری بر روی آنها مشخص گردد.

گزارش نهایی باید شامل موارد زیر باشد:

- مشخصات معماری مدل (YOLOv3 و Faster-RCNN).
- تنظیمات هایپرپارامترها (مانند تابع هزینه، بهینه ساز، و نرخ یادگیری).
- نمودارهای Loss و Accuracy برای داده های آموزش و اعتبارسنجی.
- ارزیابی نهایی شامل معیارهای Accuracy، Precision، Recall، F1-Score و ماتریس آشفتگی.
- تحلیل کامل از عملکرد و نتیجه ی نهایی مدل.