

## تمرین اول

### هدف: آشنایی با چارچوب کاری تنسورفلو و شبکه‌های عصبی پرسپترونیک چندلایه

در این تمرین قصد داریم، عملکرد شبکه‌های پرسپترونیک چندلایه را در حل مسائل دسته‌بندی<sup>۱</sup> مورد بررسی قرار دهیم. مجموعه داده مورد استفاده در این تمرین، یک مجموعه داده شامل ۴۸۸۴۲ رکورد از وضعیت درآمدی افراد مختلف در طول سال‌های متمادی است. در این مجموعه داده، اطلاعات سرشماری از جمله سن، کلاس کاری، وضعیت تحصیلی، وضعیت تاهل، جنسیت، وضعیت درآمد و خرج، کشور محل سکونت و میزان ساعات کاری افراد تهیه شده است. هدف اصلی در این مساله ارائه مدلی است که با گرفتن اطلاعات سرشماری برای هر فرد، پیش‌بینی نماید آیا این فرد در طول سال بعد قادر به تولید حداقل ۵۰ هزار دلار درآمد می‌باشد یا خیر. برای دریافت مجموعه داده و کسب اطلاعات بیشتر در مورد آن می‌توانید از لینک زیر استفاده نمایید.

<https://archive.ics.uci.edu/ml/datasets/Adult>

مراحل زیر را انجام داده و به سوالات پاسخ دهید.

۱. مجموعه داده فوق را بارگذاری نمایید. تعدادی از ویژگی‌های موجود در مجموعه داده در قالب متغیرهای عددی و تعدادی دیگر در قالب متغیرهای اسمی<sup>۲</sup> ارائه شده‌اند. برای این که بتوانیم این داده‌ها را به عنوان ورودی به یک شبکه عصبی بدهیم لازم است، تمامی داده‌ها تبدیل به مقادیر عددی شوند. با توجه به ساختار شبکه‌های عصبی پرسپترونیک چندلایه، توضیح دهید راهکار شما برای پیش‌پردازش این داده‌ها و تبدیل آن‌ها به قالب مناسب برای ورودی دادن به شبکه چیست؟
۲. داده‌ها را بارگذاری نموده و پیش‌پردازش ارائه شده خود را روی آن‌ها اعمال نمایید.
۳. با استفاده از مدل Sequential در کراس، یک شبکه عصبی پرسپترونیک چندلایه برای دسته‌بندی داده‌های موجود در مجموعه داده ایجاد نمایید. توضیح دهید ساختار شبکه عصبی خود را چگونه طراحی می‌نمایید؟ (منظور از طراحی ساختار شبکه عصبی، انتخاب تعداد لایه‌ها، تعداد نوروها در هر لایه و تابع هزینه نوروها در هر لایه است. یکی از راهکارهای مناسب برای انتخاب مقادیر بهینه برای تعداد لایه‌ها و تعداد نوروها در هر لایه، استفاده از آزمون و خطاست. در صورتی که از آزمون و خطا برای تعیین این مقادیر استفاده کردید، نحوه اجرای آزمایشات و نتایج هر آزمایش را در فایل گزارش خود ذکر نمایید و انتخاب‌های بهینه خود را بر اساس نتایج بدست آمده، ارائه دهید).
۴. تصویر گراف طراحی شده خود را با استفاده از ابزار TensorBoard رسم نموده و به فایل گزارش اضافه نمایید.
۵. مدل ساخته شده خود را با استفاده از بهینه‌ساز Adam و تابع هزینه CrossEntropy کامپایل نمایید. معیارهای Recall، Precision، Accuracy و F1 را نیز به عنوان معیارهای ارزیابی تنظیم نمایید.

<sup>1</sup> Classification

<sup>2</sup> Nominal

۶. مجموعه داده بارگذاری شده را به ۳ قسمت آموزش<sup>۳</sup>، اعتبارسنجی<sup>۴</sup> و آزمون<sup>۵</sup> تقسیم نمایید. سپس شبکه را با استفاده از مجموعه داده های آموزش و اعتبارسنجی، آموزش دهید. تعداد تکرارها<sup>۶</sup> را به حدی در نظر بگیرید که شبکه به طور کامل آموزش ببیند.

۷. مدل آموزش داده شده را با استفاده از تابع `evaluate` در کراس روی مجموعه داده آزمون، ارزیابی نموده و نتایج ارزیابی را در فایل گزارش ذکر نمایید. به علاوه ۵ مورد از داده های موجود در مجموعه داده آزمون را که شبکه پیش بینی می نماید در طول سال بعد بیش از ۵۰ هزار دلار درآمد دارند به همراه ۵ مورد از داده های موجود در مجموعه داده آزمون را که شبکه پیش بینی می نماید در طول سال بعد کمتر از ۵۰ هزار دلار درآمد دارند با استفاده از تابع `predict` در کراس پیدا کرده و در فایل گزارش ذکر نمایید.

۸. فکر می کنید افزایش تعداد لایه های شبکه چه تاثیری بر عملکرد مدل خواهد داشت؟ با طراحی و اجرای یک آزمایش، فرضیه خود را ارزیابی نمایید و نتایج آزمایش را در گزارش ذکر نمایید. (راهنمایی: با استفاده از ابزار `TensorBoard` می توانید میزان تغییرات وزن های هر لایه در مدل را مورد بررسی قرار دهید. پاسخ سوال در درک مفهوم `Vanishing Gradients` نهفته است!)

۹. با توجه به نمودار خطای آموزش و اعتبارسنجی، بگویید آیا مدلی که ارائه داده اید دچار بیش برآزش<sup>۷</sup> شده است؟ اگر خیر، چگونه می توانید این مدل را روی داده های ارائه شده بیش برآزش کنید؟ روش خود را شرح داده و پیاده سازی نمایید. نقطه بیش برآزش روی داده ها را با توجه به نمودار خطای آموزش و اعتبارسنجی در فایل گزارش خود نشان دهید.

۱۰. با توجه به تمام آزمایشات انجام شده و طراحی صورت گرفته، مدلی ارائه دهید که بهترین دقت ممکن را روی مجموعه داده آزمون کسب کند. طراحی این مدل را به همراه دقت کسب شده در فایل گزارش خود ذکر نمایید. (به بهترین دقت کسب شده نمره امتیازی تعلق خواهد گرفت)

---

## نکات مربوط به پیاده سازی و ارزیابی

در تمام پیاده سازی ها و ارزیابی های انجام شده به نکات زیر توجه داشته باشید.

۱. برای ارزیابی عملکرد شبکه از معیارهای `Precision`، `Recall` و `F1` استفاده نمایید. همین طور در هر مرحله از ارزیابی شبکه، حتما ماتریس درهم ریختگی<sup>۸</sup> را نیز رسم نمایید.

---

<sup>3</sup> Train set

<sup>4</sup> Validation set

<sup>5</sup> Test set

<sup>6</sup> epochs

<sup>7</sup> Overfitting

<sup>8</sup> Confusion Matrix

۲. برنامه شما باید قادر به ذخیره‌سازی و بارگذاری بهترین شبکه در هر یک از آزمایش‌ها باشد. در زمان تحویل تمرین باید بتوانید نتایج گزارش شده در هر آزمایش را با بارگذاری شبکه مربوطه که از قبل آموزش دیده و ذخیره شده است، بدون آموزش دوباره آن، به دست بیاورید.

کد: کد این فعالیت را با کمک چارچوب کاری تنسورفلو ۲ و با استفاده از زبان برنامه‌نویسی پایتون و کتابخانه Keras بنویسید.

گزارش: پاسخ سوالات مطرح شده را در یک فایل پی‌دی‌اف بنویسید و فایل پی‌دی‌اف را به همراه کدهای پیاده‌سازی شده در یک فایل زیپ قرار دهید. توجه نمایید در پاسخ به سوالات، تمامی استدلالات خود شامل توضیحات یا نتایج آزمایشات انجام شده را ذکر نمایید.

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

راهنمایی: در صورت نیاز میتوانید سوالات خود را در خصوص پروژه از تدریس یار درس، از طریق ایمیل زیر بپرسید.

E-mail: ann.ceit.aut@gmail.com

ارسال: فایل های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID\_HW01.zip تا تاریخ ۹۹/۰۱/۲۰ ارسال نمایید. شایان ذکر است هر روز تاخیر باعث کسر ۱۰٪ نمره خواهد شد.

---

موفق باشید

احمد اسدی