



دانشگاه اصفهان

تمرین شماره ۲

## یادگیری ماشین در تجارت الکترونیک

موعد تحویل: چهارشنبه ۲۲ آذر ۱۴۰۲

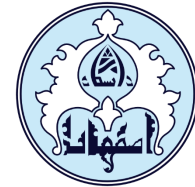
نام استاد: دکتر مرجان کائدی

دستیار حل تمرین: رضا شیری

پاییز ۱۴۰۲

## نکات مورد نیاز دربارهی تمرین

- فایل‌های ارسالی شما باید شما دو بخش کدهای پیاده‌سازی شده و همچنین یک گزارش فنی شامل توضیحات مربوط به تمرین باشد.
- در صورت مشاهده‌ی مشابهت‌های غیر طبیعی، **کسر نمره** به افراد تعلق می‌گیرد.
- در صورت استفاده از یک وبسایت یا منبع خاص، در فایل گزارش خود آدرس وبسایت یا نام منبع را ذکر کنید. در صورت عدم ذکر این موارد و پیدا شدن مشابهت‌های غیر طبیعی، کسر نمره تعلق می‌گیرد.
- شما می‌توانید حداکثر به صورت دو نفره، تمرین را تحویل دهید.
- پاسخ‌های خود را به صورت یک فایل zip به فرمت [name1]\_[family1]\_[name1]\_[family1]\_[std#1]\_[std#2]\_[assignment#] در آورید که در آن [name] نام، [family] نام‌خانوادگی، [assignment#] شماره‌ی تکلیف و [std#] شماره دانشجویی می‌باشد. (به‌طور مثال Reza\_Shiri\_953611133047\_Siavosh\_Djazmi\_993624008\_#1)
- پاسخ‌های خود را تا ساعت ۲۳:۵۹ روز چهارشنبه ۲۲ آذر در قسمت مربوط به تمرین ۱ کلاس تعریف شده در سامانه‌ی کوئرا ارسال کنید.
- مهلت پاسخ‌گویی به این تمرین تا پایان روز چهارشنبه ۲۲ آذر می‌باشد. با توجه به اختلالات مربوط به اینترنت و مشکلات دیگری که برای شما ممکن است پیش بیاید، شما می‌توانید تا **حداکثر ۵ روز** پس از اتمام مهلت ارسال تمرین، پاسخ‌های خود را **بدون کسر نمره**، در سامانه‌ی کوئرا آپلود کنید. توجه داشته باشید که **این زمان به هیچ عنوان قابل تمدید نمی‌باشد**.
- همچنین شما می‌توانید سوالات احتمالی خود را از طریق ایمیل rezamdd1998@gmail.com یا آیدی تلگرامی (@creation\_bug) بپرسید. (توجه داشته باشید این ایمیل، فقط برای پاسخ‌گویی به سوالات مربوط به تمرین است و ارسال تکالیف به این آدرس نمره‌ای را به همراه نخواهد داشت).



## تمرین شماره ۲

نام درس: یادگیری ماشین در

تجارت الکترونیک

نام استاد: دکتر مرجان کائدی

مهلت تحویل: چهارشنبه ۲۲ آذر

مجموعه داده‌ی مورد استفاده در این تمرین، مجموعه داده‌ی **Customer Classification** است که در تمرین قبلی هم مورد استفاده قرار گرفت. این مجموعه داده شامل اطلاعات مربوط به مشتریان (مانند منطقه‌ی زندگی، وضعیت بازنشستگی، وضعیت تأهل، میزان درآمد و...) است که با نام `Telecust1.csv` به همراه تمرین پیوست شده است.

هدف از این تمرین، پیش‌بینی دسته‌ی مشتریان است که نوع دسته‌ی آن‌ها در ستون `custcat` قرار دارد.

هر دو مجموعه داده‌ی ضمیمه شده را بخوانید و در صورت نیاز، پیش‌پردازش‌های مورد نیاز و دلخواه را بر روی آن‌ها انجام دهید. با استفاده از تابع `train_test_split` موجود در کتابخانه‌ی `sklearn`، ۸۰ درصد داده‌ها را به داده‌ی آموزشی و ۲۰ درصد به داده‌ی آزمایشی اختصاص دهید. (تنظیمات مقابل را بر روی آن اعمال کنید. `shuffle=True, random_state=17`)

## K-امین نزدیکترین همسایه

۱- در بخش اول تمرین شما باید روش  $k$ -امین نزدیک‌ترین همسایه ( $KNN$ )<sup>۱</sup> را پیاده‌سازی کنید. برای پیدا کردن فاصله‌ی بین دو داده می‌توانید از فاصله‌ی منتهن<sup>۲</sup> و فاصله‌ی اقلیدسی<sup>۳</sup> استفاده کنید. ماتریس درهم‌ریختگی<sup>۴</sup> و `classification_report` داده‌های آموزشی و آزمایشی را برای مقادیر مختلف  $k$  ( $k=1, 3, 5, 7$ ) گزارش و تحلیل کنید. (آزمایشات را با استفاده از هر دو معیار فاصله انجام دهید.)

## شبکه‌ی عصبی چندلایه

در این بخش شما باید یک شبکه‌ی عصبی عمیق چندلایه ( $MLP$ )<sup>۵</sup> پیاده‌سازی کنید. برای این کار می‌توانید از چارچوب‌های پایتورچ<sup>۶</sup> و تنسورفلو<sup>۷</sup> کمک بگیرید. (پیشنهاد می‌شود از چارچوب پایتورچ برای پیاده‌سازی این بخش کمک بگیرید.)

شما می‌توانید برای آشنایی با نحوه‌ی پیاده‌سازی شبکه‌های عمیق در چارچوب پایتورچ از [این لینک](#) و چارچوب تنسورفلو از [این لینک](#) استفاده کنید.

در تمام بخش‌های بعدی شما باید نمودار درصد موفقیت پیش‌بینی مدل در دوره‌های مختلف را برای مجموعه داده‌های آموزشی و اعتبارسنجی رسم کنید. ماتریس درهم‌ریختگی و `classification_report` داده‌های آموزشی، اعتبارسنجی و آزمایشی را در انتهای آموزش مدل گزارش کنید.

<sup>1</sup> K-Nearest Neighbor

<sup>2</sup> Manhattan distance

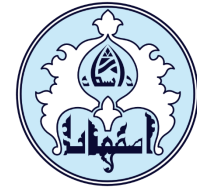
<sup>3</sup> Euclidian distance

<sup>4</sup> Confusion matrix

<sup>5</sup> Multi-Layer Perceptron

<sup>6</sup> PyTorch

<sup>7</sup> TensorFlow



## تمرین شماره ۲

نام درس: یادگیری ماشین در

تجارت الکترونیک

نام استاد: دکتر مرجان کائدی

مهلت تحویل: چهارشنبه ۲۲ آذر

۲- ۲۰ درصد داده‌های آموزشی را برای اعتبارسنجی کنار بگذارید و از بقیه‌ی داده‌ها به‌عنوان داده‌های آموزشی استفاده کنید. یک شبکه‌ی عصبی چندلایه با دو لایه‌ی پنهان، ۲۵۶ نورون و تابع فعال‌سازی<sup>۱</sup> ReLU<sup>۲</sup> بسازید و آن را با آموزش دهید. از بهینه‌گر Adam با نرخ یادگیری<sup>۳</sup> ۰.۰۰۱ برای یادگیری مدل استفاده کنید، تعداد دور<sup>۴</sup>های مورد نیاز برای یادگیری مدل را ۲۰ در نظر بگیرید و از تابع هزینه‌ی CrossEntropy برای آموزش کمک بگیرید. اندازه‌ی دسته<sup>۵</sup>ها در تکرار<sup>۶</sup>های مختلف را برابر ۳۲ در نظر بگیرید.

### هدف قسمت‌های بعد، تحلیل پارامترهای مختلف است.

در هر قسمت، پس از گزارش نتایج، نتیجه‌ی تغییر پارامترها را تحلیل کنید.

### تنظیمات مدل

- ۳- مدلهایی با تعداد لایه‌های مختلف (۲، ۳، ۴، ۵، ۶) بسازید و با تنظیمات قبلی آن‌ها را آموزش دهید.
- ۴- بهترین مدل از بین مدل‌های مرحله‌ی «۳» را انتخاب کنید و آن را با تعداد نورون‌های مختلف (۳۲، ۶۴، ۱۲۸، ۲۵۶، ۵۱۲) و تنظیمات قبلی آموزش دهید.
- ۵- بهترین مدل از بین مدل‌های مرحله‌ی «۴» را انتخاب کنید و آن را با توابع فعال‌سازی مختلف (ReLU، Tanh، Leaky ReLU، Sigmoid) و تنظیمات قبلی آموزش دهید.

### بهینه‌گر

در یادگیری عمیق، عملیات یادگیری وزن‌ها توسط بهینه‌گر<sup>۷</sup>ها انجام می‌شود. بهینه‌گرها با محاسبه‌ی مشتقات و محاسبات مربوط به آن، عملیات پس‌انتشار<sup>۸</sup> خطا را انجام می‌دهند. در این بین بهینه‌گرهای مختلفی وجود دارند که سرعت عملیات آموزش را تغییر می‌دهند و مسیر حرکت آن‌ها به سمت وزن‌های بهینه‌ی شبکه یکسان نیست. در میان این بهینه‌گرها، دو بهینه‌گر معروف به نام‌های Adam و SGD وجود دارند که برای سوال‌های بعد از آن‌ها استفاده می‌شود.

مدل بهینه از مراحل قبل را برای مراحل بعدی انتخاب کنید.

- ۶- مدل ساخته شده را با نرخ‌های یادگیری<sup>۳</sup> ( $10^{-2}$ ،  $10^{-3}$ ،  $10^{-4}$ ،  $10^{-5}$ ،  $10^{-6}$ ) برای بهینه‌گر Adam آموزش دهید.
- ۷- مدل ساخته شده را با نرخ‌های یادگیری<sup>۳</sup> ( $10^{-1}$ ،  $10^{-2}$ ،  $10^{-3}$ ،  $10^{-4}$ ) برای بهینه‌گر SGD آموزش دهید.
- ۸- دو بهینه‌گر ذکر شده را با توجه به نتایج به‌دست آمده با یکدیگر مقایسه کنید.

<sup>۱</sup> Activation function

<sup>۲</sup> Rectified Linear Unit

<sup>۳</sup> Learning rate

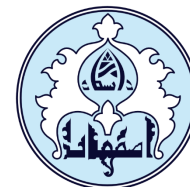
<sup>۴</sup> Epoch

<sup>۵</sup> Batch size

<sup>۶</sup> iteration

<sup>۷</sup> Optimizer

<sup>۸</sup> Back propagation



دانشکده‌ی مهندسی کامپیوتر  
گروه فناوری اطلاعات

## تمرین شماره ۲

نام درس: یادگیری ماشین در

تجارت الکترونیک

نام استاد: دکتر مرجان کائدی

مهلت تحویل: چهارشنبه ۲۲ آذر

### تنظیمات آموزش

بهترین تنظیمات بهینه‌گر از مراحل قبل را برای مراحل بعدی انتخاب کنید.

۹- مدل ساخته شده را با اندازه دسته‌ی (۱، ۴، ۸، ۳۲، ۶۴، ۱۲۸) آموزش دهید.

۱۰- مدل ساخته شده را با تعداد دور متفاوت (۵، ۱۰، ۲۰، ۵۰، ۱۰۰) آموزش دهید.

۱۱- به نظر شما رابطه‌ای بین اندازه‌ی دسته و تعداد دورهای آموزش مدل وجود دارد؟ برای پاسخ خود استدلال بیاورید.

۱۲- (این مرحله اختیاری است و می‌تواند شامل نمره‌ی اضافی برای شما باشد.) شما می‌توانید آزمایش‌های بیشتری با تنظیمات

مختلف (مانند پیش‌پردازش متفاوت، یکسان نبودن تعداد نورون‌های لایه‌های مختلف، استفاده از دیگر بهینه‌گرها، نرخ یادگیری‌های

متفاوت و...) انجام دهید و نتایج آن‌ها را گزارش کنید و نتایج مدل خود را بهبود بخشید.

**به بالاترین درصد موفقیت در بین نتایج مدل‌های دانشجویان نمره‌ی اضافی تعلق می‌گیرد.**