به نام خدا

دانشکده سامانه‌های هوشمند

مبانی محاسبات (رایانش) نرم (۰۱-۰۳۹-۰۵-۸۳)

نیم‌سال اول ۱۴۰۵-۱۴۰۴

تمرین شماره ۱

تاریخ تحویل: ۱۴۰۴/۰۸/۲۱

---

۱. (۱۵٪) [**مطالعه و بررسی**] توابع فعال‌سازی (Activation functions) زیر را که بر مبنای تاریخ توسعه مرتب شده‌اند، در نظر بگیرید:

- Sigmoid
- Tanh
- Parametric Leaky Tanh (PLTanh)
- Softmax
- Rectified Linear Unit (ReLU)
- Leaky ReLU
- PReLU
- Exponential Linear Unit (ELU)
- Trivialized Exponential Linear Unit (TeLU)
- Scaled Exponential Linear Unit (SeLU)
- Maxout
- Gelu
- Swish
- Mish

۱-۱ نمودار هر یک از آنها را به ازای متغیر x در بازه $x \in [-5, 5]$ رسم کنید.

۱-۲ تابع گرادیان هر کدام از توابع را نسبت به متغیر x در بازه $x \in [-5, 5]$ بدست آورید و آنها را رسم کنید.

۱-۳ کدام یک از توابع فعال‌سازی بالا مناسب برای مسائل دوکلاسه (Binary Classification) و کدام یک مناسب برای مسائل چندکلاسه (Multi Class Classification) مناسب هستند؟

---

۲. (۴۵٪) [**پیاده‌سازی: پرسپترون و آدالاین**] مسئلهٔ بازشناسی نویسه را برای الگوهای بیان شده در مثال ۲-۱۵ در فصل دوم کتاب، با استفاده از ساختار شبکه نشان داده شده در شکل ۲-۲۱ پیاده‌سازی

h.veisi@ut.ac.ir

به نام خدا

کنید. بدین منظور الگوهای آموزش شکل ۲-۲۰ و الگوهای آزمون (تست) شکل ۲-۲۲ که به صورت فایل‌های متنی به همراه تمرین ارائه شده است، به کار بگیرید. برای موارد زیر نتیجه را گزارش کنید. در گزارش‌ها درصد خطای بازشناسی را به صورت زیر محاسبه و گزارش کنید.

$$Error\ Rate = \frac{N^{err}}{N} \times 100 = \frac{\text{تعداد الگوهایی که اشتباهی بازشناسی شده اند}}{\text{تعداد کل الگوها}} \times 100$$

نکته: برای بهتر دیدن الگوهای ارائه شده در فایل‌های متنی، می‌توانید از فونت Courier New یا CourierPS در NotePad استفاده کنید.

۲-۱ (۲۰٪) [**پرسپترون**] آموزش شبکه را با استفاده از قانون پرسپترون و با در نظر گرفتن موارد خواسته شده در بخش‌های زیر انجام دهید. توجه شود در هر بخش جدید، بهترین پارامترهای پیدا شده در بخش‌های پیشین را به عنوان مقادیر پیش فرض در نظر بگیرید:

الف) [**مقادیر اولیه**] در این بخش قصد داریم درصد خطای بازشناسی را برای داده‌های آزمون، به ازای اجراهای مختلفی از مقادیر مختلف اولیه وزن‌ها و بایاس‌های شبکه پیدا کنیم. در این راستا این مقادیر را با روش های زیر تولید کرده و سپس درصد خطای بازشناسی و زمان همگرایی را برای داده‌های آموزش گزارش کنید. با استفاده از نتایج، در مورد اثر مقادیر اولیه بر کارایی و سرعت شبکه بحث کنید. روش‌های تولید وزن اولیه:

۱- همه صفر
۲- مقادیر تصادفی با توزیع یکنواخت
۳- مقادیر تصادفی با توزیع نرمال
۴- روش Xavier(Glorot)
۵- روش Kaiming(HE)

ب) [**مقدار آستانه**] برنامه را با بهترین روش بخش الف، برای چند مقدار مختلف آستانهٔ θ تکرار کنید و نتایج بازشناسی را برای داده‌های آزمون ارائه کنید. آیا مقدار بیشتر θ تأثیری بر درصد خطای شبکه دارد؟

h.veisi@ut.ac.ir

ج) **[نرخ یادگیری]** شبکه را با بهترین روش بخش الف و ب، برای پنج مقدار مختلف از نرخ یادگیری، ۰.۵، ۰.۱، ۰.۰۵ و ۰.۰۱ آموزش دهید و نتیجه بازشناسی بر روی داده‌های آزمون را به صورت نمودار رسم کنید. زمان همگرایی شبکه را در هر حالت گزارش کنید. با توجه به نتایج حاصل، در مورد تاثیر این پارامتر بر عملکرد شبکه بحث کنید.

۲-۲ (۱۵٪) **[آدالاین]** شبکه را با استفاده از قانون آدالاین (دلتا) و با بهترین پارامترهای پیدا شده در قسمت‌های قبل سوال (نرخ یادگیری، وزن‌های اولیه و بایاس) آموزش دهید و نتیجه بازشناسی را برای داده‌های آزمون و داده‌های آموزش گزارش کنید.

۲-۳ (۲.۵٪) **[مقایسه]** بهترین نتایج بدست آمده را روی هر دو مجموعه آموزش و آزمون در ۳ پیاده‌سازی مربوط به پرسپترون و پیاده‌سازی مربوط به آدالاین، در یک جدول بیاورید. نتایج را مقایسه کرده و علل تفاوت نتایج را شرح دهید.

۲-۴ (۲.۵٪) **[داده نویزی]** برای حالتی که برای الگوی تست، بیش از یکی از دسته‌ها انتخاب می‌شوند، چه راهکاری را پیشنهاد می‌کنید. روش خود را پیاده‌سازی کرده و نتیجه را برای شبکه پرسپترون گزارش کنید.

۲-۵ (۵٪) **[استخراج ویژگی]** در بخش‌های قبل از این تمرین، مقدار کل پیکسل‌های هر نویسه به عنوان ورودی شبکه استفاده شد. در این بخش، از روش تصویر کردن (projection) برای استخراج ویژگی استفاده می‌شود و مقدار ویژگی‌ها (به جای مقادیر پیکسل‌ها) به عنوان ورودی به شبکه داده می‌شود. در این روش، به ازای هر ردیف (و هر ستون) از هر نویسه، مجموع پیکسل‌های روشن (با مقدار یک) آن ردیف (یا ستون) شمارش شده و مقدار حاصل‌جمع به عنوان ویژگی در نظر گرفته می‌شود. با توجه به ابعاد نویسه‌ها که ۹×۷ هستند، تعداد ویژگی‌های هر نویسه ۹+۷=۱۶ خواهد بود.

شبکه آدالاین را برای آموزش با ویژگی‌های حاصل تغییر دهید و نتایج حاصل را (نرخ خطا روی مجموعه آزمون) در یک جدول گزارش کنید. آیا کارایی شبکه‌ها در این حالت، به نسبت حالتی که مقدار خود پیکسل‌ها استفاده شود، بهبود می‌یابد یا خیر؟ در هر صورت، تحلیل خود را از این نتایج بیان کنید.

h.veisi@ut.ac.ir

۳. (۴۰٪) [**پیاده‌سازی: تحلیل ریزش مشتری با شبکه عصبی**] در بسیاری از کسب‌وکارهای مبتنی بر اشتراک، مانند شرکت‌های مخابراتی، سرویس‌های پخش آنلاین، بانک‌ها و پلتفرم‌های محتوایی، مشتریان ممکن است در مقاطع زمانی مختلف تصمیم بگیرند استفاده از خدمات را متوقف کنند. این پدیده در علم داده و بازاریابی با عنوان ریزش (Churn) یا Customer Attrition شناخته می‌شود و بیانگر از دست دادن مشتریان در طول زمان است. شناسایی و پیش‌بینی مشتریانی که در آستانه‌ی ترک سرویس هستند، یکی از چالش‌های اصلی تحلیل رفتار کاربران محسوب می‌شود؛ زیرا نگه‌داشت مشتریان موجود معمولاً از جذب مشتریان جدید مقرون‌به‌صرفه‌تر است. در یادگیری ماشین، مسئله‌ی پیش‌بینی ریزش به‌صورت یک مسئله طبقه‌بندی دودویی مدل می‌شود، که در آن هدف، پیش‌بینی برچسب ۱ برای کاربران ترک‌کننده و برچسب ۰ برای کاربران وفادار است. این پیش‌بینی به شرکت‌ها کمک می‌کند تا با طراحی راهکارهای شخصی‌سازی‌شده و مداخلات هدفمند، احتمال از دست دادن کاربران را کاهش دهند.

اسپاتیفای (Spotify) یکی از بزرگ‌ترین سرویس‌های پخش موسیقی آنلاین در جهان است که کاربران می‌توانند از آن برای گوش دادن به موسیقی، ساخت پلی‌لیست، و دنبال‌کردن هنرمندان استفاده کنند. ز آنجا که کاربران به‌راحتی می‌توانند اشتراک خود را لغو کنند، پیش‌بینی احتمال ریزش برای شرکت بسیار مهم است.تحلیل داده‌های استفاده کاربران (مثل زمان گوش دادن، دفعات ورود، تعداد آهنگ ردشده، و غیره) می‌تواند الگوهای رفتاری منجر به ریزش را آشکار کند.

دادگان Spotify Churn Analysis شامل اطلاعات حدود ۸۰۰۰ کاربر سرویس پخش موسیقی اسپاتیفای است و برای تحلیل و پیش‌بینی رفتار ترک سرویس طراحی شده است. هر ردیف در این مجموعه‌داده بیانگر ویژگی‌های فردی و رفتاری یک کاربر است. متغیرها شامل مشخصات جمعیت‌شناختی مانند جنسیت (gender)، سن (age) و کشور (country)، نوع اشتراک (subscription_type) و الگوهای استفاده از سرویس مانند زمان گوش دادن روزانه (listening_time)، تعداد آهنگ‌های پخش‌شده در (songs_played_per_day)، نرخ رد کردن آهنگ‌ها (skip_rate)، نوع دستگاه مورد استفاده (device_type) و تعداد تبلیغات شنیده‌شده در هفته (ads_listened_per_week) هستند. همچنین ستون offline_listening نشان می‌دهد که آیا کاربر از

h.veisi@ut.ac.ir

حالت آفلاین استفاده کرده است یا خیر. متغیر هدف (is_churned) به‌صورت دودویی تعریف شده است، که مقدار ۱ نشان‌دهنده‌ی کاربران ترک‌کننده و مقدار ۰ نشان‌دهنده‌ی کاربران فعال است. این مجموعه‌داده ترکیبی از ویژگی‌های عددی و رده‌ای (numeric و categorical) دارد و برای ساخت مدل‌های یادگیری ماشین در زمینه‌ی پیش‌بینی ترک کاربر، تحلیل الگوهای تعامل، و بهبود راهبردهای حفظ مشتریان اسپاتیفای استفاده قرار می‌گیرد.

۳-۱ (۲.۵٪) [تحلیل مقدماتی داده] در نخستین گام، لازم است ساختار کلی مجموعه‌داده بررسی شود. دانشجو باید با بارگذاری فایل داده، ابعاد آن (تعداد سطر و ستون)، نام ستون‌ها و نوع داده‌ی هر ویژگی را مشاهده کند تا درک اولیه‌ای از ماهیت داده به‌دست آید. سپس با استفاده از توابع آماری نظیر describe() و info() شاخص‌هایی مانند میانگین، انحراف معیار، بیشینه و کمینه‌ی متغیرها استخراج می‌شود تا رفتار عددی آن‌ها مشخص گردد. پس از آن، باید وجود مقادیر گمشده (NaN) یا مقادیر پرت (outliers) شناسایی شود. در پایان، توزیع مقادیر در متغیر هدف is_churned بررسی می‌شود تا مشخص گردد آیا داده نامتوازن (imbalanced) است یا خیر. در صورت وجود نامتوازنی، دانشجو باید توضیح دهد که این پدیده چگونه می‌تواند باعث سوگیری مدل به سمت کلاس اکثریت شود و چه روش‌هایی برای مقابله با آن وجود دارد؟

۳-۲ (۲.۵٪) [تبدیل ویژگی‌ها] از آنجا که بسیاری از مدل‌های یادگیری ماشین از جمله پرسپترون و آدلاین فقط با داده‌های عددی کار می‌کنند، لازم است ویژگی‌های غیرعددی (categorical) مانند gender, country, subscription_type و device_type به داده‌های عددی تبدیل شوند. این تبدیل را می‌توان با دو روش انجام داد: Label Encoding که به هر دسته یک عدد یکتا اختصاص می‌دهد، و One-Hot Encoding که برای هر مقدار ممکن از دسته، یک ستون جداگانه ایجاد می‌کند. انتخاب روش مناسب باید با توجه به ماهیت داده انجام شود؛ برای مثال، اگر متغیر دارای ترتیب طبیعی نباشد (مثل نوع دستگاه یا کشور)، One-Hot Encoding مناسب‌تر است. در پایان این مرحله، تمامی ستون‌های داده باید عددی و قابل استفاده برای مدل‌سازی باشند.

۳-۳ (۲.۵٪) [بررسی همبستگی ویژگی‌ها] پیش از مدل‌سازی، لازم است بررسی شود کدام ویژگی‌ها با یکدیگر یا با متغیر هدف همبستگی دارند. ابتدا ستون user_id حذف می‌شود، زیرا

h.veisi@ut.ac.ir

شناسه‌ای یکتا است و هیچ ارزش تحلیلی ندارد. سپس با استفاده از تابع corr() ماتریس همبستگی میان ویژگی‌های عددی محاسبه می‌شود و به کمک نمودار حرارتی (heatmap) نمایش داده می‌شود. این ماتریس نشان می‌دهد که کدام متغیرها ارتباط مثبت یا منفی قوی با یکدیگر یا با متغیر هدف دارند. دانشجو باید تحلیل کند آیا برخی ویژگی‌ها دارای همبستگی بالا هستند (که می‌تواند منجر به افزونگی اطلاعات شود) و کدام ویژگی‌ها بیشترین تأثیر را بر متغیر is_churned دارند.

۳-۴ (۲.۵٪) [رفع نامتوازنی داده] در صورتی که کلاس‌های داده نامتوازن باشند، مدل تمایل خواهد داشت کلاس اکثریت را بیشتر پیش‌بینی کند. برای جلوگیری از این مسئله، در این تمرین از روش Random Undersampling استفاده می‌شود. در این روش، نمونه‌هایی از کلاس بزرگ‌تر به‌صورت تصادفی حذف می‌شوند تا تعداد نمونه‌های دو کلاس برابر شود. این کار موجب می‌شود مدل بدون سوگیری نسبت به هیچ کلاس خاصی آموزش ببیند. هرچند Undersampling موجب از دست رفتن بخشی از داده می‌شود، اما برای تمرین‌های آموزشی و مدل‌های سبک مانند Adaline و Perceptron گزینه‌ای ساده و مؤثر است. پس از اجرای این روش، دانشجو باید نسبت کلاس‌ها را قبل و بعد از بالانس بررسی و درباره تأثیر آن بر عملکرد مدل توضیح دهد.

۳-۵ [تقسیم داده و نرمال‌سازی] در این مرحله داده‌ی متعادل‌شده به دو بخش آموزش (train) و آزمون (test) تقسیم می‌شود تا عملکرد مدل بتواند روی داده‌های دیده‌نشده سنجیده شود. نسبت پیشنهادی تقسیم ۸۰ درصد برای آموزش و ۲۰ درصد برای آزمون است. سپس ویژگی‌ها باید استاندارد شوند تا میانگین صفر و انحراف معیار یک داشته باشند؛ این فرآیند با استفاده از StandardScaler نجام می‌شود.

۳-۶ (۳۰٪) [آموزش و مقایسه مدل‌ها] در گام نهایی، دو مدل یادگیری خطی کلاسیک، یعنی Perceptron و Adaline روی داده‌ی آماده‌شده آموزش داده می‌شوند. پرسپترون از تابع فعال‌سازی پله‌ای استفاده می‌کند و بر اساس قانون یادگیری خطی ساده عمل می‌کند، در حالی که آدلاین از تابع هزینه میانگین مربعات و به‌روزرسانی تدریجی وزن‌ها بهره می‌گیرد. هر دو مدل پس از آموزش باید روی داده‌ی آزمون ارزیابی شوند. معیارهای ارزیابی شامل Accuracy دقت کلی، Precision دقت پیش‌بینی‌های کلاس مثبت، Recall توانایی مدل در شناسایی کاربران ترک‌کرده، F1-score (میانگین

h.veisi@ut.ac.ir

موزون دقت و یادآوری)، و Confusion Matrix برای مشاهده‌ی توزیع خطاهاست. در پایان، دانشجو باید نتایج دو مدل را مقایسه کرده و درباره‌ی تفاوت در همگرایی، حساسیت به نویز، و پایداری هر مدل تحلیل کوتاهی ارائه دهد.

h.veisi@ut.ac.ir