



به نام خدا  
دانشگاه تهران  
دانشکده مهندسی  
برق و کامپیوتر



## درس داده کاوی تمرین ششم

محمد صادق صادقی <a href="mailto:Mhmssadeghi74@gmail.com">Mhmssadeghi74@gmail.com</a>	طراح
۱۶ خرداد ۱۴۰۳	تاریخ بارگذاری
۲۷ خرداد ۱۴۰۳	مهلت ارسال

## فهرست

۲.....	مقدمه
۳.....	مجموعه داده
۴.....	بخش عملی
۵.....	بخش تشریحی
۶.....	ملاحظات

### دسته بندی نودهای گراف با کمک شبکه‌ی عصبی گرافی (GNN):

در این تمرین از شما انتظار می‌رود، با کمک شبکه‌ی عصبی گرافی بر روی مجموعه داده‌ی musae-github وظیفه‌ی رده‌بندی نودهای گراف را انجام بدهید. هدف از این تمرین آشنایی با نحوه تجزیه و تحلیل و آماده‌سازی داده‌ها برای آموزش شبکه‌ی عصبی و همین طور آشنایی با چگونگی عملکرد یک شبکه‌ی عصبی گرافی و پیاده‌سازی آن خواهد بود.

داده‌های musae-github یک شبکه اجتماعی از توسعه‌دهندگان GitHub است، که در آن گره‌ها، نشان‌دهنده‌ی کاربران و یال‌ها روابط متقابل این کاربران را نشان می‌دهند. هر کدام از گره‌ها شامل مجموعه‌ای از ویژگی‌ها هستند. در این مجموعه داده هر کدام از گره‌ها دارای برچسب هستند و این برچسب مشخص می‌کند هر کاربر جزو توسعه‌دهندگان یادگیری ماشین خواهد بود یا خیر، برای دسترسی به این دیتاست می‌توانید از این [لینک](#) استفاده کنید.

در این تمرین از شما انتظار می رود با کمک الگوریتم GCNN وظیفه رده بندی گره های مجموعه ی داده ی بیان شده صورت پذیرد. هدف از این کار این است که شما تعیین کنید یک گره آیا متعلق به توسعه دهندگان یادگیری ماشین است یا خیر. بر همین اساس لازم است که هر کدام از بخش های زیر به طور کامل مورد بررسی قرار گرفته و گزارش آن ارائه گردد.

الف: در ابتدا لازم است مجموعه داده ی در دسترس توسط شما مورد بررسی قرار بگیرد و بیان گردد هر کدام از داده های موجود در برگیرنده ی چه اطلاعاتی است.

ب: در این بخش، از شما انتظار می رود تحلیلی از مجموعه داده در دسترس را بیان کنید، بر همین اساس لازم است در ابتدا نحوه ی توزیع برچسب های هر دسته از کلاس ها در قالب نمودار ارائه گردد. آیا این مجموعه داده متعادل است یا خیر؟ در صورتی که مجموعه داده متعادل نیست چه راهکاری را در جهت حل این مشکل ارائه می دهید؟

ج: در این بخش لازم است، معماری مربوط به مدل GCNN را پیاده سازی کرده و با تقسیم مجموعه داده به سه بخش آموزش، اعتبارسنجی و تست مدل را آموزش بدهید.

د: در پایان لازم است، با معیارهای ارزیابی مناسب عملکرد مدل آموزش دیده را بر روی داده های تست نمایش دهید.

### توجه:

- در این بخش شما میتوانید از کتابخانه های آماده استفاده کنید.
- به تمرین هایی که صرفا پیاده سازی کد هستند نمره ای تعلق نخواهد گرفت و گزارش ارائه شده معیار اصلی نمره شما خواهد بود.

فرض کنید شما بر روی یک سیستم توصیه‌گر برای فروش محصولات کار می‌کنید. این سامانه شامل تعداد زیادی از کاربران است و هر کاربر با تعداد زیادی از محصولات تعامل دارد (شامل مشاهده محصول، خرید و امتیازدهی)، همچنین اطلاعات دیگر مانند کاتالوگ محصولات در اختیار شما قرار دارد که شامل اطلاعات کامل و جامعی به منظور توصیف هر محصول است.

اکنون وظیفه شما طراحی یک سیستم توصیه‌گر است، که به شکل کارآمد محصولاتی را براساس تعاملات گذشته‌ی افراد به آن‌ها پیشنهاد کند. فرض کنید از شما خواسته شده است که از شبکه عصبی گرافی به منظور در نظر گرفتن رابطه‌ی بین محصولات و کاربران و همین طور سایر اطلاعات مرتبط با آن‌ها استفاده گردد.

- 1- توضیح دهید چرا شبکه‌ی عصبی گرافی می‌تواند در این وظیفه برای شما کمک‌کننده باشد؟
- 2- بیان کنید، شما گراف را به چه شکل به عنوان ورودی این شبکه‌ی عصبی گرافی در نظر خواهید گرفت.
- 3- بیان کنید چگونه بازنمایی‌هایی بدست آمده از گره‌های این شبکه‌ی عصبی گرافی می‌تواند در پیشنهاد محصولات به هرکدام از کاربران کمک‌کننده باشد؟
- 4- از نظر شما، چه محدودیت‌ها و چالش‌هایی می‌تواند، برای سامانه‌ی پیشنهادی شما وجود داشته باشد و راه‌حل‌های پیشنهادی خود را در جهت حل این مشکلات بیان کنید. (حداقل یک چالش به همراه راه‌حل آن بیان گردد)

- تمامی نتایج شما باید در یک فایل فشرده با عنوان DM\_CA6\_StudentID تحویل داده شود.
- خوانایی و دقت بررسی‌ها در گزارش نهایی از اهمیت ویژه‌ای برخوردار است. به تمرین‌هایی که به صورت کاغذی تحویل داده شوند یا به صورت عکس در سایت بارگذاری شوند، ترتیب اثری داده نخواهد شد.
- بخش اصلی نمره به گزارش شما تعلق می‌گیرد و دستیاران الزامی برای اجرای تمام کدهای شما در صورتی که در گزارش به آن‌ها اشاره‌ای نکرده باشید ندارند. لطفاً تمام موارد مورد نیاز را در گزارش ذکر کنید.
- کدهای نوشته شده برای هر بخش را با نام مناسب مشخص کرده و به همراه گزارش تکلیف ارسال کنید. همه‌ی کدهای پیوست گزارش بایستی قابلیت اجرای مجدد داشته باشند. در صورتی که برای اجرا مجدد آنها نیاز به تنظیمات خاصی می‌باشد بایستی تنظیمات مورد نیاز را نیز در گزارش خود ذکر کنید.
- برای تحویل تمرین از چارچوب قرارداده شده در سامانه و کانال تلگرام استفاده کنید.
- در صورت قصد ارسال تمرین به صورت دیگر (انگلیسی، latex و ...)، لطفاً پیش از ارسال با دستیارمسئول تمرین هماهنگ کنید.
- توجه کنید این تمرین باید به صورت تک نفره انجام شود و پاسخ‌های ارائه شده باید نتیجه فعالیت فرد نویسنده باشد (هم‌فکری خارج از چارچوب و به اتفاق هم نوشتن تمرین نیز ممنوع است). در صورت مشاهده تقلب برای همه‌ی افراد مشارکت کننده، نمره تمرین، صفر در نظر گرفته خواهد شد.
- در صورت بروز هرگونه مشکل با ایمیل زیر در ارتباط باشید:

<mailto:mhmssadeghi74@gmail.com>

مهلت تحویل: ۱۴۰۳/۳/۲۷