



تمرین سری ششم درس مباحث ویژه

نام مدرس: دکتر محمدی

دستیار آموزشی مرتبط: نیکوکاران، باباپور

مهلت تحویل: ۱۵ آبان ۱۴۰۰

۱- در این سوال قصد داریم با دو نمونه از توابع فعال سازی که اخیراً معرفی شده‌اند، یعنی [Swish](#) و [Mish](#) آشنا شویم. نسخه تغییر یافته مقاله این دو تابع فعال سازی که نکات مهم در آن مشخص شده‌اند در فوق ارجاع داده شده‌اند. با استفاده از این دو مقاله به سه دسته از سوالات زیر پاسخ دهید.

سوالات عمومی:

أ. رابطه هر یک از این دو تابع فعال سازی را ذکر کنید و به کمک کتابخانه‌های [NumPy](#) و [Matplotlib](#) نمودار این توابع را رسم نمایید.

برای اطمینان از درستی نمودارهای رسم شده به نمودارهای ارائه شده در مقالات توجه فرمایید.

ب. مشتق این توابع فعال سازی را محاسبه کنید و با کمک کتابخانه‌های فوق، این توابع را رسم نمایید. درستی این نمودارها نیز از طریق نمودارهای موجود در مقالات قابل بررسی است.

ت. ابتدا بگویید تابع فعال سازی ReLU چه مزیت‌هایی نسبت به دو تابع فعال سازی که تا آن زمان استفاده می‌شده‌اند، یعنی Sigmoid و

Tanh داشته است و سپس مشخص کنید معرفی دو تابع فعال سازی اخیر یعنی [Swish](#) و [Mish](#) کدام مشکلات تابع فعال سازی ReLU

را برطرف کرده‌اند و چه شباهت‌هایی با این تابع فعال سازی دارند. در این بخش می‌توانید از مفاهیم Dying ReLU، حدود هر یک از

توابع فوق و مشکل مقداردهی اولیه وزن‌ها کمک بگیرید.

سوالات اختصاصی تابع فعال سازی [Swish](#):

ث. نویسنده در بخشی از این مقاله به معرفی تابع فعال سازی $\text{Swish-}\beta$ می‌پردازد. بیان کنید این تابع فعال سازی چه مزیت‌هایی دارد.

سوالات اختصاصی تابع فعال سازی [Mish](#):

ج. دلایلی که نویسنده تابع فعال سازی [Mish](#) را بهتر از تابع فعال سازی [Swish](#) می‌داند را ذکر کنید. در این بخش می‌توانید از پارامتر $\Delta(x)$

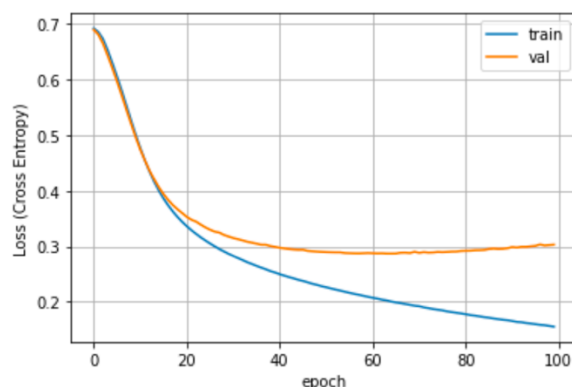
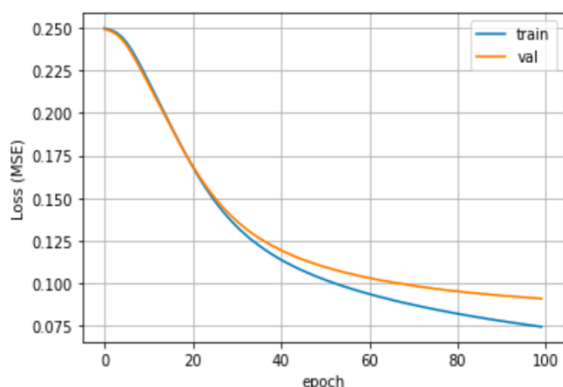
موجود در رابطه (۲) و همچنین تصویر ۴ در مقاله [Mish](#) استفاده نمایید.

۲- دو نمودار زیر، روند آموزش دو مدل در یک مسئله که دارای دو کلاس می‌باشد را بر روی دو نوع داده آموزشی (رنگ آبی) و داده آزمایشی (رنگ

نارنجی) نشان می‌دهد. نمودار سمت چپ روند آموزش مدلی را نشان می‌دهد که از تابع ضرر MSE و نمودار سمت راست روند آموزش مدلی را

نشان می‌دهد که از تابع ضرر Binary Cross Entropy استفاده می‌کند. لازم به ذکر است در این مدل از بهینه‌ساز SGD برای آموزش مدل‌ها

استفاده شده است. با توجه به این دو نمودار به سوالات زیر پاسخ دهید.



أ. به نظر شما چرا مقدار تابع ضرر در Epoch نخست برای این دو مدل نزدیک به ۰.۲۵ و ۰.۷ شده است؟

ب. توضیح دهید چرا در نمودار مرتبط با مدلی که از تابع ضرر Binary Cross Entropy استفاده می‌کند، اختلاف دو نمودار مربوط به داده‌های آموزشی و آزمایشی بیشتر از نمودار دیگر است.

ت. با فرض آنکه قابلیت توقف آموزش در هر Epoch را در اختیار داشتیم، چه زمانی را برای توقف آموزش این دو مدل بهتر می‌دانید.

۳- همانطور که در کلاس گفته شد، تابع فعال‌سازی ReLU انواع مختلفی دارد که از معروف‌ترین آن‌ها می‌توان به Parametric Leaky ReLU، ELU، ReLU اشاره کرد. در این سوال قصد داریم آزمایشی در رابطه با مقادیر متفاوت پارامتر α در تابع فعال‌سازی Leaky ReLU داشته باشیم. در این سوال با اندکی تغییر در مجموعه داده [MNIST](#) نسخه بهبود یافته این مجموعه داده را آماده می‌کنیم و سپس نشان می‌دهیم تنظیم مقدار α چه تاثیری در آموزش مدل دارد. برای انجام آزمایش مراحل زیر را انجام دهید:

- ابتدا به کمک api فراهم شده در کتابخانه [Keras](#) مجموعه داده [MNIST](#) را بارگذاری نمایید.
- مجموعه داده [MNIST](#) را به گونه‌ای تغییر دهید که برای تصویر هر عدد در مجموعه داده علاوه بر تصویر اصلی، تصویری با دست‌خط مشابه و برچسب مشابه اما پس زمینه سفید و رنگ خط مشکی نیز وجود داشته باشد. برای این کار تنها لازم است به ازای هر تصویر مقدار هر پیکسل از عدد ۲۵۵ کم شود. در این صورت تعداد داده‌های آموزشی از ۶۰ هزار به ۱۲۰ هزار و تعداد داده‌های آزمایشی از ۱۰ هزار به ۲۰ هزار افزایش می‌یابد.
- داده‌های آموزشی را با استفاده از تابع [shuffle](#) پیاده‌سازی شده در کتابخانه [scikit-learn](#) بهم بریزید.
- برای اطمینان از درستی پیاده‌سازی مراحل فوق، ابعاد داده‌های آموزشی و داده‌های آزمایشی را چاپ نمایید. همچنین ۱۰ تصویر نخست موجود در مجموعه داده آموزشی را به همراه برچسب آن‌ها نمایش دهید. توجه فرمایید که در این تصاویر باید تصاویری با پس‌زمینه سفید و مشکی به صورت تصادفی وجود داشته باشند.



تمرین سری ششم درس مباحث ویژه

نام مدرس: دکتر محمدی

دستیار آموزشی مرتبط: نیکوکاران، باباپور

مهلت تحویل: ۱۵ آبان ۱۴۰۰

- همانطور که در قسمت قبل مشاهده کردید، ابعاد داده‌های آموزشی و داده‌های آزمایشی به ترتیب (20000, 28, 28) و (120000, 28, 28) می‌باشد. برای استفاده از لایه‌های کانولوشنی بایستی ابعاد داده‌های آموزشی و آزمایشی به ترتیب به (20000, 28, 28, 1) و (120000, 28, 28, 1) تغییر نماید. با استفاده از تابع [reshape](#) این تغییر را ایجاد نمایید.
- مقدار هر پیکسل در هر تصویر این مجموعه داده بین صفر تا ۲۵۵ می‌باشد. تبدیل این محدود به محدوده صفر تا یک می‌تواند به روند آموزش کمک کند، لذا این تغییر را نیز به مجموعه داده اضافه نمایید. همچنین برچسب‌های این مجموعه داده دارای مقادیری بین صفر تا ۹ می‌باشند. با تغییر این نمایش به نمایش categorical به آموزش مدل کمک بیشتری می‌کنیم. برای این کار از این [لینک](#) استفاده کنید.
- در این بخش قصد داریم با آزمایش مقادیر مختلف پارامتر α برای تابع فعال‌سازی Leaky ReLU، بهترین مقدار این پارامتر را برای مسئله فوق بیابیم، لذا تابعی پیاده‌سازی نمایید که پارامتر α تابع فعال‌سازی را به عنوان ورودی دریافت نماید و مدل را به عنوان خروجی تحویل نماید. این مدل را از نوع Sequential ساخته و سپس از لایه‌هایی با مشخصات زیر استفاده نمایید:
 - لایه Conv2D با ۸ فیلتر و ابعاد پنجره (7, 7) و تابع فعال‌سازی Leaky ReLU. توجه شود که پارامتر ورودی α ورودی تابع به این تابع فعال‌سازی داده شود.
 - لایه Conv2D با ۸ فیلتر و ابعاد پنجره (5, 5) و تابع فعال‌سازی Leaky ReLU. توجه شود که پارامتر ورودی α ورودی تابع به این تابع فعال‌سازی داده شود.
 - Flatten
 - لایه Dense با ۱۰ واحد و تابع فعال‌سازی softmax.
- برای اطمینان از درستی مدل ساخته شده، با استفاده از تابع پیاده‌سازی شده، مدلی با پارامتر α دلخواه ساخته و خلاصه‌ای (summary) از مدل نمایش دهید. این خلاصه باید مشابه تصویر زیر باشد.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 22, 22, 8)	400
conv2d_1 (Conv2D)	(None, 18, 18, 8)	1608
flatten (Flatten)	(None, 2592)	0
dense (Dense)	(None, 10)	25930
Total params: 27,938		
Trainable params: 27,938		
Non-trainable params: 0		

- پارامتر α را برابر با ۱، -۰.۵، صفر، ۰.۵ و ۱ قرار دهید و برای هر یک از این مقادیر مدل را مطابق شرایط زیر آموزش دهید:



تمرین سری ششم درس مباحث ویژه

نام مدرس: دکتر محمدی

دستیار آموزشی مرتبط: نیکوکاران، باباپور

مهلت تحویل: ۱۵ آبان ۱۴۰۰

○ تابع ضرر: Categorical Crossentropy

○ بهینه‌سازی: Adam

○ تعداد Epoch: ۵

○ اندازه Batch: ۱۲۸

- نمودار دقت و مقدار تابع ضرر را برای هر یک از این مقادیر نمایش دهید. آیا مورد عجیبی در رابطه با یکی از این مقادیر مشاهده نمی‌کنید؟ توضیح دهید. همچنین مشخص کنید کدام مقدار بهترین عملکرد را دارد. توضیح دهید.
 - مدلی مشابه با مدل قسمت‌های پیشین پیاده‌سازی کنید اما این بار تابع فعال‌سازی Leaky ReLU موجود در دو لایه نخست را با تابع فعال‌سازی [PReLU](#) جایگزین نمایید و نمودارهای دقت و مقدار تابع ضرر را رسم نمایید. همچنین مقدار α هر فیلتر لایه‌های کانولوشنی را چاپ نمایید. برای این کار ابتدا به توضیحات داده شده در بخش انتهایی لینک فوق توجه فرمایید. مطابق این توضیحات توجه فرمایید که حتما پارامتر `shared_axes` را در این دو تابع فعال‌سازی برابر با `[1, 2]` قرار دهید. سپس برای چاپ مقادیر α هر فیلتر لایه‌های کانولوشنی، ابتدا بوسیله `model.layers` به تمامی لایه‌های مدل و سپس به کمک تابع `layer.get_weights()` به وزن‌های هر لایه دسترسی پیدا کنید. عنصر اول خروجی تابع `get_weights` وزن‌های نورون، عنصر دوم مقادیر `bias` و عنصر سوم در صورت وجود برابر با مقادیر α می‌باشد. آیا این مقادیر نتیجه آزمایش انجام شده بر روی مقادیر متفاوت پارامتر α در تابع فعال‌سازی Leaky ReLU را تایید می‌نمایند؟ توضیح دهید. (توضیح این بخش شامل نمره امتیازی می‌باشد)
- ** توصیه می‌شود از امکانات [TensorBoard](#) برای رسم نمودارها استفاده شود.

نکات تکمیلی:

۱. لطفاً پاسخ سوالات (تئوری و توضیحات پیاده‌سازی) را به طور گویا و به زبان فارسی و در صورت امکان تایپ همراه با سورس کدهای نوشته شده، فقط در یک فایل فشرده شده به شکل `HW6_YourStudentID.zip` قرار داده و بارگذاری نمایید.
۲. منابع استفاده شده را به طور دقیق ذکر کنید.
۳. برای سهولت در پیاده‌سازی‌ها و منابع بیشتر، زبان پایتون پیشنهاد می‌شود. لطفاً کدهای مربوطه را در فرمت `ipynb` ارسال نمایید و هر کدام از موارد خواسته شده در بالا را در یک سلول جدید پیاده‌سازی نمایید.
۴. ارزیابی تمرین‌ها براساس صحیح بودن راه حل‌ها، گزارش مناسب، بهینه بودن کدها و کپی نبودن می‌باشد.
۵. در مجموع تمام تمرین‌ها، تنها ۷۲ ساعت تاخیر در ارسال پاسخ‌ها مجاز است اما پس از آن به صورت خطی از نمره شما کسر خواهد شد (معادل با روزی ۵۰ درصد).



تمرین سری ششم درس مباحث ویژه

نام مدرس: دکتر محمدی

دستیار آموزشی مرتبط: نیکوکاران، باباپور

مهلت تحویل: ۱۵ آبان ۱۴۰۰

۶. تمرین‌ها باید به صورت انفرادی انجام شوند و حل گروهی تمرین مجاز نیست.

۷. پرسش و پاسخ در رابطه با تمرین‌ها را می‌توانید در گروه مربوطه مطرح کنید.

موفق و سربلند باشید