بسم الله الرحمن الرحیم

دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر

یادگیری عمیق
نیم‌سال دوم ۰۳-۰۲
مدرس: مهدیه سلیمانی

تمرین پنجم

- مهلت ارسال پاسخ تا ساعت **۲۳:۵۹** روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر تمرین‌های نظری بدون کسر نمره تا سقف ۵ روز و تمرین‌های عملی تا سقف ۱۰ روز وجود دارد. محل بارگزاری جواب تمرین‌های نظری بعد از ۳ روز و تمرین‌های عملی بعد از ۵ روز بسته خواهد شد و پس از گذشت این مدت، پاسخ‌های ارسال‌شده پذیرفته نخواهند شد.
- هم‌فکری در انجام تمرین مانعی ندارد، فقط توجه داشته باشید که پاسخ تمرین حتما باید توسط خود شخص نوشته شده باشد. همچنین در صورت همفکری در هر تمرین، در ابتدای جواب تمرین نام افرادی که با آن‌ها هم‌فکری کرده اید را حتما ذکر کنید.
- برای پاسخ به سوالات نظری در صورتی که از برگه خود عکس تهیه می‌کنید، حتما توجه داشته باشید که تصویر کاملا واضح و خوانا باشد. درصورتی که خوانایی کافی را نداشته باشد، تصحیح نخواهد شد.
- محل بارگذاری سوالات نظری و عملی در هر تمرین مجزا خواهد بود. به منظور بارگذاری بایستی تمارین تئوری در یک فایل pdf با نام HW5_[First-Name]_[Last-Name]_[Student-Id].pdf و تمارین عملی نیز در یک فایل مجزای زیپ با نام HW5_[First-Name]_[Last-Name]_[Student-Id].zip بارگذاری شوند.
- در صورت وجود هرگونه ابهام یا مشکل، در کوئرای درس آن مشکل را بیان کنید و از پیغام دادن مستقیم به دستیاران آموزشی خودداری کنید.
- طراحان این تمرین : آقایان سامتی و هادیان

بخش نظری (۳۵ نمره)

سوال اول: (۱۵ نمره)

۱. با توجه به مقاله‌های SimCLR، MoCo و BYOL به سوالات زیر پاسخ دهید.

(آ) اهمیت وجود negative sampleها برای بدست آوردن بازنمایی چیست؟ چطور در روش BYOL این نیاز برطرف شده است؟

(ب) در روش SimCLR نویسندگان چه تمهیداتی را برای آموزش مدل با توجه به بزرگ بودن batch size در نظر گرفته‌اند؟

۲. با توجه به به مقاله‌های DINO و DINOv2 به سوالات زیر پاسخ دهید.

(آ) شبکه teacher دقیقا مشابه شبکه student و بدون آموزش پیشینی از ابتدا، به صورت Exponential Moving Average از شبکه student آپدیت می‌شود. در ابتدا چه چیزی باعث بهتر بودن بازنمایی بدست آمده از شبکه teacher و در نتیجه انتقال دانش به شبکه student می‌شود؟

(ب) در فرایند آموزش چگونه شبکه teacher به سمت توجه کردن به شئ و نادیده گرفتن پس زمینه در تصویر تشویق می‌شود؟

(ج) برای جلوگیری از collapse راهکار نویسندگان هر مقاله چیست؟

(د) نکات پیاده‌سازی در مقاله DINOv2 را مختصر توصیح دهید.

(ه) تفاوت‌های اصلی DINO با BYOL را مشخص کنید و توضیح دهید این تفاوت‌ها چطور در نتایج چشمگیر مدل DINO اثر داشته؟

سوال دوم: (۸ نمره)

۱. برای یک گراف بدون جهت و بدون برچسب روی یال‌ها، تابعی که در هر لایه از یک شبکه عصبی گراف محاسبه می‌کنیم باید ویژگی‌های خاصی را رعایت کند تا بتوان از همان تابع (با اشتراک‌گذاری وزن‌ها) در گره‌های مختلف گراف استفاده کرد. فرض کنید برای یک گره مشخص i در گراف، $h_i^{\ell-1}$ پیام خودی (یعنی حالتی که در لایه قبلی برای این گره محاسبه شده است) برای این گره از لایه قبلی باشد، در حالی که پیام‌های لایه قبلی از n_i همسایه‌های گره i با $m_{i,j}^{\ell-1}$ نشان داده می‌شوند که j از ۱ تا n_i می‌باشد. ما از w با زیرنویس‌ها و بالانویس‌ها برای نشان دادن وزن‌های قابل یادگیری استفاده خواهیم کرد. اگر بالانویسی وجود نداشته باشد، وزن‌ها در سراسر لایه‌ها به اشتراک گذاشته می‌شوند. فرض کنید که همه ابعاد به درستی کار می‌کنند. توضیح دهید کدام یک از این‌ها توابع معتبر برای محاسبه پیام بعدی h_i^ℓ برای این گره هستند. برای هر انتخابی که معتبر نیست، به طور مختصر دلیل آن را ذکر کنید.

توجه: اعتبار به این معنی است که آن‌ها باید Invariance و Equivariance را که برای استفاده به عنوان یک GNN روی یک گراف بدون جهت نیاز داریم، رعایت کنند.

(آ) $h_i^\ell = w_1 h_i^{\ell-1} + w_2 \frac{1}{n_i} \sum_{j=1}^{n_i} m_{i,j}^{\ell-1}$

(ب) $h_i^\ell = \max(w_1^\ell h_i^{\ell-1}, w_2 m_{i,1}^{\ell-1}, w_3 m_{i,2}^{\ell-1}, \ldots, w_{n_i-1} m_{i,n_i}^{\ell-1})$ که در آن max به صورت مؤلفه‌ای بر روی بردارها عمل می‌کند.

(ج) $h_i^\ell = \max(w_1^\ell h_i^{\ell-1}, w_2 m_{i,1}^{\ell-1}, w_2 m_{i,2}^{\ell-1}, \ldots, w_2 m_{i,n_i}^{\ell-1})$ که در آن max به صورت مؤلفه‌ای بر روی بردارها عمل می‌کند.

۲. یک شبکه عصبی گراف (GNN) برای دسته‌بندی گره‌ها در یک گراف بی‌جهت $G = (V, E)$ در نظر بگیرید که در آن V مجموعه رئوس و E مجموعه یال‌ها است. GNN با استفاده از یک مکانیزم ارسال پیام به‌صورت تکراری ویژگی‌های گره‌ها را به‌روز می‌کند. فرض کنید $\mathbf{H}^{(t)}$ ماتریس ویژگی گره‌ها در تکرار t باشد که هر سطر $\mathbf{h}_v^{(t)}$ بردار ویژگی گره v را نشان می‌دهد.

عملیات ارسال پیام در تکرار t را به صورت زیر تعریف کنید:

$$\mathbf{h}_v^{(t+1)} = \sigma\left(\mathbf{W}\mathbf{h}_v^{(t)} + \sum_{u \in \mathcal{N}(v)} \mathbf{W}'\mathbf{h}_u^{(t)}\right),$$

که در آن:

(آ) $\mathbf{W}$ و $\mathbf{W}'$ ماتریس‌های وزن قابل یادگیری هستند،

(ب) $\mathcal{N}(v)$ مجموعه همسایگان گره v است،

(ج) σ یک تابع فعال‌سازی غیرخطی است.

ثابت کنید که GNN فوق نسبت به هر جایگشت گره‌ها هموردا است. به عبارت دیگر، نشان دهید که برای هر ماتریس جایگشت $\mathbf{P}$، ویژگی‌های گره‌ها $\mathbf{H}^{(t)}$ پس از t تکرار، معادله زیر را ارضا می‌کند:

$$\mathbf{P}\mathbf{H}^{(t+1)} = \mathbf{H}_{\mathbf{P}}^{(t+1)}$$

که در آن $\mathbf{H}_{\mathbf{P}}^{(t+1)}$ ویژگی‌های گره‌هایی هستند که با اعمال همان عملیات GNN بر روی گراف جایگشت‌داده‌شده به دست می‌آیند.

سوال سوم: (۶ نمره)

با مطالعه‌ی مقاله‌ی Universal adversarial perturbations در مورد آشفتگی خصمانه‌ی فراگیر به سوالات زیر پاسخ دهید:

۱. به صورت خیلی مختصر و در حد یک الی دو جمله توضیح دهید که آشفتگی خصمانه‌ی فراگیر چیست.

۲. چرا یافتن چنین آشفتگی‌ای مهم است؟

۳. با داشتن داده‌های D و تابع g که میزان موفقیت حمله را اندازه‌گیری می‌کند (هر چه $g(x)$ بیشتر باشد، حمله موفق‌تر است) یافتن آشفتگی خصمانه‌ی فراگیر را به صورت یک مسئله‌ی بهینه‌سازی مقید به کرانی بر روی نرم p آشفتگی بنویسید.

سوال چهارم: (۶ نمره)

۱. در درس با مدل‌های CLIP, SimVLM و CoCa آشنا شدید. دو شباهت و دو تفاوت این مدل‌ها را بیان کنید و موارد مربوطه را برای هر مدل به صورت واضح مشخص کنید.

۲. در مورد مدل CLIP به سوالات زیر پاسخ دهید:

(آ) ماژول‌های موجود در این مدل را نام برده و عملکرد هر کدام را توضیح دهید. برای هر کدام از این ماژول‌ها از چه مدلی استفاده شده است؟

(ب) می‌دانیم تابع هزینه‌ای که برای آموزش این مدل استفاده شده است از دو بخش تشکیل شده است که بخش اول آن به شکل زیر است:

$$\mathcal{L}_1 = -\frac{1}{N}\log\frac{e^{sim(x_i,y_i)/\tau}}{\sum_{j=1}^{N} e^{sim(x_i,y_j)/\tau}}$$

در صورتی که داشته باشیم $s_{ij} = \frac{x_i.y_j}{\tau\|x_i\|\|y_j\|}$ ، مشتق $\mathcal{L}_1$ نسبت به x_i یعنی $\frac{\partial\mathcal{L}_1}{\partial x_i}$ را محاسبه کنید.

بخش عملی (۶۵ نمره)

توجه: لطفا در کلیه سوال‌های عملی نوت بوک تکمیل شده خود را به همراه سایر موارد در کوئرا بارگذاری کنید و از ارسال لینک و به اشتراک گذاری نوت بوک خودداری فرمایید.

سوال اول: (۲۰ نمره)

هدف از این سوال، آشنایی با مدل DINOv2 است. در این سوال از شما خواسته می‌شود با اضافه کردن یک لایه ترنسفورمر بر روی ویژگی‌های استخراج شده از داینو یک دسته‌بند بسازید. هدف این مدل دسته‌بندی تصاویر ماهواره‌ای برای تشخیص وجود یا عدم وژود پنل‌های خورشیدی است. همچنین در ادامه با استفاده از ماتریس توجه به دست آمده از آن می‌توانید اندازه پنل‌ها را تخمین بزنید.

سوال دوم: (۲۵ نمره)

نوت‌بوک StableDiffusion.ipynb شامل سه بخش پیاده‌سازی StableDiffusionPipeline به صورت Classifier-free guidance، اضافه کردن guidance اضافی برای تقویت رنگ آبی در تصویر تولیدی و در آخر fine tune کردن مدل برای آموزش یک مفهموم جدید شامل یک جفت توکن متنی و تصویر مربوطه به مدل با روش معرفی شده در مقاله DreamBooth است. در بخش آخر می‌توانید تصاویر مورد نظر خودتان را استفاده کنید.

توجه: نوتبوک برای محیط Google Colab بهینه شده است و توصیه می‌شود از این محیط برای تکمیل نوتبوک استفاده شود.

سوال سوم: (۲۰ نمره)

در این قسمت قرار است تا مباحث امنیت در یادگیری ماشین را به صورت عملی پیاده‌سازی کنید. برای انجام این بخش به نوت بوک `Adversarial_attacks_training.ipynb` مراجعه کنید. در این نوت بوک ابتدا دو نوع حمله FGSM و PGD را یکی با استفاده از کتابخانه و یکی به صورت from scratch پیاده‌سازی می‌کنید و سپس به کمک هردوی آن‌ها، آموزش خصمانه روی مدل انجام می‌دهید. تمامی بخش‌های داخل نوت‌بوک را تکمیل کنید و به سوالات مطرح شده پاسخ دهید.