به نام خدا

رباتیک اجتماعی و شناختی

مدرس : علیرضا طاهری

نیمسال دوم ۰۲-۱۴۰۲

دانشگاه صنعتی شریف
دانشکده‌ی مهندسی مکانیک

تمرین سری دوم	" شبکه‌های عصبی پیچشی"	مهلت ارسال: ۴ اردیبهشت ۱۴۰۳

- هم فکری و هم کاری شما در انجام تمارین مانعی ندارد اما پاسخ ارسالی هر فرد در نهایت حتما باید توسط خود او حل و نوشته شده باشد.
- در صورت هم فکری و یا استفاده از منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید. کمک گرفتن از LLM ها در حل تمارین مجاز است!!
- نتایج و پاسخ های خود را در یک فایل فشرده (ترجیحاً به نام HW1-Name-StudentNumber) در سامانه قرار دهید. پیشنهاد می شود برای بخش نرم افزاری از زبان برنامه نویسی پایتون در یکی از محیط های Jupyter notebook و یا Google Colab برای کدنویسی و تست کدهای خود استفاده نموده و فایل کدها را به فرم IPYNB ارسال کنید.

سوالات تشریحی (۲۵۰ نمره)

۱- با دانش خود و یا بهره گیری از جستجوهای اینترنتی یا سایر منابع، به پرسش‌های زیر پاسخ دهید:

الف) برای بدست آوردن loss تاکنون، روش هایی همانند MSE را آموختیم. در خصوص روش های دیگری مانند nRMSE و rRMSE تحقیق کنید و ویژگی های این روش ها را بیان کنید. این روش ها را با روش MSE که پیش تر در کلاس آموختید، مقایسه کنید.

ب) تحقیق کنید که توابع فعال سازی LeakyReLU، PReLU، RReLU، ELU، SELU و Gelu چه ویژگی هایی دارند. این توابع را با ReLU که تاکنون آموخته اید، مقایسه کنید.

پ) در شبکه‌های عصبی گاهی از دراپ‌اوت[1] دوبعدی و سه‌بعدی استفاده می‌شود. کاربرد و مزایای این دو روش را توضیح داده و بگویید چه تفاوتی با دراپ‌اوت معمولی دارند.

ت) همانطور که می‌دانید در شبکه‌های عصبی عمیق بعضا از روشی به نام نرمال‌سازی دسته‌ای[2] استفاده می‌شود. تحقیق کنید نرمال‌سازی دسته‌ای چگونه به آموزش شبکه‌های عصبی کمک می‌کند. هم‌چنین به طور مختصر توضیح دهید چه پارامترهای قابل یادگیری در روش نرمال‌سازی دسته‌ای وجود دارد.

[1] Dropout

[2] Batch Normalization

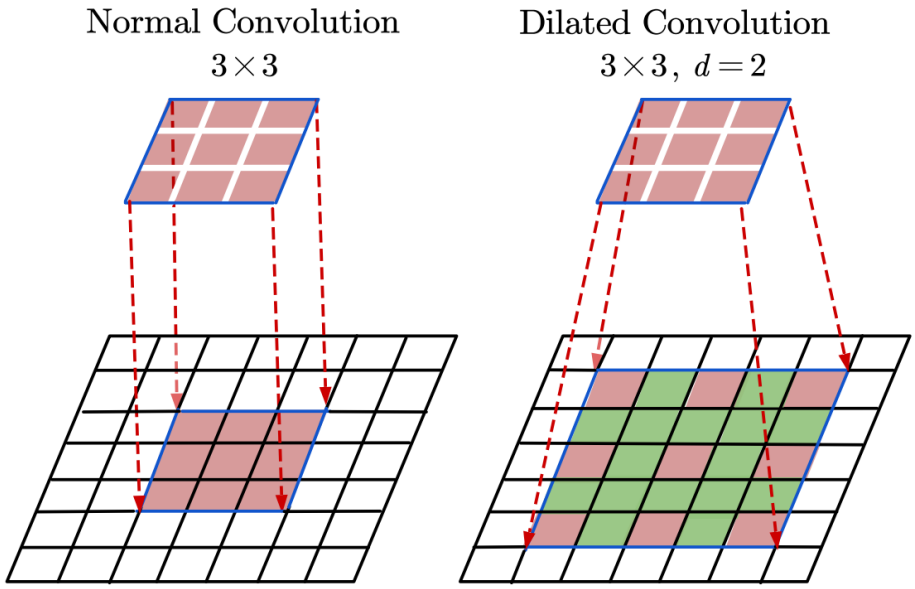
ث) در خصوص شبکه‌های خودرمزگذار[3] با لایه‌های کانولوشنی تحقیق کنید. ساختار، کاربرد و ویژگی‌های این شبکه‌ها را بیان نمایید.

۲- همانطور که می‌دانید، در شبکه‌های پیچشی معمولاً از لایه‌های کانولوشن ساده استفاده می‌شود. نوع دیگری از لایه‌های کانولوشن نیز وجود دارد که لایه‌های کانولوشن گسترش‌یافته[4] نام دارند. در این لایه‌ها، همانطور که در شکل ۱ مشاهده می‌شود، فیلتر هنگام اعمال بر ورودی و انجام ضرب کانولوشنی، بین خانه‌های ورودی فاصله می‌اندازد و با طول گام[5] بزرگ‌تری حرکت می‌کند. توجه کنید که این طول گام با مفهوم stride در شبکه‌های پیچشی تفاوت دارد.

الف) در مورد مزایا و معایب لایه‌های کانولوشن گسترش‌یافته تحقیق کنید و حداقل دو مورد از هر کدام را بیان کنید.

ب) در لایه‌های کانولوشن گسترش‌یافته، مفهوم میدان دید[6] اهمیت بیشتری پیدا می‌کند. فرض کنید سه لایه کانولوشن گسترش‌یافته با سایز فیلتر $k \times k$ و طول گام d بر روی یک ورودی اعمال می‌شود. محدوده ورودی را که عنصر i, j ام خروجی مشاهده می‌کند، به صورت پارامتری مشخص کنید (مقدار stride را ۱ در نظر بگیرید).

پ) فرض کنید یک لایه با m فیلتر با اندازه $k \times k \times 3$ با پارامتر گسترش d به یک ورودی با اندازه $N \times N \times 3$ اعمال شود. ابعاد خروجی این لایه را بر حسب N, m, k, d محاسبه کنید.



شکل ۱: فیلتر کانولوشن معمولی و کانولوشن گسترش‌یافته

---

[3] Auto Encoder

[4] Dilated Convolution

[5] Dilation Rate

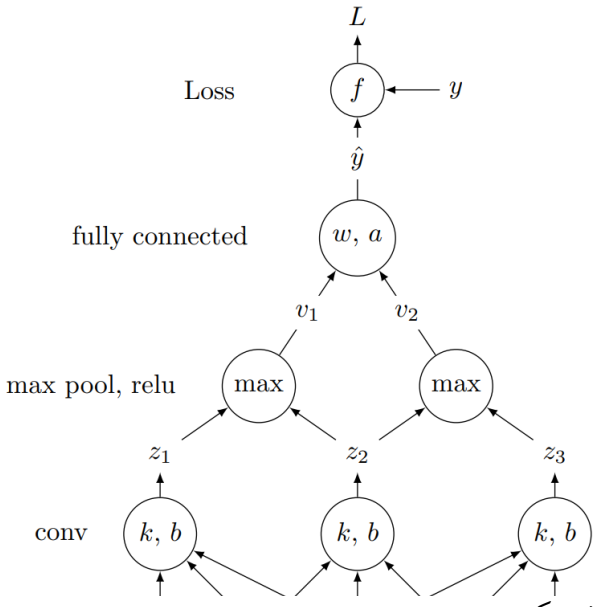
[6] Receptive Field

۳- شبکه کانولوشنی با ساختار ذکر شده در جدول زیر را در نظر بگیرید. برای هر کدام از لایه‌های موجود در جدول، ابعاد خروجی هر لایه و تعداد پارامترهای آن را محاسبه کنید. ابعاد خروجی هر لایه به صورت $H \times W \times C$ بیان می‌شود که حروف به ترتیب بیانگر ارتفاع، عرض و عمق خروجی می‌باشد. نحوه نمایش لایه‌ها به صورت زیر است:

- CONV*x-N(S, P)* بیانگر یک لایه کانولوشن است که دارای N فیلتر با ابعاد $x \times x \times D$ می‌باشد، که D عمق خروجی لایه قبل است. همچنین S بیانگر stride و P بیانگر padding می‌باشد. در صورت عدم بیان این دو مقدار، هر دو را یک در نظر بگیرید.
- POOL-*n* بیانگر یک لایه max-pooling با اندازه $n \times n$ و پدینگ ۰ است.
- FC-*N* بیانگر یک لایه fully-connected با N نورون است.

Layer	Output dimension	Number of parameters
Input	32 × 32 × 3	
CONV3-8		
ReLU		
POOL-2		
BATCHNORM		
CONV5-16(3, 2)		
ReLU		
POOL-2		
FLATTEN		
FC-10		

۴- شبکه کانولوشنی یک بعدی زیر را در نظر بگیرید. در شرایطی که همه متغیرها اسکالر هستند، به سوالات زیر پاسخ دهید.



$$L = \frac{1}{2}(y - \hat{y})^2$$

$$\hat{y} = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + a$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \max\{z_1, z_2, 0\} \\ \max\{z_2, z_3, 0\} \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 & 0 & 0 \\ 0 & k_1 & k_2 & k_3 & 0 \\ 0 & 0 & k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix}$$

شکل ۲: شبکه کانولوشن یک بعدی

الف) پارامترهای این شبکه را مشخص کنید.

ب) مقادیر $\frac{\partial L}{\partial w_1}$ و $\frac{\partial L}{\partial w_2}$ و $\frac{\partial L}{\partial a}$ را بدست آورید.

پ) با فرض اینکه داشته باشیم:

$$\frac{\partial L}{\partial v_1} = \delta_1 \qquad\qquad \frac{\partial L}{\partial v_1} = \delta_2$$

مقادیر $\frac{\partial L}{\partial z_i}$ را محاسبه کنید.

ت) با فرض اینکه داشته باشیم:

$$\frac{\partial L}{\partial z_1} = \alpha_1 \qquad \frac{\partial L}{\partial z_2} = \alpha_2 \qquad \frac{\partial L}{\partial z_3} = \alpha_3$$

مقادیر $\frac{\partial L}{\partial k_i}$ و $\frac{\partial L}{\partial b}$ را بدست آورید.

ث) در حالت کلی لایه کانولوشن یک بعدی زیر را در نظر بگیرید:

$$\begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} k_1 & \cdots & k_d & & \\ & k_1 & \cdots & k_d & \\ & & & \ddots & \\ & & & k_1 & \cdots & k_d \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b \\ \vdots \\ b \end{bmatrix}$$

با فرض اینکه داشته باشیم:

$$\frac{\partial L}{\partial z_i} = \alpha_i$$

مقادیر $\frac{\partial L}{\partial k_i}$ و $\frac{\partial L}{\partial b}$ را بدست آورید.

سوالات عملی (۷۵۰ نمره)

۱- هدف از این تمرین مقایسه عملکرد شبکه‌های MLP و CNN در یک مسئله طبقه‌بندی مشابه می‌باشد. در ابتدا قصد داریم با استفاده از مجموعه دادگان EMNIST و با ساختار مبتنی بر MLP، یک بار طبقه‌بندی داده‌ها را انجام دهیم. سپس آموزش را با استفاده از شبکه‌های پیچشی و با تعداد پارامترهای تقریباً مشابه تکرار می‌کنیم، تا با مزایا و معایب هر کدام از این شبکه‌ها آشنا شویم (توجه شود که از طبقه‌بندی byclass، که شامل کل داده‌های موجود در این دیتاست است، استفاده شود. این طبقه‌بندی ۶۲ کلاس دارد که شامل اعداد ۰ تا ۹ و حروف کوچک و بزرگ انگلیسی است).

الف) برای آشنایی با دادگان، یک تصویر از هر کلاس را به عنوان نمونه نمایش دهید.

ب) یک شبکه MLP را طراحی کرده و آموزش دهید. برای جلوگیری از بیش‌برازش[7] از تکنیک‌های Regularization مانند Dropout و L2 Regularization استفاده کنید. توجه کنید که در طول آموزش بهترین مدل را ذخیره کنید. سعی کنید تعداد پارامترهای شبکه با تعداد پارامترهای شبکه پیشنهادی شما در قسمت CNN تقریباً مشابه باشد. تعداد پارامترهای شبکه و نیز Accuracy، Recall و Precision را گزارش نمایید.

پ) تنسوربورد[8] ابزاری است که به ما امکان مشاهده چگونگی تغییرات تابع خسارت و یا چگونگی تغییر وزن‌ها را در طول زمان می‌دهد. از این ابزار استفاده کنید و نمودار دقت و تابع خسارت را برای دادگان آموزش و تست رسم کنید.

ت) یک شبکه پیچشی با معماری دلخواه را طراحی کرده و آموزش دهید. مانند بخش قبل می‌توانید از تکنیک‌هایی نظیر Dropout و Batch Normalization استفاده کنید. هایپرپارامترها را مشابه بخش قبل در نظر بگیرید. در طول آموزش بهترین مدل خود را ذخیره کنید. تعداد پارامترهای شبکه و نیز Accuracy، Recall و Precision را گزارش نمایید و با استفاده از تنسوربورد نمودارهای دقت و تابع خسارت را برای دادگان آموزش و تست رسم کنید.

ث) نتایج به دست آمده را با نتایج بخش قبل مقایسه کنید.

ج) فاکتوریزیشن کرنل‌ها روشی است که در آن بجای استفاده از یک کرنل با سایز بزرگ‌تر از چند کرنل متوالی با سایز کوچک‌تر استفاده می‌شود (برای مثال در این روش یک فیلتر ۳ × ۳ به دو فیلتر ۱ × ۳ و ۳ × ۱ متوالی تبدیل می‌شود). معماری شبکه پیچشی خود را به این منظور به روز کنید و شبکه را مجدداً آموزش دهید. تعداد پارامترها را با بخش قبل مقایسه کنید. بطور کلی مزایای استفاده از این روش را بیان کنید.

۲- هدف از این تمرین انجام مسئله طبقه‌بندی برای تشخیص حالت چهره افراد (نظیر شادی، غم، خشم و ...) با استفاده از دادگان FER-2013 می‌باشد.

الف) یک شبکه عصبی پیچشی با معماری دلخواه طراحی کنید و آن را با دادگان FER-2013 آموزش دهید، بطوری که تا حد امکان طبقه‌بندی ۷ کلاسه شناسایی حالت چهره افراد را به خوبی انجام دهد. عکس‌ها در دو پوشه آموزش و تست قرار دارند و نام هر پوشه معادل برچسب آن کلاس می‌باشد.

- Accuracy، Recall و Precision را گزارش کنید.

---

[7] Overfitting

[8] Tensorboard

- نمودار خسارت و دقت را برای دادگان آموزش و تست رسم کنید.
- ماتریس درهم ریختگی[9] را رسم نمایید.

ب) یک عکس از این پایگاه داده انتخاب کنید. به دلخواه، تعدادی از فیلترهای لایه‌های پیچشی و همچنین نتیجه اعمال آن فیلترها بر عکس انتخابی خود را به صورت گرافیکی رسم نمایید. آیا فیلتر های انتخابی شما در حال آموزش ویژگی خاص و قابل بیانی از تصویر می باشند؟

پ) شبکه نهایی خود را جهت ارزیابی میزان دقت در شناسایی حالت چهره (خود یا اطرافیان) از طریق وب‌کم لپ‌تاپ آزمایش کرده و نتیجه را گزارش دهید (بد نیست فیلمی در حد یک دقیقه از عملکرد این شبکه بر روی چهره خودتان را نیز بارگذاری کنید). ورودی شبکه‌تان عکس‌های گرفته شده در هر فریم توسط وب‌کم است.

۳- در یادگیری عمیق، مدل‌های شبکه عصبی می‌توانند بسیار بزرگ باشند. در بسیاری از مواقع برای کاربردهای کوچک به خصوص در شرایطی که با محدودیت در منابع محاسباتی مواجه هستیم، استفاده از این مدل‌های بزرگ مقرون به صرفه نیست. در چنین مواقعی مجبور هستیم تا با فدا کردن مقداری از عملکرد و دقت، به استفاده از مدل‌های کوچک‌تر و با تعداد پارامترهای کمتر روی بیاوریم و دادگان خود را با این مدل کوچک‌تر آموزش دهیم. طبیعتاً با کوچک‌تر کردن مدل، یادگیری بسیاری از خواص پیچیده دادگان برای مدل دشوار می‌شود. در این شرایط یکی از کارهایی که در راستای بهبود دقت مدل می‌توانیم انجام دهیم، پیاده‌سازی روش Knowledge Distillation است که در آن از یک مدل بزرگ‌تر به عنوان آموزگار استفاده می‌شود. به عبارتی دیگر در این روش، از یک مدل بزرگ‌تر که از پیش روی دادگان مرجع آموزش دیده است، استفاده می‌کنیم تا مدل کوچک‌تر را آموزش دهیم. برای مطالعه بیشتر می‌توانید به این مرجع مراجعه کنید. هم‌چنین در این مقاله به طور دقیق‌تر به این موضوع پرداخته شده است. به طور خلاصه، در این روش، لاجیت‌ها[10]ی مدل بزرگ‌تر به عنوان برچسب و به جای برچسب‌های اصلی دادگان ورودی، به مدل کوچک‌تر داده می‌شود تا آن را یاد بگیرد. تابع خسارت نهایی به صورت زیر است:

$$\mathcal{L}(x;W) = \alpha \times \mathcal{H}(y, \sigma(z_s; T=1)) + (1-\alpha) \times \mathcal{H}(\sigma(z_t; T=\tau), \sigma(z_s, T=\tau))$$

که در این رابطه x بیانگر ورودی، W پارامترهای مدل کوچک‌تر، y برچسب‌های واقعی، H تابع خسارت Cross-Entropy، σ تابع Softmax، α ضریبی برای ترکیب تابع خسارت عادی و distiller، و T نیز بیانگر Temperature اعمال شده در تابع Softmax است که بصورت زیر می‌باشد:

$$\sigma(z_i; T) = \frac{e^{\frac{z_i}{T}}}{\sum_j e^{\frac{z_j}{T}}}$$

در این تمرین قصد داریم یک شبکه عصبی با معماری MobileNetV2 را برای مسئله طبقه‌بندی با مجموعه دادگان CIFAR-10 آموزش دهیم.

الف) در این بخش قصد داریم از مفهوم Transfer Learning استفاده کنیم. برای این کار مدل از پیش آموزش دیده ResNet-50 روی ImageNet را آماده کنید. سپس لایه fully-connected نهایی آن را با یک لایه با سایز مناسب برای دادگان CIFAR-10

---

[9] Confusion Matrix

[10] مقادیر خروجی شبکه پس از عبور از لایه fully-connected نهایی

جایگزین کنید. حال پارامترهای دیگر شبکه را ثابت نگه دارید و تنها لایه آخر را بر روی دادگان CIFAR-10 آموزش دهید و نتایج را ارزیابی و ارائه نمایید.

ب) مدلی که در بخش قبل آموزش دادید را بعنوان مدل آموزگار انتخاب کنید و با آن یک مدل MobileNetV2 را از صفر بر روی دادگان CIFAR-10 آموزش دهید و نتایج را ارزیابی کنید. برای انتخاب هایپرپارامتر α و τ آزمون انجام دهید و تا جای امکان بهترین مقادیر را انتخاب کنید (توجه شود که برای انتخاب هایپرپارامترها نیاز نیست فرایند آموزش را با تعداد epoch زیاد انجام دهید. برای این کار دو یا سه بار آزمون با تعداد epoch کم کافیست. نتایج این آزمون‌ها را در جدولی ارائه کنید و بهترین هایپرپارامتر را انتخاب نمایید).

پ) در این مرحله مدل MobileNetV2 را از صفر و بدون آموزگار بر روی CIFAR-10 آموزش داده و ارزیابی کنید. دلیل تفاوت را توضیح دهید.

ت) در صورتی که در بخش (الف) به جای آموزش لایه آخر، کل مدل را Fine-tune می‌کردیم چه اتفاقی می‌افتد. این آزمایش را انجام دهید و ارزیابی نمایید و دلیل تفاوت را گزارش کنید.

نکات مورد توجه در زمینه ارائه فایل های سوالات عملی:

- ترجیحاً صفحات html از صفحه گوگل کولب یا جوپیتر که هم دارای کدها و هم پاسخ ها در زیر سلول های مربوطه می باشند را برای ما ذخیره و ارسال نمایید.
- پیشنهاد می شود بهترین مدل خود در هر سوال عملی را با فرمت h5. ذخیره کنید تا در صورت نیاز ما در آینده به آن ها، دسترسی ها مقدور باشد.
- در بخش عملی الزامی به ارائه گزارش نبوده و تنها کافی است توضیحات و فرضیات مورد نظر و همچنین بحث روی نتایج حاصله، به صورت مختصر و قابل فهم در کد پایتون نوشته شود.