



به نام خدا
درس مبانی یادگیری عمیق
تمرین سری چهارم
استاد درس : دکتر مرضیه داوودآبادی
دستیاران : الناز رضایی، سید محمد موسوی، کمیل فتحی
دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر
نیمسال اول تحصیلی ۱۴۰۲ - ۱۴۰۳

مهلت تحویل : ۱۴۰۲/۰۹/۱۷

لطفا به نکات موجود در سند قوانین انجام و تحویل تمرین ها دقت فرمایید.

۱. در این سؤال قصد داریم تا مدلی طراحی کنیم که توانایی دسته‌بندی داده‌های موجود در دیتاست *MNIST* و پیدا کردن کمترین میزان خطا را به صورت خودکار داشته باشد. در صورت پیدا کردن بهترین مقدار برای هر *Hyper parameter*، مدل بهینه و مناسبی ساخته خواهد شد که توانایی دسته‌بندی داده‌ها با کمترین میزان خطا را دارد. حال گام‌های الف تا ج را طی کنید تا شبکه مناسب را طراحی نمایید (۱۵ نمره).
الف) طراحی شبکه هم‌گشتی^۱ با استفاده از *KerasTuner* :

- ابزار *KerasTuner* را معرفی کنید.
- توضیح دهید چگونه می‌توان از این ابزار برای بهینه‌سازی شبکه هم‌گشتی جهت دسته‌بندی داده‌ها استفاده کرد.
- *Tuner* موجود در *KerasTuner* را معرفی کنید. شما در پیاده‌سازی خود از کدام یک استفاده می‌نمایید؟ چرا؟

ب) استفاده از *KerasTuner* بر روی دیتاست *MNIST* :

- دیتاست *MNIST* را معرفی کنید.
- توضیح دهید چگونه می‌توان از شبکه هم‌گشتی بهینه‌سازی شده با *KerasTuner* برای دسته‌بندی تصاویر این دیتاست استفاده کرد.

^۱Convolutional

- اهمیت و تأثیر استفاده از *dropout* و *pooling* در بهبود عملکرد شبکه را توضیح دهید (۵ نمره امتیازی).

(ج) پیاده‌سازی شبکه موردنظر:

- باتوجه به توضیحات داده شده، مدل پیشنهادی خود را که باید شامل تعدادی لایه *Convolutional* و *Fully – connected* باشد را پیاده‌سازی کنید. در صورت نیاز می‌توان از لایه *pooling* و *dropout* نیز استفاده نمود. در جدول زیر بخشی از اطلاعات موردنیاز برای *tune* مدل ارائه شده است همچنین باتوجه به دانش خود مدل را تا حد امکان بهبود دهید. اما موارد زیر را حتماً با استفاده از *KerasTuner*، تنظیم و نتایج را گزارش و تحلیل نمایید.

<i>HyperParameter</i>	<i>MaxLayer/value</i>
<i>Convolutional</i>	۵
<i>Dense</i>	۵
<i>Filters</i>	۲۵۶
<i>Neurons</i>	۲۵۶
<i>Learningrate</i>	۰.۰۰۱
<i>Optimizer</i>	–

- نحوه انتخاب اندازه فیلترها در لایه‌های *Convolutional* را توضیح دهید.
- چگونه می‌توان از *pooling* و *dropout* در شبکه هم گشتی برای جلوگیری از *over fitting* و افزایش دقت استفاده کرد؟ (۵ نمره امتیازی)

۲. به فایل *medical.ipynb* مراجعه کرده و باتوجه به توضیحات قسمتهای مربوطه را تکمیل کنید (۲۰ نمره).

۳. در فایل *consumption.ipynb* یک کد جهت پیشبینی مصرف گاز قرار داده شده است که ورودی ها و خروجی های آن به شکل زیر است:

ورودی ها: تاریخ، ساعت، ماه میلادی و دما
خروجی: مصرف گاز

با وجود استفاده از *lstm* دقت r^2 منفی به دست می آید. علت و مشکل کد را پیدا کرده و پیشنهادی برای اصلاح آن بیان کنید (۱۰ نمره).

۴. الف) تعریفی مختصر از شبکه های هم گشتی و شبکه های بازگشتی بیان کرده و بگویید هریک برای چه کاربردها و مسائلی بهتر عمل می کند؟ چرا؟ (۵ نمره)

ب) این شبکه ها را از لحاظ تعداد پارامتر و قابلیت موازی سازی با یکدیگر مقایسه کنید (۵ نمره).

۵. الف) شبکه هم گشتی زیر را در نظر بگیرید. فرض کنید ورودی یک تصویر رنگی با اندازه ۲۵۶ در ۲۵۶ است. اندازه خروجی و تعداد پارامترها را برای هر لایه به دست آورید (۵ نمره).

Layer1: Conv(64, (3,3), stride=1, padding='same')

Layer2: Dilated-Conv(32, (5,5), stride=2, dilation rate=2, padding='valid')

Layer3: Max-pool (size=(2,2), stride=2)

Layer4: Conv(128, (3,3), stride=1, padding='same')

Layer5: Dilated-Conv(64, (5,5), stride=2, dilation rate=4, padding='valid')

Layer6: Max-pool (size=(2,2), stride=2)

Layer7: Conv(256, (3,3), stride=1, padding='same')

Layer8: Dilated-Conv(128, (5,5), stride=2, dilation rate=8, padding='valid')

Layer9: Max-pool (size=(2,2), stride=2)

ب) در $2Dconv$ چه میزان $padding$ باید اضافه شود تا ابعاد خروجی تغییر نکند؟ (پاسخ را به صورت رابطه ای از سائز فیلتر یعنی f بیان کنید) (۵ نمره).

۶. به سوالات زیر پاسخ دهید (۱۵ نمره).

الف) کدام یک از عبارات زیر در مورد نرمال سازی دسته ای^۲ درست است؟

- نرمال سازی دسته ای تنها پردازش یک دسته را سریع تر می کند و زمان آموزش را کاهش می دهد و درعین حال تعداد به روزرسانی ها را ثابت نگه می دارد. این به شبکه اجازه می دهد تا زمان مشابهی را صرف انجام به روزرسانی های بیشتر کند تا به حداقل برسد.
- نرمال سازی دسته ای توزیع خروجی را نرمال می کند تا در ابعاد یکنواخت تر باشد.
- به شبکه اجازه می دهد تا وزن های ما را به مقادیر کوچک تر نزدیک به صفر مقداردهی کند.

ب) کد موجود در فایل `forward - batchnorm.py` را به کمک `numpy` تکمیل کنید.

ج) نقش ابر پارامتر ϵ در نرمال سازی دسته ای چیست؟ (۵ نمره امتیازی)

د) استفاده از نرمال سازی دسته ای با اندازه یک چه مشکل یا مشکلاتی دارد؟

²Batch Normalization

ه) یک شبکه کاملاً متصل با ۱۰ ورودی و ۲۰ خروجی را در نظر بگیرید که نرمال سازی دسته ای به آن اضافه می کنیم. تعداد پارامترهای قابل آموزش را محاسبه کنید.

۷. در این سوال قصد داریم الگوریتم $Grad - CAM$ را بر روی مجموعه داده $MNIST$ پیاده سازی نماییم. برای این کار مراحل زیر را انجام دهید (۲۰ نمره).

- ابتدا به کمک api فراهم شده در کتابخانه $Keras$ مجموعه داده $MNIST$ را بارگذاری نمایید.
- داده های آموزشی را $shuffle$ کنید و ابعاد داده ها را چاپ کنید.
- ۱۰ تصویر نخست موجود در مجموعه داده آموزشی را به همراه برجسب آن ها نمایش دهید.
- مقادیر پیکسل ها را به محدوده ۰ تا ۱ تغییر دهید.
- برجسب ها را به نمایش $categorical$ تغییر دهید.
- یک شبکه عصبی با ویژگی های زیر طراحی و آموزش دهید:

□ لایه $Conv2D$ با ۳۲ فیلتر (۳، ۳) و تابع فعال سازی $ReLU$ با پارامتر $padding$ برابر $same$.

□ لایه $MaxPooling2D$ با ابعاد پنجره (۲، ۲).

□ لایه $Conv2D$ با ۶۴ فیلتر (۳، ۳) و تابع فعال سازی $ReLU$ با پارامتر $padding$ برابر $same$.

□ لایه $MaxPooling2D$ با ابعاد پنجره (۲، ۲).

□ لایه $Conv2D$ با ۶۴ فیلتر (۳، ۳) و تابع فعال سازی $ReLU$ با پارامتر $padding$ برابر $same$.

□ لایه $MaxPooling2D$ با ابعاد پنجره (۲، ۲).

□ لایه $Flatten$.

□ لایه $Dense$ با ۱۲۸ واحد و تابع $ReLU$.

□ لایه $Dense$ با ۱۰ واحد و تابع $Softmax$.

- مدل ساخته شده را با مشخصات زیر آموزش دهید:

□ تابع ضرر: $CategoricalCrossentropy$

□ بهینه ساز: $Adam$

□ تعداد $Epoch$: ۱۵

□ اندازه $Batch$: ۶۴

- الگوریتم $Grad - CAM$ را بر روی آخرین لایه هم گشتی اجرا کرده و خروجی آن را برای ۱۰ تصویر نمونه نمایش دهید.
- خروجی الگوریتم را تحلیل کنید و نتیجه‌ی استفاده از $Grad - CAM$ در تشخیص ویژگی‌های مهم در تصاویر $MNIST$ را بررسی کنید.