دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر

پروژه درس رایانش عصبی و یادگیری عمیق

# پروژه‌ی چهارم

**هدف: استفاده از شبکه‌های کانولوشنی در تشخیص حالت چهره‌ی انسان**

**کد:** پیاده سازی این پروژه را به زبان پایتون انجام دهید؛ در این فعالیت مجاز به استفاده از tensorflow یا pytorch یا jaxمی‌باشید. فایل‌های کد خود را بر اساس شماره سوال و زیر قسمت خواسته شده‌ی آن نام گذاری کنید (برای مثال می‌توان نام گذاری قسمت اول برای سوال سوم تمرین را بصورت P3_a_ preprocessing.py در نظر گرفت). فایل‌های ارسالی‌تان بایستی با فرمت py. یا ipynb.(با حفظ خروجی هر سلول) باشد.

**گزارش:** ملاک اصلی انجام فعالیت، گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این فعالیت یک فایل گزارش در قالب pdf تهیه کنید که دارای فهرست بوده و پاسخ‌ها بترتیب در آن قرار گرفته اند و نام، نام خانوادگی و شماره دانشجویی‌تان در قسمت چپ سربرگ تمامی صفحات تکرار شده است. علاوه بر خواسته‌ی مستقیم هر سوال، مقتضی است که نمودار های خطا (loss) و صحت (accuracy) را به ازای مجموعه داده‌های آموزش و اعتبارسنجی رسم نمایید.
همچنین در صورت امکان ماتریس درهمریختگی را بصورت رنگ‌آمیزی شده به همراه اعداد متناظر برای مجموعه داده‌های آموزش، آزمون و اعتبارسنجی نیز تولید نمایید. لازم به ذکر است که در هر آموزش بایستی موارد مهم تنظیم شده نظیر تابع خطا، بهینه‌ساز(به همراه پارامترهای تنظیم شده‌ی آن مانند نرخ یادگیری)، معماری شبکه‌ی آموزشی(کتابخانه‌ها و ابزارهایی برای بصری‌سازی موجود است)، تعداد گام آموزشی، اندازه دسته(Batch Size)، آمارگان تفکیک مجموعه داده (به آموزش، آزمون و اعتبارسنجی)، پیش‌پردازش‌های اعمالی بروی دادگان ورودی و... ذکر گردد.

**تذکر:** مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و با تمامی طرفین برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری صرفا با ارجاع به آن بلامانع است، اما کپی کردن آن غیرمجاز است.

**راهنمایی:** در صورت نیاز میتوانید سوالات خود را در خصوص پروژه از تدریسیارهای درس، از طریق ایمیل زیر یا در گروه تلگرامی بپرسید.(لینک گروه تلگرامی)

Email: ann.ceit.aut@gmail.com CC: m.ebadpour@aut.ac.ir

**توجه:** می‌توانید از منابع و بسترهای سخت افزاری برخط رایگان نظیر Google Colab یا Kaggle استفاده نمایید.

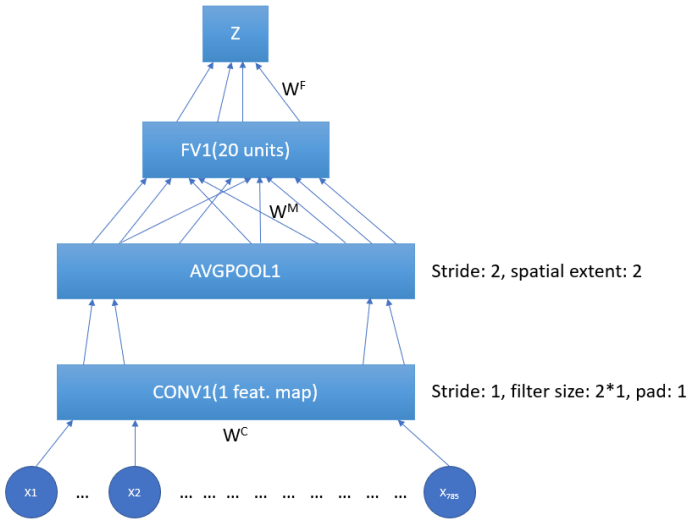
**تاخیر مجاز:** در طول ترم، ده روز زمان مجاز تاخیر برای ارسال پروژه‌ها در اختیار دارید(بدون کسر نمره). این تاخیر را می‌توانید بر حسب نیاز بین پروژه‌های مختلف تقسیم کنید که مجموع آن نباید بیشتر از ده روز شود. پس از استفاده از این تاخیر مجاز، هر روز تاخیر باعث کسر ۱۰٪ نمره‌ی کسب شده‌ی آن تمرین خواهد شد.

**ارسال:** فایل های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW04.zip تا تاریخ ۱۴۰۲/۰۲/۲۱ صرفا از طریق سایت کورسز ارسال نمایید. ارسال از طریق تلگرام، ایمیل و سایر راه‌های ارتباطی مجاز نبوده و تصحیح صورت نخواهد گرفت.

1 . نحوه اشتراک گذاری پارامترها در لایه‌های کانولوشنی، باعث ویژگی equivariance نسبت به translation می‌شود. این ویژگی را شرح دهید و کاربرد آنرا توضیح دهید. (۵ امتیاز)

2 . شبکه‌های عمیق از عدم تفسیرپذیری رنج می‌برند؛ تلاش برای حل این مشکل دو ایده‌ی deconvolutional[1] و up-convolutional[2] مطرح شده است. بررسی کنید و توضیح دهید هرکدام از دو روش به چه صورت منجر به تفسیرپذیری می‌شوند. (۱۲ امتیاز)

3 . معماری شبکه کانولوشنی شکل زیر را در نظر بگیرید:




- ابعاد ورودی 1 * 785 و خروجی شبکه 1 * 1
- لایه ورودی X با zero-padding با طول 1
- لایه کانولوشنی یک بعدی CONV1 با یک کرنل 1 * 2 و تابع فعال‌سازی ReLU
- لایه average-pooling (AVGPOOL1)
- لایه تمام متصل FC1 با تابع فعال‌سازی ReLU
- لایه خروجی Z که به لایه FC1 کاملا متصل است و تابع فعال‌سازی sigmoid

[1] https://arxiv.org/abs/1412.6806
[2] https://arxiv.org/abs/1506.02753

وزن لایه FC1 به Z را با $W_i^F$، بایاس Z را با $b^F$، وزن لایه AVGPOOL1 به FC1 را با $W_{ji}^A$، بایاس FC1 را با $b_i^M$، بردار $W^C$ برابر $[W_1^C, W_2^C]$ و بایاس لایه کانولوشنی را با $b^C$، نشان می‌دهیم. داده‌های مجموع آموزش به صورت $X^i$ و خروجی مورد انتظار به صورت $Y^i$ است. همچنین خروجی‌های لایه‌های شبکه به ترتیب $c(X^i)$، $a(X^i)$، $f(X^i)$ و $z(X^i)$ می‌نامیم. در این صورت تابع هزینه به صورت زیر تعریف می‌شود:

$$cost(X,Y) = \sum_n cost\left(X^{(n)}, Y^{(n)}\right) = \sum_n (-Y^{(n)} \log\left(z\left(X^{(n)}\right)\right) - (1 - Y^{(n)}) \log\left(1 - z(X^{(n)})\right))$$

با توجه به مفروضات بالا به پرسش‌های زیر پاسخ دهید:

الف) تعداد پارمترهای شبکه بالا را با ذکر جزئیات محاسبه کنید. (۵ امتیاز)

ب) برای فقط یک نمونه آموزشی مقدار $\frac{\partial Cost}{\partial W_1^C}$ و $\frac{\partial Cost}{\partial W_{ji}^A}$ را با جزئیات محاسبه کنید. (۱۵ امتیاز)

4. کانولوشن متسع[3] روشی برای افزایش میدان پذیرش (Receptive Field) شبکه‌های کانولوشنی است که به صورت زیر تعریف می‌شود: (دقت شود خروجی تنها برای اندیس‌هایی که کرنل و تصویر همپوشانی کامل دارند، محاسبه می‌شود)

$$(K \star_D I)(i,j) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} K(m,n) I(i + Dm, j + Dn)$$

الف) در یک شبکه کانولوشنی با یک لایه کانولوشن K * K با طول گام یک، عرض میدان پذیرش را بدست آورید. (۵ امتیاز)

ب) برای ورودی $I \in \mathbb{R}^{M \times N}$ و کرنل $K \in \mathbb{R}^{F \times F}$، نشان دهید خروجی عملگر متسع دارای ابعاد $(M - DF + D) \times (N - DF + D)$ است. متغیر D به معنی نرخ dilation است. (۵ امتیاز)

ج) نشان دهید کانولوشن متسع معادل کانولوشن با کرنل متسع شده $K' = K \otimes A$ است. ماتریس A را مشخص کنید. (عملگر $\otimes$ به معنی kronecker product است) (۵ امتیاز)

5. شبکه‌های کانولوشنی با توجه به توانایی آنها در استخراج و یادگیری خودکار ویژگی‌ها، مقاومت نسبت به تغییرات و کارایی آنها در مقابل پیچیدگی‌های وظیفه‌ی بازشناسی چهره، یک عنصر اساسی در اکثر این سیستم‌ها هستند. در این تمرین قصد داریم که با استفاده از شبکه‌های عصبی کانولوشنی به تحلیل احساسات چهره[4] و طبقه بندی آنها از روی تصویر

[3] dilated convolution
[4] Facial Expression Recognition

بپردازیم. مجموعه داده‌ی این تمرین شامل ۱۲۰۰ تصویر نمونه‌گیری شده از هر کلاس مجموعه داده‌ی AffectNet می‌باشد.

مجموعه داده‌ی AffectNet شامل ۴۵۰ هزار تصویر چهره با ۸ حالت مختلف می‌باشد که شکل ۱ نمونه‌هایی از آن را نشان می‌دهد.


شکل ۱: نمونه‌هایی از مجموعه داده‌ی AffectNet

**الف) پیش پردازش و داده‌افزایی**[5]: مجموعه داده را از این لینک دانلود کنید و از هر کلاس سه نمونه را نمایش دهید. برای افزایش سرعت آموزش تمامی تصاویر را به بازه‌ی $[0,1]$ نرمالسازی کنید. همچنین داده‌ها را با پردازش مناسب افزونه کنید. توضیح دهید که به نظر شما استفاده از چه پردازش‌هایی در این حالت مناسب است و چرا در این مساله نیاز به داده‌افزایی وجود دارد؟ از هر کلاس سه نمونه‌ی افزونه‌شده را نمایش دهید و همچنین تعداد کل نمونه‌ها پیش و پس از داده‌افزایی را در گزارش خود بیاورید.(۱۰ امتیاز)

**ب)یادگیری انتقالی**[6] یک رویکرد رایج در هوش مصنوعی است که از یک مدل از قبل آموزش دیده برای یک وظیفه‌ی متفاوت اما مرتبط استفاده می‌کند و آن را با وظیفه‌ی جدید تطبیق می‌دهد. با استفاده از شبکه‌ی پیش آموزش دیده‌ی VGG16 وظیفه‌ی بازشناسی حالت چهره را بر روی مجموعه داده‌ی ارائه شده انجام دهید. برای فرآیند آموزش از داده‌های موجود در پوشه‌ی train استفاده کنید. نمودار خطا و دقت در فرآیند آموزش و نمودار ROC و ماتریس در هم ریختگی[7] را برای داده‌های موجود در پوشه‌ی validation گزارش کنید.(۳۵ امتیاز)

[5] Data Augmentation
[6] Transfer Learning
[7] Confusion Matrix

به کارگیری شبکه‌های از پیش آموزش دیده به طور خاص در زمانی که داده‌ی کمی وجود دارد مزایای زیادی دارد اما این شبکه‌ها با توجه به معماری از پیش تعریف شده و نسبتا سنگین آنها برای استفاده در ابزارهای کاربردی مانند تلفن همراه مناسب نیستند. مدل‌های موجود در تلفن‌های همراه باید نیازهای ذخیره‌سازی را به حداقل برسانند و در عین حال افت عملکرد قابل توجهی نداشته باشند. برای دستیابی به این امر در این مقاله سه معماری سبک از سه شبکه‌ی کانولوشنی مطرح یعنی AlexNet،VGG و MobileNet مطرح شده است. نتایج به دست آمده نشان می‌دهد که این سه معماری عملکرد مشابهی نسبت به آخرین مدل‌های پیشرو در این زمینه دارند.

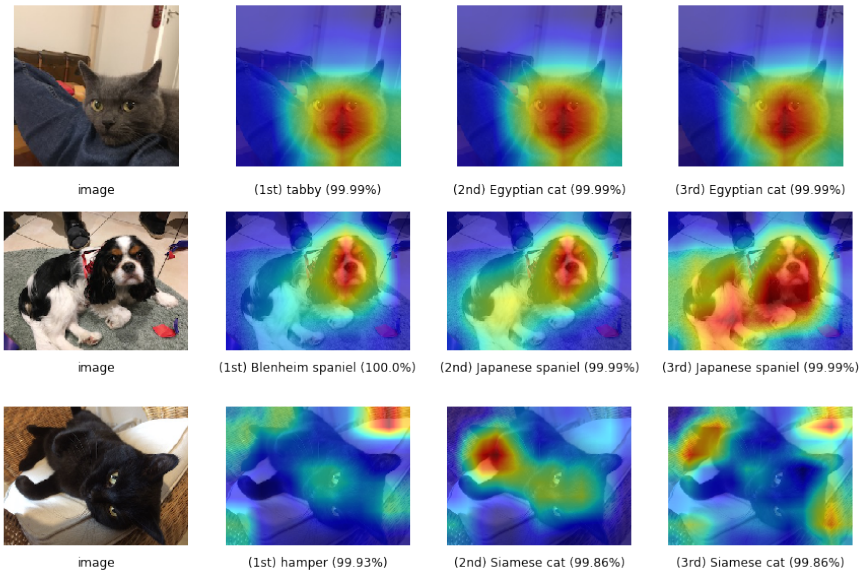
ج) معماری مطرح شده برای شبکه‌ی VGG که جزئیات آن در شکل ۲ آمده است را پیاده‌سازی کنید. این مدل را بر روی مجموعه داده‌ی ارائه شده آموزش دهید و نمودار خطا و دقت آن را رسم کنید. همچنین با استفاده از داده‌ی موجود در پوشه‌ی validation مدل را تست کنید و نمودار ROC و ماتریس در هم ریختگی آن را گزارش کنید. تعداد پارامترهای این مدل و عملکرد آن را با مدل قسمت قبل مقایسه و تحلیل کنید.(۳۵ امتیاز)

| Type | Shape | Output |
|---|---|---|
| 2×Conv | $3 \times 3 \times 16$ | $128 \times 128 \times 16$ |
| MaxPool | $2 \times 2$ | $64 \times 64 \times 16$ |
| 2×Conv | $3 \times 3 \times 32$ | $64 \times 64 \times 32$ |
| MaxPool | $2 \times 2$ | $32 \times 32 \times 32$ |
| 2×Conv | $3 \times 3 \times 64$ | $32 \times 32 \times 64$ |
| MaxPool | $2 \times 2$ | $16 \times 16 \times 64$ |
| 2×Conv | $3 \times 3 \times 128$ | $16 \times 16 \times 128$ |
| MaxPool | $2 \times 2$ | $8 \times 8 \times 128$ |
| 2×Conv | $3 \times 3 \times 128$ | $8 \times 8 \times 128$ |
| MaxPool | $2 \times 2$ | $4 \times 4 \times 128$ |
| Flatten | 2048 | – |
| 2×Dense | 1024 | – |
| Dense | 8 or 2 | 1 label or 2 floats |

شکل ۲: معماری شبکه‌ی VGG ارائه شده در مقاله

د)برای درک هر چه بهتر عملکرد شبکه‌های کانولوشنی ابزارهای متنوعی وجود دارد. یکی از این ابزارها نقشه‌ی فعالسازی کلاس[8] یا به اختصار CAM است که یک نمونه از آن در شکل ۳ آمده است. بررسی کنید که استفاده از این ابزار چه بینشی برای بهبود شبکه‌های کانولوشنی فراهم می‌آورد. برای دو نمونه‌ی به اشتباه دسته بندی شده و دو نمونه‌ی به درستی دسته بندی شده‌ی به ازای هر کلاس در مدل سوال ۳ نقشه‌ی فعالسازی کلاس را به دست آورید و با تحلیل نتایج به دست آمده، رویکردی برای بهبود شبکه‌ی پیشنهادی سوال ۳ ارائه دهید.(۲۰ امتیاز)

[8] Class Activation Map




شکل ۳: نقشه‌ی فعالسازی کلاس برای مساله‌ی دسته بندی سگ و گربه