



مبانی یادگیری ماشین - پروژه نهایی

ددلاین: ۳۰ دی ماه، ساعت ۸ صبح

مدرس: دکتر حامد ملک

نیمسال اول ۱۴۰۱-۱۴۰۲

- ۳..... پروژه دونفره
- ۳..... هدف پروژه
- ۳..... دیتاست
- ۳..... مدل
- ۴..... پروژه تک نفره
- ۴..... هدف پروژه
- ۴..... فاز اول: جمع آوری و پیش پردازش داده
- ۵..... فاز دوم: انتخاب و طراحی مدل
- ۵..... فاز سوم: آموزش مدل
- ۵..... پروژه MLOps
- ۶..... نکات

توضیح: برای پروژه نهایی دو موضوع انتخاب شده است که باید یکی از آن ها را به اختیار خودتان انتخاب کنید. پروژه اول گروهی و در قالب گروه های دو نفره و پروژه دوم انفرادی انجام می شود. همچنین پروژه MLOps برای هر دو گروه اجباری است. دقت بفرمایید که در هر دو پروژه فقط مجاز به استفاده از کتابخانه های PyTorch، Scikit-learn و tensorflow هستید.

پروژه دونفره

هدف پروژه

در این پروژه باید به کمک روش‌های گفته‌شده، یک سیستم توصیه‌گر فیلم بسازید.

دیتاست

دیتاست موجود متعلق به ۹۰۰۰ فیلم است که توسط ۷۰۰ کاربر امتیازدهی شده‌اند. در مجموع ۱۰۰,۰۰۰ نمره‌دهی در دیتاست ثبت شده است (برخی کاربران به برخی فیلم‌ها امتیازی نداده‌اند).

این دیتاست شامل اطلاعاتی در مورد فیلم‌های دیتاست هم هست:

- فایل movies_metadata.csv شامل اطلاعاتی نظیر ژانر فیلم، زبان فیلم، عنوان و خلاصه‌ای از داستان است.

- فایل rating.csv که شامل امتیاز کاربران به فیلم‌ها است.

- فایل keywords.csv هم شامل کلمات کلیدی مرتبط با یک فیلم است.

- فایل credits.csv هم شامل بازیگران و دست‌اندرکاران یک فیلم است.

در دیتاست موجود اطلاعات زیادی موجود است که ممکن است نیازی به استفاده از همه‌ی آنها نباشد. مثلاً، در فایل movies_metadata.csv لینک صفحه‌ی IMDB هر فیلم نیز موجود است، اما احتمالاً نیازی به استفاده از آن در ساخت سیستم توصیه‌گر نخواهیم داشت. بنابراین، تمیز کردن دیتاست می‌تواند بخشی از پروژه محسوب شود.

مدل

- پیاده‌سازی به کمک content-based filtering

در این روش به کمک شباهت میان فیلم‌های مختلف و فیدبک‌هایی که از رفتار مردم در گذشته دارید، فیلم‌ها رو پیشنهاد می‌دهید.

- پیاده‌سازی به کمک Collaborative filtering

برای رفع برخی از محدودیت‌های روش content-based filtering، این روش شباهت‌های بین کاربران و موارد به طور همزمان برای ارائه توصیه‌ها استفاده می‌کند. در این بخش، باید به کمک collaborative filtering سیستم توصیه‌گر را بسازید.

- مدل ensemble

در این بخش می‌توانید به کمک مدل‌های ساخته شده در بخش قبل، یک مدل توصیه‌گر ترکیبی بسازید.

پروژه تک‌نفره

هدف پروژه

هدف این پروژه، تشخیص نوع ماشین براساس تصویری از آن ماشین است.

کلاس‌های خروجی مد نظر:

- تیبیا
- پژو ۲۰۶
- پراید
- پژو پارس
- سمند
- دنا

فاز اول: جمع‌آوری و پیش‌پردازش داده

برای آموزش شبکه عصبی نیاز است تا داده جمع‌آوری کنید. این کار را می‌توان با استفاده از یک crawler به راحتی انجام داد. (پیشنهاد می‌شود که به ازای هر کلاس حداقل ۵۰۰ تا تصویر داشته باشید.)

نکاتی که در پیش‌پردازش کمک‌کننده است.

- تقسیم دیتاست به بخش‌های train و validation و test
- از آنجا که داده‌هایتان از اینترنت گرفته شده است، مطمئن شوید داده‌ها ابعاد یکسانی دارند. (برای مثال $100 * 100$)
- مطمئن شوید که کلاس‌های متفاوت، عکس‌های برابری داشته باشند تا موقع یادگیری bias رخ ندهد.
- پیشنهاد می‌شود که label‌هایتان را به صورت ترتیبی encode نکنید و از یک encoding غیر ترتیبی استفاده کنید.
- انواع تبدیل‌ها را برای داده‌ها تست کنید مانند تبدیل به grayscale.

فاز دوم: انتخاب و طراحی مدل

پیشنهاد می‌شود که از مدل‌های متفاوت استفاده کنید و دقت‌ها را مقایسه کنید. این کار را با استفاده از دیتای validation انجام دهید و در نهایت نتایج خود را با کمک نمودار نمایش دهید و در تحلیل کردن استفاده کنید. در این پروژه، فقط مجاز به طراحی با استفاده از کتابخانه‌ی tensorflow و pytorch می‌باشید.

فاز سوم: آموزش مدل

برای آموزش مدل و بدست آوردن بهترین نتیجه نیاز است تا مقادیر ابر پارامترها (تمامی مدل‌ها)، تعداد لایه‌ها و توابع فعال سازی (در شبکه عصبی) خود را تغییر دهید و بررسی کنید که با چه تغییراتی توانستید دقت خود را بهبود دهید در گزارش پروژه تغییرات داده شده و نتیجه‌های خود را ذکر کنید و با استفاده از نمودار تغییر در دقت بر اساس پارامترهای متفاوت را نمایش دهید.

پروژه MLOps

در این مرحله نیاز است تا برای پروژه‌ای که در بخش قبل پیاده‌سازی کردید، با استفاده از فریم‌ورک [MLflow](#) بخش MLOps را انجام دهید. لازم به ذکر است که پروژه MLOps برای هر دو گروه یکسان و اجباری است.

۱. ثبت رکورد دیتاست‌ها و مدل‌های مختلف

متادیتای دیتاست استفاده‌شده برای هر بار آموزش و همچنین مدل‌های متفاوت خود را ثبت کنید و در نهایت مدل/مدل‌های اصلی‌تان را ثبت کنید تا بعداً به راحتی بتوانید استفاده کنید.

۲. پیاده‌سازی API

یک API پیاده‌سازی کنید تا بتوانیم با ارسال ورودی به مدل، خروجی مدل را دریافت کنیم.

۳. مانیتور کردن مدل - امتیازی

متریک‌های مناسب با توجه به مدل‌تان انتخاب کنید و سپس آن‌ها را مانیتور کنید تا بتوان هنگامی که مدل deploy شده‌است، آن را در گذر زمان مشاهده کرد.

۴. بروزرسانی مدل - امتیازی

فرآیندی را پیاده‌سازی کنید تا پس از آنکه مدل deploy شد، با داده‌هایی که از طریق API دریافت می‌کند بصورت batch بروزرسانی شود.

۵. آپلود کردن در Hugging Face – امتیازی

در سایت [Hugging Face](#) برای مدل‌تان یک ریپازیتوری ساخته و سپس Model Card بسازید و مدل را آنجا آپلود کنید.

نکته: میزان نمره دریافتی شما از بخش امتیازی متأثر از عملکرد شما و نظر گروه حل تمرین خواهد بود.

نکات

۱. دیتاست پروژه دونفره از [اینجا](#) قابل دسترسی است.
۲. استفاده از زبان غیر از پایتون مجاز **نیست**.
۳. پروژه نهایی تحویل خواهد داشت. نکات تحویل متعاقباً اعلام خواهد شد.
۴. گزارش پروژه خود را در یک فایل PDF به صورت دقیق بنویسید.

موفق باشید
گروه حل تمرین