

به نام او

پروژه اول درس مبانی هوش محاسباتی، «شبکه‌های عصبی»



استاد درس: دکتر عبادزاده

زمستان ۱۴۰۳ - دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر



نکاتی در مورد این تمرین نیاز به توجه و دقت دوستان دارد.

- ۱- هرگونه کپی کردن باعث عدم تعلق نمره به تمامی افراد مشارکت کننده در آن می شود.
- ۲- استفاده از چت بات ها در صورت کشف باعث عدم تعلق نمره به پروژه می شود.
- ۳- آخرین مهلت ارسال تمرین، ساعت ۲۳:۵۵ دقیقه روز یکشنبه ۲۶ اسفند می باشد.
- ۴- لطفاً گزارش ارسالی خود را به صورت فایل PDF و به صورت «شماره دانشجویی-PRJ1» مانند PRJ1_40031000 نام گذاری کنید (در صورت عدم رعایت نمره کسر می شود).
- ۵- تمامی کدها نیز به همراه خروجی حاصل شده باید در سامانه بارگذاری شود و در هنگام ارائه نیز همین کد تحویل گرفته خواهد شد.
- ۶- در صورت هرگونه سؤال یا مشکل می توانید با تدریس یاران درس از طریق discussion کانال در ارتباط باشید.

فاز اول: پیاده سازی شبکه عصبی با کتابخانه PyTorch

کتابخانه PyTorch یکی از محبوب ترین کتابخانه در حوزه شبکه عصبی در زبان برنامه نویسی پایتون است. از جمله ویژگی های این کتابخانه نسبت به دیگر کتابخانه ها، شی گرا بودن آن است که روند پیاده سازی را تسهیل می کند. در این بخش قصد داریم با استفاده از این کتابخانه یک شبکه عصبی پرسپترون چندلایه بر روی مجموعه داده گان **Breast Cancer Wisconsin (Diagnostic)** آموزش دهیم.

این مجموعه داده گان متشکل از ۵۶۹ داده مربوط به سرطان سینه است. هر داده در این مجموعه داده گان شامل ۳۰ ویژگی است. هدف این بخش از پروژه آموزش مدلی برای تشخیص برچسب هر داده یعنی "خوش خیم" و یا "بدخیم" است.

برای استفاده از این مجموعه داده گان علاوه بر فایل قرار داده شده در سامانه، می توانید از کتابخانه `ucimlrepo` نیز استفاده کنید.

تسک اول: تشخیص کلاس به کمک پرسپترون چندلایه

مدل پرسپترون چندلایه ای را پیاده سازی کنید.

این شبکه باید `final train accuracy` حداقل ۹۰٪ و `final test accuracy` حداقل ۸۵٪ داشته باشد. فایل نوت بوک این تسک به منظور راحتی شما دوستان عزیز نوشته است و صرفا نیاز است تا آن را کامل کنید.

فاز دوم: پیاده سازی MyTorch

در فاز اول شما با پیاده سازی شبکه های عصبی به کمک کتابخانه PyTorch آشنا شدید. حالا وقتش است که عمیق تر وارد این حوزه شویم و خودمان یک کتابخانه برای انجام این کار پیاده سازی کنیم. در ادامه شما را با بخش های مختلف این پروژه آشنا می کنیم.

کلاس Tensor

این کلاس اساس کار شما می باشد و تمامی محاسبات پایه ای مورد نیاز برای انجام عملیات های مورد نیاز در شبکه عصبی در این کلاس پیاده سازی می شود. یکی از فیلدهای مهم این کلاس `data` می باشد که از نوع `numpy.ndarray` است. به عبارتی شما عملیات ماترسی خود را با این نوع داده انجام می دهید. پیشنهاد می شود

در صورت عدم آشنایی با این کتابخانه پیش از شروع پروژه کمی آن را مطالعه کنید. در فایل `tensor.py` با بررسی کد و خواندن کامنت‌ها بخش‌هایی که کامل نشده‌است را پیاده‌سازی کنید.

کلاس Model

یک کلاس `abstract` برای مدل‌هایی که بعداً تعریف می‌کنید می‌باشد. مدل شما باید در این چارچوب پیاده‌سازی شود.

ماژول Layer

در این بخش لایه `fully-connected` را کامل خواهید کرد. لایه‌ی `fully-connected` یک لایه متراکم می‌باشد که ورودی‌های آن در وزن‌ها ضرب شده و جمع آن‌ها به خروجی می‌رود. شکل زیر ایده این پیاده‌سازی را به شما می‌دهد. برای پیاده‌سازی این قسمت فایل `linear.py` را کامل کنید.

ماژول Activation

در این بخش شما باید توابع فعال‌سازی را پیاده‌سازی کنید. تابع `step` به عنوان نمونه پیاده‌سازی شده است اما سایر توابع نیاز به کامل شدن دارند. توابع `sigmoid`، `relu`، `leaky relu`، `softmax` اجباری می‌باشند و تابع `tanh` امتیازی است.

ماژول Loss

در این بخش شما باید توابع هزینه را پیاده‌سازی کنید. از شما خواسته شده است که دو تابع `MSE` و `CE` را تکمیل کنید.

ماژول Optimizer

در این بخش شما باید بهینه‌سازها را پیاده‌سازی کنید. در فصل اول درس شما با روش گرادیان کاهشی آشنا شده اید و باید پیاده‌سازی آن را در `sgd.py` کامل کنید. دقت داشته باشید شما در هر گام باید پارامترهای مدل را با انجام عملیات ریاضی مربوط به آن الگوریتم بروزرسانی کنید. پیاده‌سازی مابقی بهینه‌سازها امتیازی می‌باشد و برای پیاده‌سازی آن‌ها باید تحقیق کنید.

تسک دوم: آموزش مدل با استفاده از کتابخانه MyTorch

در این تسک شما باید مدل پرسپترون چندلایه خود را که تشخیص نوع سرطان در دیتاست Breast Cancer Wisconsin (Diagnostic) است با استفاده از کتابخانه MyTorch انجام دهید.

این شبکه باید حداقل $\text{final train accuracy}$ ۷۰٪ و $\text{final test accuracy}$ 60٪ داشته باشد. این تسک به همراه خروجی های آن باید در یک نوت بوک با نام task2.ipynb نوشته شده و همراه دیگر کدها آپلود شود. برای نوشتن کد این فایل می توانید از تسک اول کمک بگیرید.

نکات پیاده سازی:

- فایل های پروژه بر روی کورسز آپلود شده اند برای انجام تسک ها به تکمیل قسمت های TODO در نوت بوک و فایل هایی که در اختیار شما قرار داده شده است بپردازید. البته برای تسک دوم همانطور که بیان شد باید یک نوت بوک با نام task2 ایجاد و پیاده سازی کنید. لازم به ذکر است که در صورت نیاز به تغییر سایر بخش ها باید به صورت خلاصه دلیل خود را برای تغییر آن قسمت توضیح دهید.
- پیشنهاد می شود برای اجرای سریع تر از Google Colab استفاده کرده و به T4GPU متصل شوید.
- پیاده سازی شبکه در فاز دوم با استفاده از کتابخانه های کمکی همانند PyTorch و ... هیچ نمره ای ندارد.
- برای دریافت نمره ی هر بخش به خصوص قسمت های امتیازی، تسلط بر چگونگی عملکرد و همچنین پیاده سازی قطعه کد ضروری است.
- چون ممکن است بین تاریخ آپلود و ارائه پروژه فاصله زیادی باشد، پیشنهاد می شود حتما پیش از ارائه، کد خود را مرور کنید.