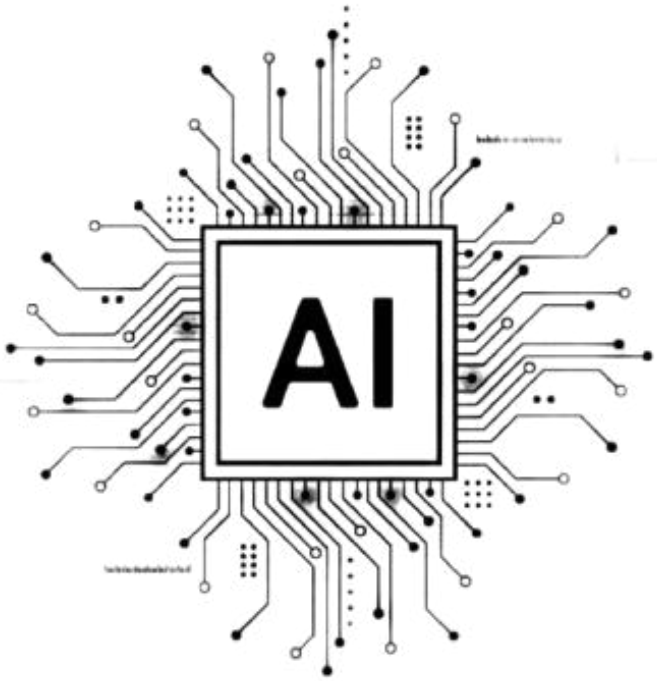
بِسْمِ اللهِ الرَّحْمَنِ الرَّحِيمِ

دفترچه سوالات مرحله دوم المپیاد هوش مصنوعی از ابتدا تاکنون

(همراه با کلید)



amoozz.ir

sampaadia.ir

- برای هدایت به صفحه راهکارهای مؤثر برای موفقیت در المپیاد: از برنامه‌ریزی تا مدیریت زمان بر روی این متن کلیک کنید

- برای هدایت به صفحه امتیاز و تسهیلات کسب مدال در المپیادهای علمی دانش آموزی چیست؟ بر روی این متن کلیک کنید

- برای هدلیت‌به صفحه نمره کف قبولی المپیاد چیست؟ بر روی این متن کلیک کنید

سایر مطالب مرتبط:

- آشنایی با المپیاد بین‌المللی هوش مصنوعی IOAI
- منابع و مراجع المپیاد هوش مصنوعی

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

باسمه تعالی
جمهوری اسلامی ایران
وزارت آموزش و پرورش
باشگاه دانش‌پژوهان جوان

مبارزهٔ علمی برای جوانان، زنده‌کردن روح جست‌وجو و کشف واقعیّت‌هاست. «امام خمینی (ره)»

دفترچهٔ سؤالات مرحلهٔ دوم

اولین دورهٔ المپیاد هوش مصنوعی(چند گزینه‌ای)

سال تحصیلی ۱۴۰۴-۱۴۰۳

تاریخ : ۱۴۰۴/۱/۲۸ - ساعت : ۱۴:۰۰ - مدت : ۱۲۰ دقیقه - نوع : چندگزینه‌ای

توضیحات مهم

استفاده از هرنوع ماشین‌حساب ممنوع است.

۱- بلافاصله پس از آغاز آزمون، تعداد سؤالات داخل دفترچه و همهٔ برگه‌های دفترچهٔ سؤالات را بررسی نمایید. در صورت هرگونه نقص در دفترچه، در اسرع وقت مسئول جلسه را مطّلع کنید.

۲- یک برگ پاسخ‌برگ در اختیار شما قرار گرفته که مشخّصات شما بر روی آن نوشته شده است. در صورت نادرست بودن آن، در اسرع وقت مسئول جلسه را مطّلع کنید. ضمناً مشخّصات خواسته‌شده در پایین پاسخ‌برگ را با مداد مشکی بنویسید.

۳- پاسخ‌برگ توسط دستگاه تصحیح می‌شود؛ پس آن را تا نکنید و تمیز نگه دارید و به علاوه، پاسخ هر پرسش را با مداد مشکی نرم در محلّ مربوط علامت بزنید. لطفاً خانهٔ مورد نظر را کاملاً سیاه کنید.

٤- دفترچهٔ سؤال باید همراه پاسخ‌برگ تحویل داده شود.

٥- پاسخ درست به هر پرسش ۳ نمرهٔ مثبت و پاسخ نادرست ۱ نمرهٔ منفی دارد.

٦- شرکت‌کنندگان در دورهٔ تابستانی از بین دانش‌آموزان دهم و یازدهم انتخاب می‌شوند.

۷- از مخدوش کردن بارکدها و مربع‌ها در چهارگوشهٔ صفحه در دفترچهٔ پاسخ‌برگ جداً خودداری کنید. در غیر این صورت برگهٔ شما تصحیح نخواهد شد.

۸- همراه داشتن هر گونه کتاب، جزوه، یادداشت و لوازم الکترونیکی نظیر تلفن همراه، ساعت هوشمند، دستبند هوشمند و لپتاپ ممنوع است. همراه داشتن این قبیل وسایل حتی اگر از آن استفاده نکنید یا خاموش باشد تقلب محسوب خواهد شد.

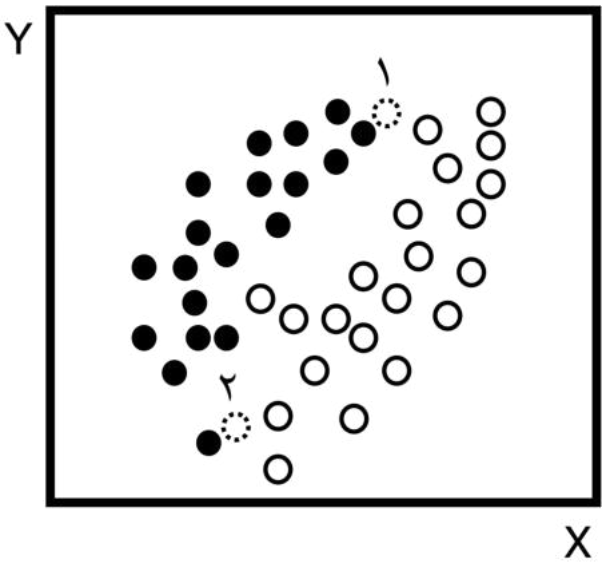
۹- این دفترچه شامل ۱۵ سوال و با احتساب جلد ۷ برگ است.

۱۰- آزمون تشریحی بلافاصله بعد از اتمام این آزمون شروع می‌شود.

کلّیهٔ حقوق این سؤالات برای باشگاه دانش‌پژوهان جوان محفوظ است.
آدرس پایگاه اینترنتی: ysc.medu.gov.ir

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

۱ شکل ۱ داده‌های مساله‌ی دسته‌بندی (Classification) به دو دسته‌ی سیاه و سفید را نشان می‌دهد. داده‌هایی که با خط چین مشخص شده‌اند داده‌هایی هستند که هنوز دسته‌ی آن‌ها مشخص نشده است. در صورتی که از دسته‌بند KNN (K-Nearest Neighbours) با $k = ۳$ یا $k = ۱$ برای دسته‌بندی استفاده کنیم، داده‌های مشخص شده‌ی ۱ و ۲ به کدام دسته‌ها اختصاص پیدا می‌کنند؟



شکل ۱: داده‌های سوال دسته‌بندی KNN

۱) برای $k = ۱$ داده‌ی ۱ سیاه و داده‌ی ۲ سیاه. برای $k = ۳$ داده‌ی ۱ سیاه و داده‌ی ۲ سفید.
۲) برای $k = ۱$ داده‌ی ۱ سیاه و داده‌ی ۲ سیاه. برای $k = ۳$ داده‌ی ۱ سیاه و داده‌ی ۲ سیاه.
۳) برای $k = ۱$ داده‌ی ۱ سیاه و داده‌ی ۲ سیاه. برای $k = ۳$ داده‌ی ۱ سفید و داده‌ی ۲ سفید.
۴) برای $k = ۱$ داده‌ی ۱ سفید و داده‌ی ۲ سفید. برای $k = ۳$ داده‌ی ۱ سیاه و داده‌ی ۲ سفید.

۲ فضای دو بعدی $(x_۱, x_۲)$ را در نظر بگیرید. برای حل مسئله‌ی دسته‌بندی (Classification) در این فضا از درخت تصمیم (Decision Tree) استفاده می‌شود. هر گره در این درخت تصمیم یک شرط به صورت $x_i < C$ را مورد بررسی قرار می‌دهد که x_i یکی از ویژگی‌های $x_۱$ یا $x_۲$ و C یک مقدار ثابت برای این گره است. پیچیدگی درخت را تعداد گره‌های درخت در نظر بگیرید. همچنین فرض کنید که درخت حاضر از بین درخت‌هایی که عملکرد مشابه دارند کمترین پیچیدگی را دارد.

دو تبدیل زیر را در نظر بگیرید:

۱. داده‌های ورودی را بهنجار (Normalize) کنیم. به این معنی که داده‌ها میانگین صفر و واریانس واحد داشته باشند.
۲. داده‌های ورودی ۴۵ درجه حول مبدا مختصات دوران دهیم.

فرض کنید پس از اعمال هر کدام از تبدیلات بالا یک درخت تصمیم دیگر برای دسته‌بندی روی داده‌های جدید آموزش ببیند. کدام یک از گزاره‌های زیر درباره‌ی این درخت آموزش داده‌ی شده‌ی جدید درست است؟

۱) در صورت اعمال تبدیل اول، با یک درخت با پیچیدگی کمتر می‌توانیم به عملکرد درخت اولیه دست پیدا کنیم.
۲) در صورت اعمال تبدیل دوم، نمی‌توان درباره‌ی این که درخت دوم پیچیده‌تر یا ساده‌تر از درخت اولیه است نظر

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

قطعی داد.
۳) در صورت اعمال تبدیل دوم، با درختی با پیچیدگی کمتر می‌توانیم به عملکرد درخت اولیه برسیم.
۴) در صورت اعمال تبدیل اول، نمی‌توان درباره‌ی این‌که درخت دوم پیچیده‌تر یا ساده‌تر از درخت اولیه است نظر قطعی داد.

۳ یک شرکت تماشای فیلم آنلاین می‌خواهد یک مدل یادگیری ماشین برای پیش‌بینی احتمال لغو اشتراک مشترکان خود آموزش دهد. هدف این مدل شناسایی آن دسته از مشترکانی که احتمالا اشتراکشان را تمدید نمی‌کنند (که نمونه‌های مثبت نامیده می‌شوند). در این صورت شرکت می‌تواند با ارائه‌ی پیشنهادها و تخفیف‌های ویژه به آن‌ها از لغو اشتراکشان جلوگیری کند.

برای این کار شرکت دو مدل یادگیری ماشین ۱ و ۲ را طراحی کرده است که عملکردی مشابه دارند اما نمودار واسنجی (Calibration Curve) آن‌ها متفاوت است. این نمودارها در شکل ۲ آمده است. با توجه به این نمودارها در عمل شرکت استفاده از کدام مدل را ترجیح می‌دهد؟

توضیحات: برای کشیدن نمودار واسنجی به صورت زیر عمل می‌شود:

۱. ابتدا کاربران را بر اساس احتمال پیش‌بینی شده برای لغو اشتراک آن‌ها توسط مدل از کم به زیاد مرتب می‌کنیم.
۲. سپس آن‌ها را به K گروه تقسیم می‌کنیم.
۳. در هر گروه نسبت تعداد کاربرانی که اشتراکشان را لغو کردند به کل کاربران را حساب می‌کنیم. این نسبت در محور افقی نمایش داده شده است.

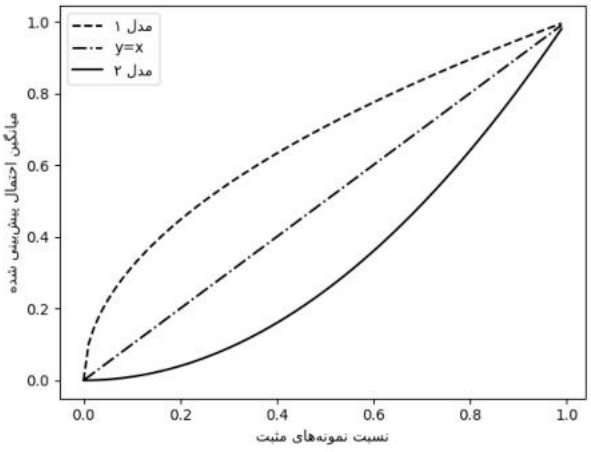
$$\forall i \in [۱, K]: \quad x_i = \frac{\text{تعداد کاربرانی در گروه i که اشتراکشان را لغو کردند}}{\text{تعداد کل کاربران گروه i}} \qquad (۱)$$

۴. در هر گروه میانگین پیش‌بینی احتمال لغو اشتراک توسط کاربران را توسط مدل را حساب می‌کنیم. این نسبت در محور عمودی نمایش داده شده است.

$$\forall i \in [۱, K]: \quad y_i = \frac{\sum_{s \in \text{i گروه}} p_s}{\text{تعداد کل کاربران گروه i}} \qquad (۲)$$

در رابطه‌ی بالا p_s احتمال پیش‌بینی شده برای لغو اشتراک کاربر sام است.

۵. در نهایت مجموعه نقاط بدست آمده $\{(x_۱, y_۱), (x_۲, y_۲), ..., (x_k, y_k)\}$ را روی نمودار نمایش می‌دهیم.



شکل ۲: نمودار واسنجی مربوط به مدل‌های پیش‌بینی احتمال خروج مشتریان

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

۱) مدل ۱
۲) مدل ۲
۳) این که کدام مدل را انتخاب کند وابسته به شرایط کسب‌وکار است.
۴) هر دو مدل انتخاب خوبی هستند چرا که مادامی که عملکرد مشابهی داشته باشند نمودار واسنجی آن‌ها اهمیتی ندارد.

۴ وزارت آموزش و پرورش قصد پایش وضعیت دانش آموزان در فضای مجازی را دارد و می‌خواهد آن‌ها را از نظر رفتاری به گروه‌هایی تقسیم کند به طوری که دانش آموزان هر گروه ویژگی‌های رفتاری شبیه به یکدیگر و متفاوت از ویژگی‌های رفتاری گروه‌های دیگر داشته باشند. برای این کار آن‌ها داده‌های حساب اجتماعی تعداد زیادی از دانش‌آموزان را جمع‌آوری کرده و آن را به یک متخصص هوش مصنوعی دادند. این متخصص ابتدا برای هر دانش آموز از روی این داده‌ها یک بردار بازنمایی (Embedding) بدست آورد. این بردار به گونه‌ای است که اگر دو دانش‌آموز ویژگی‌های رفتاری مشابهی در شبکه‌ی اجتماعی داشته باشند، بردارهای بازنمایی آن‌ها هم شبیه به یکدیگر خواهد بود. معیار شباهت دو بردار را فاصله‌ی اقلیدسی در نظر بگیرید. رابطه‌ی فاصله‌ی اقلیدسی در زیر آمده است:

$$||p-q||_۲^۲ = \sum_{i=۱}^{d}(p_i - q_i)^۲ \qquad (۳)$$

در رابطه‌ی بالا p و q دو بردار دلخواه d بعدی هستند. p_i و q_i هم به ترتیب درایه‌های iام بردارهای p و q هستند. این متخصص، برای تقسیم‌بندی دانش‌آموزان به K گروه تابع هزینه‌ی (Loss Function) زیر را کمینه کرد.

$$\mathcal{L} = \sum_{k=۱}^{K}\sum_{x\in C_k}||x-\mu_k||_۲^۲$$
$$\mu_k = \frac{۱}{|C_k|}\sum_{x\in C_k} x \qquad (۴)$$

در رابطه‌ی بالا K تعداد گروه‌ها، μ_k مرکز گروه k و C_k مجموعه داده‌های موجود در گروه k است.
جدول ۱ مقدار این تابع هزینه را به ازای تعداد گروه‌های گوناگون نشان می‌دهد. با توجه به این جدول، بهترین تعداد گروه‌ها برای گروه‌بندی یا خوشه‌بندی دانش‌آموزان کدام است؟

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

جدول ۱: تعداد گروه و تابع هزینه‌ی متناظر با آن

تعداد گروه	مقدار تابع هزینه
۱	۱۰۰
۲	۷۰
۳	۵۰
۴	۴۰
۵	۳۷
۶	۳۳
۷	۳۰
۸	۲۴
۹	۲۰
۱۰	۱۸

۱) ۲ ۲) ۴ ۳) ۷ ۴) ۱۰

۵ فرض کنید می‌خواهید یک مسئله‌ی دسته‌بندی ۵ دسته‌ای را حل کنید. به شما گفته می‌شود که دسته‌های مختلف نسبت به یکدیگر به صورت خطی جدایی پذیر (Linearly Separable) هستند و شما برای حل مسئله تنها مجاز به استفاده از دسته‌بندهای دودویی (Binary Classifier) هستید. آیا به ازای همه‌ی انواع داده‌ی آموزش با خواص ذکر شده، کمترین تعداد دسته‌بند مورد نیاز برای حل این مسئله را می‌توان به صورت یکتا تعیین کرد؟ در صورتی که می‌توان این کار را انجام داد این تعداد چندتا است؟

۱) بله، ۳ ۲) بله، ۴ ۳) بله، ۵ ۴) خیر، به‌ازای داده‌های آموزش مختلف، کمینه‌ی تعداد دسته‌بند می‌تواند فرق کند.

۶ وزارت راه و شهرسازی برای تشخیص این که آیا یک خانه فروش می‌رود یا خیر از یک دسته‌بند خطی (Linear Classifier) استفاده می‌کند. ویژگی‌های (Features) زیر به عنوان ورودی به این دسته‌بند داده می‌شود:

- $x_۱$: قیمت خانه به تومان
- $x_۲$: مساحت خانه به متر مربع
- $x_۳$: شهری یا روستایی بودن؛ در صورت شهری بودن مقدار $x_۳ = ۱$ و در صورت روستایی بودن مقدار $x_۳ = ۰$ است.

در نهایت دسته‌بند طراحی شده به صورت زیر کار می‌کند:

$$\mathbb{I}[\sigma(w_۱x_۱ + w_۲x_۲ + w_۳x_۳ + b) > ۰/۵] \quad (۵)$$

که در معادله‌ی بالا σ نماینده‌ی تابع سیگموید است.

$$\sigma(x) = \frac{۱}{۱ + e^{-x}} \quad (۶)$$

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

مقادیر w_i وزن‌های مدل خطی، مقدار b نماینده‌ی بایاس (عرض از مبدا) و $\mathbb{I}$ تابع نشانگر (Indicator Function) است که به صورت زیر تعریف می‌شود:

$$\mathbb{I}(x) = \begin{cases} ۱ & \text{شرط x درست باشد.} \\ ۰ & \text{شرط x غلط باشد.} \end{cases} \quad (۷)$$

همچنین برای آموزش پارامترهای دسته‌بند از تابع هزینه‌ی زیر استفاده شده است:

$$\mathcal{L} = \frac{۱}{N}\sum_{i=۱}^{N}[-y^i\ log(\hat{y}^i) + (-۱)(۱-y^i)\ log(۱-\hat{y}^i)] + \lambda(w_۱^۲ + w_۲^۲ + w_۳^۲ + b^۲) \quad (۸)$$

که در معادله بالا، N تعداد داده‌ها، y^i برچسب داده iام ($y^i = ۱$ به معنای فروش رفتن و $y^i = ۰$ معادل با فروش نرفتن خانه) و $\hat{y}^i = \sigma(w_۱x_۱^i + w_۲x_۲^i + w_۳x_۳^i + b)$ خروجی پیش‌بینی شده است. برای جلوگیری از بیش‌برازش (Overfitting)، مقدار λ بزرگ انتخاب شده است.

پس از آموزش مدل، کارشناسی که از مدل استفاده می‌کند با تعجب درمی‌یابد که تنها ویژگی $x_۱$ (قیمت خانه) در نتیجه‌ی نهایی مدل اثرگذار است. این در صورتی است که او می‌داند مساحت خانه‌های شهری و روستایی در فروش رفتن خانه تاثیرگذار هستند. کدام یک از گزینه‌های زیر نمی‌تواند به عنوان راه‌حل در نظر گرفته شود؟

۱) کم کردن ضریب منظم‌سازی (Regularization) (λ)
۲) حذف ضریب منظم‌سازی (λ)
۳) حذف پارامتر عرض از مبدا (b)
۴) هنجارسازی (Normalization) ویژگی‌های ورودی به گونه‌ای که داده‌های ورودی میانگین صفر و واریانس واحد داشته باشند.

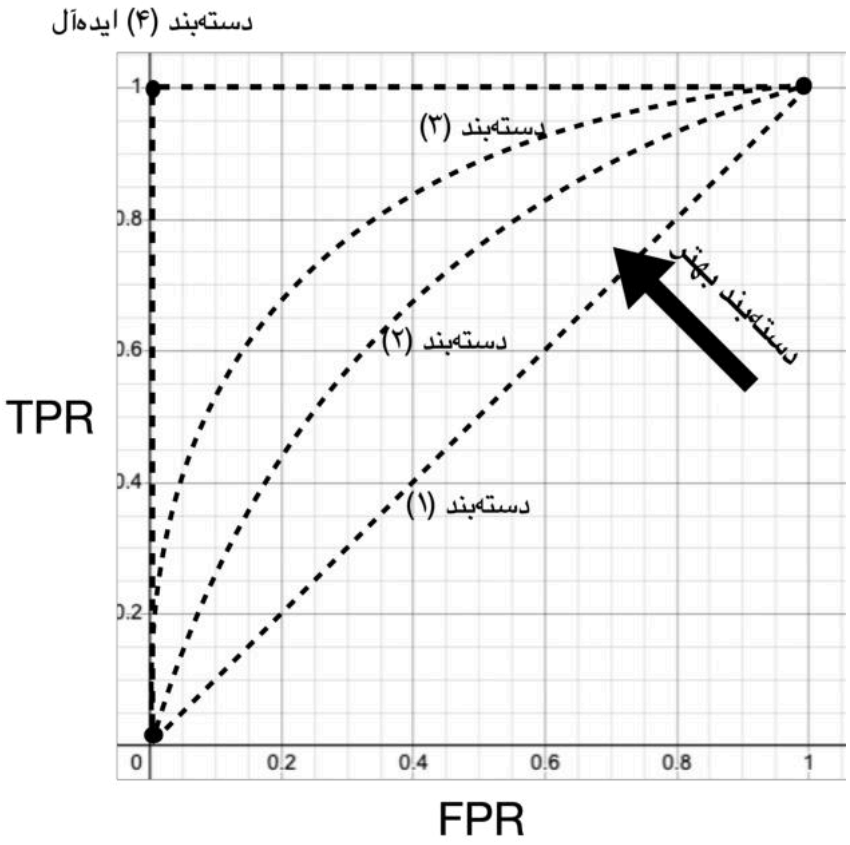
۷ مسئله‌ی دسته‌بندی (Classification) داده‌ها به دو دسته‌ی مثبت و منفی را در نظر بگیرید. مدلی که آموزش دادیم به این صورت عمل می‌کند که احتمال مثبت بودن داده p را خروجی می‌دهد. سپس برای انتساب نهایی داده به دسته منفی یا مثبت از یک آستانه (Threshold) $۰ \leqslant T \leqslant ۱$ استفاده می‌کند. به طوری که اگر $p < T$ باشد داده به دسته‌ی منفی و اگر $p \geqslant T$ باشد داده به دسته‌ی مثبت نسبت داده می‌شود. می‌خواهیم عملکرد این مدل دسته‌بند دودویی (Binary Classifier) را مورد بررسی قرار دهیم. برای این کار معیارهای زیر را تعریف می‌کنیم:

$$\text{TPR} = \frac{\text{تعداد داده‌هایی که به درستی مثبت تشخیص داده شده‌اند}}{\text{تعداد کل داده‌های مثبت}} \quad (۹)$$

$$\text{FPR} = \frac{\text{تعداد داده‌هایی که به اشتباه مثبت تشخیص داده شده‌اند}}{\text{تعداد کل داده‌های منفی}} \quad (۱۰)$$

نمودار ROC نموداری است که مقدار TPR را روی محور عمودی و مقدار FPR را روی محور افقی نمایش می‌دهد. دسته‌بند دودویی ایده‌ال دسته‌بندی است که $\text{TPR} = ۱$ و $\text{FPR} = ۰$ داشته باشد. حال فرض کنید به ازای تمامی مقایر $۰ \leqslant T \leqslant ۱$ مقادیر TPR و FPR را حساب می‌کنیم و آن را در نمودار ROC نشان می‌دهیم. حاصل کار نمودارهایی شبیه به نمودارهای شکل ۳ خواهد بود.

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید



شکل ۳: نمودار ROC برای دسته بند تصادفی به ازای هر آستانه مقدار TPRو FPR یکسانی دارد. که با خط $(y = x)$ نشان داده شده است.

در این نمودار هر چه یک دسته‌بند بهتر باشد به ازای آستانه‌های مختلف میزان TPR بیشتر و میزان FPR کمتری دارد. بنابراین نمودار دسته‌بند بهتر به سمت نقطه‌ی (۰, ۱) کشیده می‌شود (دسته‌بند ۴ در شکل ۳). دسته‌بندی که به صورت تصادفی و با احتمال برابر به ازای همه‌ی آستانه‌ها نمونه‌ها را به دسته‌ی مثبت و منفی تخصیص دهد، مقدار TPR و FPR آن با هم برابر خواهد بود (دسته‌بند ۱ در شکل ۳).

معیار AUC را به صورت مساحت زیر نمودار بدست آمده برای مدل تعریف می‌کنیم. به این ترتیب اگر یک دسته‌بند ایده‌ال باشد مساحت زیر نمودار ROC آن برابر با مساحت مربع واحد و ۱ خواهد بود. هر چه دسته‌بند بهتر باشد نمودار آن به سمت نقطه‌ی (۰, ۱) متمایل‌تر و مساحت زیر نمودار آن به عدد ۱ نزدیک‌تر است. مثلا در شکل ۳ دسته‌بند ۳ مقدار AUC بیشتر و عملکرد بهتری نسبت به دسته‌بند ۲ دارد. دسته‌بند تصادفی هم معیار AUC برابر با ۰/۵ دارد.

پرسش: فرض کنید برای بیماری سرطان ریه تعدادی تصاویر سی‌تی‌اسکن ریه همراه با برچسب مثبت (بیمار) و منفی (سالم) در اختیار داریم و احتمال مثبت بودن برچسب تصاویر را بوسیله‌ی یک مدل هوشمند آموزش‌دیده برای تشخیص سرطان ریه پیش‌بینی کرده‌ایم. روی این داده‌ها مقدار AUC را محاسبه کردیم. حال از بین داده‌های منفی با احتمال برابر و یکسان $0 < q < 1$ نمونه‌برداری (Sampling) می‌کنیم و دوباره معیار AUC را روی داده‌های مثبت و داده‌های نمونه‌برداری شده‌ی منفی حساب می‌کنیم. کدام گزینه درباره‌ی تغییرات مقدار AUC بین این دو ارزیابی درست است؟

۱) مقدار AUC پس از نمونه‌برداری بیشتر از مقدار AUC پیش از نمونه‌برداری است.
۲) مقدار AUC پس از نمونه‌برداری کمتر از مقدار AUC پیش از نمونه‌برداری است.
۳) مقدار AUC پس از نمونه‌برداری با مقدار AUC پیش از نمونه‌برداری برابر است.

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

۴) درباره‌ی تغییرات AUC پیش و پس از نمونه‌برداری چیزی نمی‌توان گفت.

۸ شرکتی بزرگ برای یکی از پروژه‌های محرمانه‌ی خود یک شبکه‌ی عصبی برای مسئله‌ی دسته‌بندی با K دسته (K-Class Classification) آموزش می‌دهد. به دلیل محرمانه بودن پروژه دسترسی به ضرایب شبکه‌ی عصبی و داده‌های ورودی امکان‌پذیر نیست. در هر گام تنها می‌توان بردار خروجی شبکه را قبل از تابع softmax (Logits) در اختیار داشت. از شما خواسته شده تا با گرفتن این K عدد یک برنامه بنویسید تا پس از اعمال تابع softmax خروجی را به تابع هزینه‌ی انتروپی متقاطع (Cross Entropy Loss Function) بدهد و مقدار تابع هزینه را برگرداند. رابطه‌ی تابع softmax به این صورت است:

$$p_i = \frac{e^{a_i}}{\sum_{j=۱}^{K} e^{a_j}} \quad (۱۱)$$

که a_i مقدار خروجی شبکه به ازای دسته‌ی i (قبل از اعمال تابع softmax) است و p_i احتمال نسبت داده شده به دسته‌ی iام. همچنین تابع هزینه‌ی انتروپی متقاطع در زیر آمده است:

$$\mathcal{L} = \frac{۱}{N} \sum_{n=۱}^{N} -log(p_{y_n}^n) \quad (۱۲)$$

که N تعداد داده‌ها و $p_{y_n}^n$ خروجی تابع softmax به ازای دسته‌ی واقعی داده‌ی nام (y_n) است.

در فرایند آموزش متوجه می‌شوید که برخی از مقادیر a_i بزرگ هستند. همین امر سبب می‌شود تا محاسبه‌ی e^{a_i} توسط رایانه امکان‌پذیر نباشد. کدام یک از گزینه‌های زیر می‌تواند بدون تاثیر بر فرایند آموزش شبکه، مشکل را حل کند؟

۱) به جای a_i ها از b_i استفاده کنیم:

$$b_i = \frac{a_i}{\max\limits_{j \in [۱,K]} a_j} \quad (۱۳)$$

۲) نمی‌توان اقدام موثری انجام داد و باید درخواست کرد تا داده‌های ورودی بهنجار (Normalize) شوند طوری‌که میانگین داده‌ها برابر با صفر و واریانس آن‌ها واحد شود.

۳) به جای a_i ها از b_i استفاده کنیم:

$$b_i = a_i - \max\limits_{j \in [۱,K]} a_j \quad (۱۴)$$

۴) نمی‌توان اقدام موثری انجام داد و باید با کاهش تعداد لایه‌های مدل و استفاده از ضرایب اولیه متفاوت تلاش کرد که بر این موضوع غلبه کنیم.

۹ یک شرکت ویدیویی به دنبال توسعه‌ی یک سامانه‌ی توصیه‌گر (Recommender System) فیلم برای مشتریانش است. این شرکت پس از بررسی‌های فراوان توانسته همه‌ی آدم‌های کره‌ی زمین را به ۱۰ دسته‌ی متفاوت و با جمعیت مساوی (در سلیقه‌ی تماشای فیلم) تقسیم کند. این شرکت ۱۰ فیلم برتر را انتخاب کرده و می‌خواهد بوسیله‌ی سامانه‌ی توصیه‌گر خود به افراد مختلف فیلمی را پیشنهاد دهد که احتمال این که فرد آن را دوست داشته باشد بیشینه

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

باشد. اینکه هر دسته‌ای از افراد چه فیلمی را می‌پسندند در جدول زیر آمده است. در این جدول عدد ۱ در هر خانه به این معنی است که افراد دسته‌ی موردنظر فیلم مشخص شده را دوست دارند و عدد ۰ به این معنی است که از آن فیلم خوش‌شان نمی‌آید. به طور مثال افراد دسته‌ی سوم فیلم پنجم را دوست دارند اما فیلم دهم را دوست ندارند.

	فیلم اول	فیلم دوم	فیلم سوم	فیلم چهارم	فیلم پنجم	فیلم ششم	فیلم هفتم	فیلم هشتم	فیلم نهم	فیلم دهم
دسته 1	1	1	0	1	1	1	0	0	0	1
دسته 2	0	1	0	0	0	1	1	1	0	1
دسته 3	1	0	0	0	1	0	0	0	1	0
دسته 4	0	0	1	0	0	0	0	1	1	1
دسته 5	1	1	1	1	1	1	0	0	0	1
دسته 6	1	0	1	0	1	0	1	0	1	0
دسته 7	1	1	1	1	0	0	1	0	1	0
دسته 8	0	1	0	1	1	1	1	0	0	0
دسته 9	0	1	1	1	1	0	1	0	0	0
دسته 10	0	1	1	0	0	0	1	1	1	0

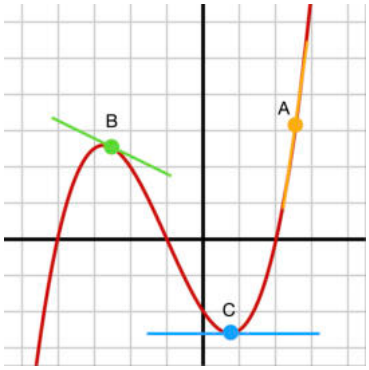
این سامانه‌ی توصیه‌گر به صورت تعاملی عمل می‌کند. نحوه‌ی کار آن بدین صورت است که در هر گام با توجه به اطلاعاتی که از کاربر دارد (دسته‌هایی که ممکن است کاربر عضو یکی از آن‌ها باشد) فیلمی را به او پیشنهاد می‌دهد که بیشترین احتمال دوست داشته شدن توسط کاربر را داشته باشد. سپس این که کاربر آن فیلم را دوست دارد یا ندارد را به عنوان بازخورد دریافت می‌کند (با فشردن دکمه‌ی پسندیدم یا نپسندیدم توسط کاربر) و اطلاعات خود از کاربر (یعنی دسته‌هایی که ممکن است کاربر عضو یکی از آن‌ها باشد) را بروز می‌کند. در صورتی که دو فیلم احتمال یکسانی برای توصیه‌شدن به کاربر داشته باشند فیلمی که عدد یا شماره کمتری دارد به کاربر توصیه می‌شود. فرض کنید می‌دانیم دسته‌ی افراد ۱ تا ۵ مرد و دسته‌ی افراد ۶ تا ۱۰ زن است. در این صورت کدام یک از گزاره‌های زیر درست است؟

۱) در صورتی که هیچ اطلاعاتی از کاربر در دسترس نباشد، فیلم سوم به او پیشنهاد خواهد شد.
۲) در صورتی که کاربر پسر باشد، اگر اولین پیشنهاد را دوست نداشته باشد حتما دومین پیشنهاد را دوست خواهد داشت.
۳) اگر جنسیت کاربر را بدانیم فیلم اول یا سوم به او پیشنهاد خواهد شد.
۴) کاربر دختر حتما اولین پیشنهاد را دوست ندارد.

۱۰ برای آموزش شبکه‌های عصبی (Neural Networks) از روش گرادیان کاهشی (Gradient Descent) استفاده می‌شود. در این روش ابتدا خروجی شبکه به ازای تعدادی از داده‌های ورودی محاسبه می‌شود، سپس با استفاده از این خروجی‌ها و خروجی‌های مطلوب برای این داده‌ها مقدار تابع هزینه (Cost Function) محاسبه می‌شود و در نهایت با محاسبه‌ی مشتق تابع هزینه نسبت به وزن‌های شبکه، وزن‌های جدید بدست می‌آید.
مفهوم ریاضی مشتق را می‌توان به صورت شیب در نظر گرفت. مثلا در تابع شکل ۴ مقدار مشتق در نقطه‌ی

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

A یک مقدار مثبت بزرگ است و شیب خط مماس به تابع در نقطه‌ی A هم بسیار زیاد است. مقدار مشتق تابع در نقطه‌ی B مقدار منفی با اندازه‌ی کوچک است. شیب خط مماس بر تابع در نقطه‌ی B هم همین‌گونه است و مقدار مشتق تابع در نقطه‌ی C برابر با صفر است که خط مماس بر تابع در این نقطه هم این امر را نشان می‌دهد.



شکل ۴: در تابع درجه‌ی ۳ بالا، مقدار مشتق (شیب) در نقطه‌ی A یک مقدار مثبت بزرگ، در نقطه‌ی B یک مقدار منفی کوچک و در نقطه‌ی C برابر با صفر است.

فرض کنید تابع هزینه یک کوهستان بزرگ باشد. قله‌های این کوهستان جاهایی است که مقدار تابع هزینه زیاد است و دره‌های این کوهستان معادل با جاهایی است که مقدار تابع هزینه کم است. به ازای هر مقدار دهی از وزن‌های شبکه‌ی عصبی، مقدار خاصی را برای تابع هزینه بدست می‌آوریم. مثلا در ابتدای فرایند آموزش که وزن‌ها به صورت تصادفی مقداردهی شده در یک نقطه‌ی تصادفی از کوهستان هستیم و مقدار تابع هزینه لزوما کم نیست. هر بار که مشتق تابع هزینه نسبت به وزن‌های شبکه را محاسبه می‌کنیم در حقیقت شیب کوهستان در جایی که ایستاده‌ایم را در نظر می‌گیریم. بروزرسانی (Update) وزن‌ها هم معادل با حرکت رو به پایین است و به ازای وزن‌های جدید می‌خواهیم تابع هزینه محاسبه شده مقدار کمتری داشته باشد. با تکرار این کار، گام به گام در این کوهستان رو به دره (جایی که تابع هزینه مقدار کمتری دارد) حرکت می‌کنیم. (شکل ۵)

شکل ۵: الگوریتم گرادیان کاهشی همانند حرکت در جهت عکس شیب کوهستان است.

هنگام محاسبه‌ی مشتق تابع هزینه نسبت به وزن‌های شبکه‌ی عصبی، به ازای هر لایه مشتق تابع فعالساز (Activation Function) آن لایه در عبارت مشتق کلی به صورت ضرب ظاهر می‌شود. به عنوان یک مثال ساده فرض کنید که شبکه‌ی عصبی معادل با تابع ریاضی زیر باشد:

$$f = W_2\sigma(W_1X) \qquad (۱۵)$$

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

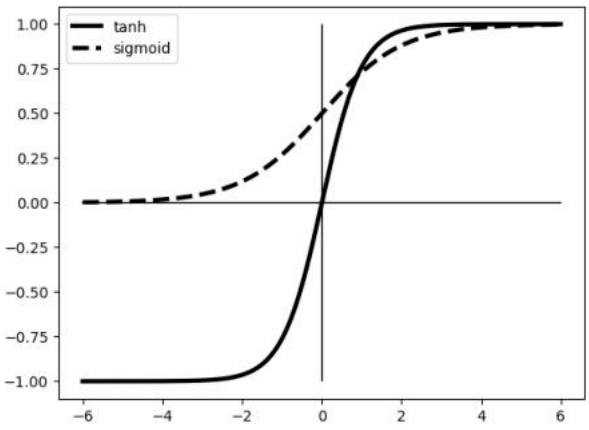
به طور ساده این عبارت یک شبکه‌ی عصبی یک لایه را نشان می‌دهد که ابتدا ورودی X را در وزن‌های لایه‌ی اولیه $W_۱$ ضرب می‌کند. سپس تابع فعالسازی σ را روی آن‌ها حساب می‌کند و خروجی تابع فعالسازی را در $W_۲$ ضرب می‌کند. در این صورت برای محاسبه‌ی مشتق تابع f نسبت به وزن‌های پیش از تابع فعالسازی $W_۱$ خواهیم داشت:

$$\frac{\partial f}{\partial W_۱} = \frac{\partial f}{\partial \sigma}\frac{\partial \sigma}{\partial W_۱} \tag{۱۶}$$

در عبارت بالا $\frac{\partial \sigma}{\partial W_۱}$ مشتق تابع فعالساز لایه‌ی اول نسبت به وزن $W_۱$ است که به صورت ضرب ظاهر شده است. در نهایت رابطه‌ی بروزرسانی وزن‌ها به صورت زیر خواهد بود.

$$W_{t+۱} \leftarrow W_t - \eta \Delta(\text{loss}, W_t) \tag{۱۷}$$

در رابطه‌ی بالا W_t مقدار وزن در مرحله‌ی tام، $W_{t+۱}$ مقدار وزن پس از بروزرسانی، η نرخ یادگیری (Learning Rate) (یک عدد ثابت) و $\Delta(\text{loss}, W_t)$ مشتق تابع هزینه نسبت به وزن در نقطه W_t است.



شکل ۶: نمودار تابع هزینه‌ی سیگموید و تانژانت هایپربولیک

نمودار دو تا از این توابع فعالساز در شکل ۶ آمده است. با توجه به توضیحات بالا فرایند آموزش شبکه‌ی عصبی تماما متصل (Fully Connected) در کدام یک از گزینه‌های زیر به درستی صورت می‌گیرد؟

۱) تمامی وزن‌های شبکه برابر با ۰/۵ مقداردهی (Initialize) شده باشند.
۲) شبکه‌ی عصبی عمیق و چندلایه با تابع فعالساز sigmoid که ورودی آن با توزیع نرمال $x \sim \mathcal{N}(۱۵, ۱)$ و وزن‌های آن با توزیع نرمال $w \sim \mathcal{N}(۰, ۱)$ مقداردهی اولیه شده باشند.
۳) شبکه‌ی عصبی عمیق و چندلایه با تابع فعالساز tanh که ورودی آن با توزیع نرمال $x \sim \mathcal{N}(۱۵, ۱)$ و وزن‌های آن با توزیع نرمال $w \sim \mathcal{N}(۰, ۱)$ مقداردهی اولیه شده باشند.
۴) شبکه‌ی عصبی عمیق و چندلایه با تابع فعالساز tanh که ورودی آن با توزیع نرمال $x \sim \mathcal{N}(۰, ۱)$ و وزن‌های آن با توزیع نرمال $w \sim \mathcal{N}(۰, ۱)$ مقداردهی اولیه شده باشند.

۱۱ تابع

$$\begin{aligned} &f : \mathbb{R} \to \mathbb{R} \\ &f(x) = x^۳ - ۶x^۲ + ۱۱x - ۶ \end{aligned} \tag{۱۸}$$

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

را در نظر بگیرید. می‌خواهیم کمینه‌ی (Minimum) این تابع را با استفاده از الگوریتم گرادیان کاهشی (Gradient Descent) پیدا کنیم. رابطه‌ی بروزرسانی برای این تابع در الگوریتم گرادیان کاهشی به صورت زیر است:

$$x_{t+۱} = x_t - \eta(۳x_t^۲ - ۱۲x_t + ۱۱) \quad (۱۹)$$

فرض کنید مقدار $\eta = ۰/۱$ باشد. دو حالت زیر را در نظر بگیرید.

۱. نقطه‌ی شروع $x_۰ = ۴$
۲. نقطه‌ی شروع $x_۰ = ۱$

کدام گزینه درباره‌ی نتیجه‌ی اجرای الگوریتم درست است؟

۱) در حالت اول الگوریتم به کمینه‌ی محلی (Local Minima) همگرا می‌شود.
۲) در حالت اول همگرایی (Convergence) حاصل نشده و مقدار x_t ها دائما کاهش می‌یابد.
۳) در حالت دوم الگوریتم به کمینه‌ی محلی همگرا می‌شود.
۴) در حالت دوم در نهایت x_t ها حول بهینه‌ی محلی (Local Optima) نوسان می‌کنند.

۱۲ کدام مورد در خصوص آموزش شبکه‌ی عصبی نادرست است؟

۱) با شروع از یک مجموعه مقادیر اولیه‌ی یکسان برای وزن‌ها و بایاس‌های شبکه، روش گرادیان کاهشی (Gradient Descent) در حالتی که از همه‌ی داده‌ها استفاده شود (Full Batch) در اجراهای مختلف در نهایت به یک مجموعه‌ی یکسان از وزن‌ها و بایاس‌ها می‌رسد.
۲) یک گام بهینه سازی در روش گرادیان کاهشی تصادفی (Stochastic Gradient Descent) (حتی با فرض طول گام آموزشی به اندازه‌ی کافی کوچک) ممکن است باعث افزایش تابع هزینه (Loss Function) کلی (که روی همه‌ی نمونه‌های آموزش به دست می‌آید) شود.
۳) طول گام یادگیری (Learning Step) در روش گرادیان کاهشی تصادفی معمولا کوچک‌تر از طول گام نسبت به حالتی است که همه‌ی داده‌ها در نظر گرفته شوند.
۴) بررسی نمودار تابع هزینه در طول زمان روی داده‌های آموزش (Train) می‌تواند باعث تشخیص پدیده‌ی بیش‌برازش (Overfitting) شود.

۱۳ آقای جواهریان مسئول درس انشا در وزارت آموزش و پرورش است. او می‌خواهد از یک مدل هوشمند برای تشخیص اصالت انشای دانش‌آموزان استفاده کند. برای این کار او تعدادی از متن‌های انشای نوشته شده توسط دانش‌آموزان مدارس مختلف را به عنوان انشای اصیل و تعداد زیادی انشای نوشته شده توسط سامانه‌های هوش مصنوعی را به عنوان انشای بی‌اصالت جمع‌آوری می‌کند. سپس این داده‌ها را برای آموزش یک مدل عمیق در اختیار یکی از دانش آموزان خود قرار می‌دهد. این دانش آموز ابتدا متن‌ها را به سه دسته‌ی آموزش (Train)، اعتبارسنجی (Validation) و آزمون (Test) تقسیم می‌کند. در حین فرایند آموزش، او تعدادی از ابرپارامترهای (Hyper-parameter) مدل شامل تعداد لایه‌های مدل، تعداد نورون‌های هر لایه، ضرایب جملات منظم‌سازی (Regularization) و ... را آن قدر تغییر می‌دهد تا عملکرد مدل روی داده‌ی آموزش و اعتبارسنجی بسیار خوب شود. اما در کمال تعجب در می‌یابد که زمانی که این مدل را روی داده‌های آزمون امتحان می‌کند عملکرد بسیار ضعیف‌تری نسبت به داده‌های اعتبار سنجی دارد. کدام یک از گزینه‌های زیر راه‌حلی را بیان می‌کند که حتما نمی‌تواند به حل مشکل پیش آمده کمک کند؟

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

۱) تعدادی داده‌ی جدید به داده‌های اعتبارسنجی اضافه کند سپس فرایند آموزش و پیدا کردن ابرپارامترها را تکرار نماید.
۲) از روش اعتبارسنجی متقابل k بخشی (K-fold Cross Validation) استفاده کند.
۳) همه‌ی داده‌ها را روی هم ریخته، به صورت تصادفی بر (Shuffle) بزند و دوباره به سه‌دسته‌ی آموزش، اعتبارسنجی و آزمون تقسیم کند.
۴) تعداد حالات بیشتری از ترکیب ابرپارامترها را امتحان کند.

۱۴ هنگام آموزش یک شبکه‌ی عمیق، اگر مقدار نرخ یادگیری (Learning Rate) بسیار بالا تنظیم شود، کدام یک از اتفاقات زیر به احتمال زیاد رخ می‌دهد؟

۱) همگرایی سریع و رسیدن به بهینه‌ی کلی (Global Optima)
۲) نوسان شدید وزن‌ها که منجر به عدم همگرایی مدل می‌شود.
۳) کاهش بیش‌برازش (Overfit) به دلیل بروزرسانی‌های بزرگ
۴) افزایش دقت در مدل بدون نیاز به تنظیمات اضافه

۱۵ زمانی که از اندازه‌ی بسته‌ی (Batch Size) کوچک در فرایند آموزش استفاده می‌شود، چه تاثیری بر روند بروزرسانی گرادیان و فرایند همگرایی مدل مشاهده می‌شود؟

۱) بروزرسانی‌های بسیار هموار (Smooth) و بدون نویز
۲) بهبود عملکرد مدل بدون تاثیر بر سرعت همگرایی
۳) نویز در بروزرسانی‌ها که می‌تواند از گیر افتادن در بهینه‌ی محلی (Local Optima) جلوگیری کند اما همزمان همگرایی را ناپایدارتر کند.
۴) افزایش مصرف حافظه و کاهش سرعت آموزش

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

المپیاد هوش مصنوعی

تاریخ آزمون: ۱۴۰۳/۱۱/۱۳

مهر حفاظت آزمون

کد ملی: استان:

نام و نام خانوادگی: جنسیت داوطلب:

منطقه حوزه: کد حوزه:

کد داوطلبی: □□□□□□□□□□□□

لطفاً داخل کادر چیزی ننویسید و گزینه‌ها را با مداد مشکی نرم و به طور کامل پر کنید.

	۱	۲	۳	۴
۱	●			
۲		●		
۳			●	
۴		●		
۵				●
۶			●	
۷			●	
۸			●	
۹		●		
۱۰				●
۱۱	●			
۱۲				●
۱۳				●
۱۴		●		
۱۵			●	
۱۶				
۱۷				
۱۸				
۱۹				
۲۰				
۲۱				
۲۲				
۲۳				
۲۴				
۲۵				
۲۶				
۲۷				
۲۸				
۲۹				
۳۰				

	۱	۲	۳	۴
۳۱				
۳۲				
۳۳				
۳۴				
۳۵				
۳۶				
۳۷				
۳۸				
۳۹				
۴۰				
۴۱				
۴۲				
۴۳				
۴۴				
۴۵				
۴۶				
۴۷				
۴۸				
۴۹				
۵۰				
۵۱				
۵۲				
۵۳				
۵۴				
۵۵				
۵۶				
۵۷				
۵۸				
۵۹				
۶۰				

	۱	۲	۳	۴
۶۱				
۶۲				
۶۳				
۶۴				
۶۵				
۶۶				
۶۷				
۶۸				
۶۹				
۷۰				
۷۱				
۷۲				
۷۳				
۷۴				
۷۵				
۷۶				
۷۷				
۷۸				
۷۹				
۸۰				
۸۱				
۸۲				
۸۳				
۸۴				
۸۵				
۸۶				
۸۷				
۸۸				
۸۹				
۹۰				

اینجانب ..... به کد ملی ..... دفترچه‌ی سوالات المپیاد هوش مصنوعی شامل ۲۰ سوال را به طور کامل دریافت نمودم.

امضا و اثر انگشت:

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

باسمه تعالی
جمهوری اسلامی ایران
وزارت آموزش و پرورش
باشگاه دانش‌پژوهان جوان

مبارزهٔ علمی برای جوانان، زنده‌کردن روح جست‌وجو و کشف واقعیّت‌هاست. «امام خمینی (ره)»

دفترچهٔ سؤالات مرحلهٔ دوم

اولین دورهٔ المپیاد هوش مصنوعی(تشریحی)

سال تحصیلی ۱۴۰۴-۱۴۰۳

تاریخ: ۱۴۰۴/۱/۲۸ - ساعت: ۱۶:۰۰ - مدت: ۱۵۰ دقیقه - نوع: تشریحی

توضیحات مهم

استفاده از هرنوع ماشین‌حساب ممنوع است.

۱- مشخصات خود را با اطلاعات بالای هر صفحه تطبیق دهید در صورتی که حتی یکی از صفحات پاسخ نامه با مشخصات شما هم‌خوانی ندارد بلافاصله مراقبین را مطلع نمایید.

۲- پاسخ هر سوال را در محل تعیین شده خود بنویسید. چنانچه همه یا قسمتی از جواب سوال را در محل پاسخ سوال دیگری بنویسید به شما نمره ای تعلق نمی گیرد.

۳- با توجه به آنکه برگه های پاسخ نامه به نام شما صادر شده است امکان ارائه هیچ‌گونه برگه اضافه وجود نخواهد داشت. لذا توصیه می‌شود ابتدا سوالات را در برگه چرک نویس، حل کرده و آنگاه در پاسخ‌برگ پاکنویس نمایید.

٤- عملیات تصحیح توسط مصححین پس از برش سربرگ به صورت ناشناس انجام خواهد شد. لذا از درج هرگونه نوشته یا علامت مشخصه که نشان دهنده صاحب برگه باشد، خودداری نمایید. در غیر این صورت تقلب محسوب شده و در هر مرحله ای که باشید از ادامه حضور در المپیاد محروم خواهید شد.

٥- از مخدوش کردن **بارکدها و مربع‌ها در چهارگوشهٔ صفحه در دفترچهٔ پاسخ‌برگ** جداً خودداری کنید. در غیر این صورت برگهٔ شما تصحیح نخواهد شد.

٦- همراه داشتن هر گونه کتاب جزوه یادداشت و لوازم الکترونیکی نظیر تلفن همراه، ساعت هوشمند، دستبند هوشمند و لپ تاپ ممنوع است همراه داشتن این قبیل وسایل حتی اگر از آن استفاده نکنید یا خاموش باشد تقلب محسوب خواهد شد.

۷- این دفترچه شامل ۴ سوال و با احتساب جلد ۳ برگ است.

کلّیهٔ حقوق این سؤالات برای باشگاه دانش‌پژوهان جوان محفوظ است.
آدرس پایگاه اینترنتی: ysc.medu.gov.ir

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

۱ صدا و سیما قصد دارد تا برای دسته‌بندی (Classification) فیلم‌ها در ژانرهای مختلف از یک شبکه‌ی عصبی (Neural Network) استفاده کند. این شبکه به این صورت کار می‌کند که می‌تواند برچسب‌های مختلفی را به یک فیلم نسبت دهد. مثلا فیلم *"الف"* می‌تواند طنز و هیجان‌انگیز باشد و فیلم *"ب"* می‌تواند جنایی، معمایی و رازآلود باشد. فرض کنید که در کل K دسته‌ی متفاوت (K ژانر فیلم متفاوت) داشته باشیم. صداوسیما از محمدجواد درخواست می‌کند تا این شبکه را آموزش دهد. محمدجواد مدل‌سازی خود را به این صورت انجام می‌دهد:

- به ازای هر داده یک بردار K بعدی را به عنوان برچسب در نظر می‌گیرد. در این بردار اگر یک فیلم ژانر iام را داشته باشد درایه‌ی iام برابر با یک ($y_i = ۱$) و در غیر این صورت صفر ($y_i = ۰$) است.
- در خروجی شبکه‌ی عصبی یک لایه با K دسته قرار می‌دهد. سپس این خروجی را به تابع softmax می‌دهد تا احتمال هر دسته را محاسبه کند. رابطه‌ی تابع softmax به صورت زیر است:

$$p_i = \frac{e^{a_i}}{\sum_{j=۱}^{K} e^{a_j}} \quad (۱)$$

در رابطه‌ی بالا p_i احتمال داشتن برچسب iام و a_i خروجی (Logit) هر دسته از شبکه‌ی عصبی قبل از اعمال تابع softmax است.

- برای آموزش شبکه از تابع هزینه‌ی (Loss Function) زیر استفاده می‌کند:

$$\mathcal{L} = \frac{۱}{N} \sum_{s=۱}^{N} \sum_{i=۱}^{K} -y_i^s log(p_i^s) \quad (۲)$$

در رابطه‌ی بالا y_i^s مقدار درایه‌ی iام برای داده‌ی sام و p_i^s خروجی softmax برای این داده است. N هم تعداد کل داده‌ها است.

با این توضیحات به سوالات زیر پاسخ دهید:

۱. فرض کنید شبکه‌ای که در اختیار داریم این قدر بزرگ است که توانایی بیش‌برازش (Overfit) به داده‌ی آموزش (Train) را دارد. آیا در این صورت اگر تعداد دورهای آموزش (Epoch) را بالا ببریم مقدار تابع هزینه روی داده‌های آموزش صفر خواهد شد؟ توضیح دهید.

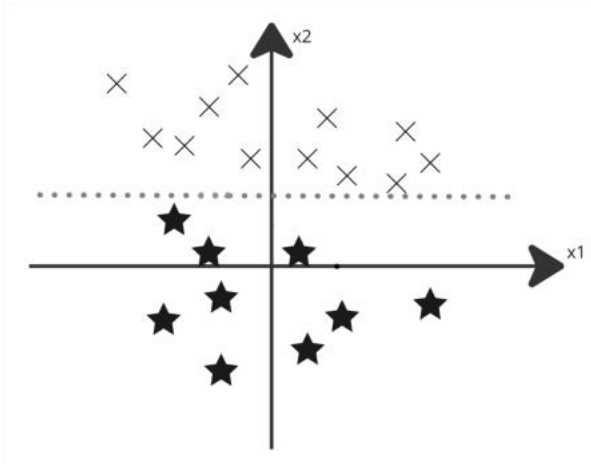
۲. فرض کنید با روش مدل‌سازی محمدجواد یک مدل را آموزش دادیم. نشان دهید در این مدل در صورتی که فیلم A تعداد برچسب بیشتری نسبت به فیلم B داشته باشد (تعداد ژانرهای بیشتری داشته باشد)، آنگاه مقدار تابع هزینه برای فیلم A بیشتر از فیلم B خواهد بود.

۳. آیا می‌توانید مدل‌سازی بهتری نسبت به آنچه محمدجواد برای این مسئله انجام داده، ارائه کنید. در صورت پاسخ مثبت این کار را انجام دهید و توضیح دهید مدل‌سازی جدید شما چه مشکلی را در مدل‌سازی محمدجواد رفع می‌کند؟

۲ یک مدل رگرسیون لجستیک (Logistic Regression) قادر است تا یک خط تصمیم مرزی برای داده‌های شکل ۱ به صورت زیر ارائه کند.

$$w_۲x_۲ + w_۱x_۱ + w_۰ = ۰ \quad (۳)$$

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید



شکل ۱: داده‌های سوال تصمیم‌گیری خطی

همان‌طور که از شکل پیداست داده‌ها به صورت خطی جدایی پذیرند. فرض کنید تابع هزینه (Loss Function) این مدل را $\mathcal{L}_\cdot$ بنامیم. تابع هزینه‌ی جدید $\mathcal{L}_i$ را به صورت زیر تعریف می‌کنیم.

$$\mathcal{L}_i = \mathcal{L}_\cdot + \lambda w_i^۲, \quad i \in \{۰, ۱, ۲\} \tag{۴}$$

منظور از w_i آن است که می‌توانیم یکی از ضرایب را (به دلخواه) انتخاب کنیم. اگر فرض کنیم λ مقدار بزرگی انتخاب شود به ازای انتخاب کدام یک از $\mathcal{L}_i$ ها مدل می‌تواند این دو دسته را از هم جدا کند؟ توضیح دهید.

۳ مریم به این موضوع علاقه‌مند است تا مدل‌های درخت تصمیم‌گیری (Decision Tree) را برای مسائل دسته‌بندی تبدیل به شبکه عصبی بکند. برای این کار او از شبکه‌های عصبی پرسپترون چند لایه (Multi-Layer Perceptron) با تابع فعالسازی (Activation Function) پله‌ای برای لایه‌های پنهان (Hidden Layers) استفاده می‌کند:

$$\sigma(x) = \begin{cases} ۱ & x \geqslant ۰ \\ ۰ & x < ۰ \end{cases} \tag{۵}$$

همچنین او از یک لایه softmax برای بردار احتمال دسته‌بندی استفاده می‌کند. در نهایت مدل دسته‌ی با بیشترین احتمال را به عنوان خروجی تعیین می‌کند.

درخت تصمیم‌گیری T به مریم داده شده است. گره‌های این درخت همگی دارای شروطی به شکل $x_i \geqslant C$ و یا $x_i \leqslant C$ هستند که C در هر گره یک عدد ثابت است. این درخت تصمیم‌گیری، داده‌ای به شکل چهار ویژگی عددی حقیقی را به عنوان ورودی می‌گیرد و آنها را در بین ۳ دسته تقسیم‌بندی می‌کند. مریم برای این درخت تصمیم‌گیری، شبکه عصبی‌ای با دو لایه پنهان پیدا کرده است که معادل آن عمل می‌کند. به این صورت که اگر بردار $\mathbf{x} = [x_۱, x_۲, x_۳, x_۴]^T$ به عنوان ورودی به شبکه عصبی داده شود، شبکه عصبی به شکل زیر بردار احتمال $\mathbf{y} = [y_۱, y_۲, y_۳]^T$ را به عنوان بردار احتمال دسته‌ها خروجی می‌دهد. روابط ریاضی این شبکه به صورت زیر است:

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

$$\mathbf{h}_1 = \sigma\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} -5 \\ 4 \\ -7 \\ -1 \\ 3 \\ 10 \end{bmatrix}\right),$$

$$\mathbf{h}_2 = \sigma\left(\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & -1 & 0 & 0 \end{bmatrix}\mathbf{h}_1 + \begin{bmatrix} -3 \\ -3 \\ -2 \\ -1 \\ -1 \\ -1 \\ 0 \end{bmatrix}\right),$$

$$\mathbf{y} = \text{softmax}\left(\begin{bmatrix} 10 & 0 & 0 & 0 & 10 & 0 & 10 \\ 0 & 0 & 10 & 0 & 0 & 10 & 0 \\ 0 & 10 & 0 & 10 & 0 & 0 & 0 \end{bmatrix}\mathbf{h}_2\right)$$

۱. فرض کنید درخت T گره تکراری یا پوچ نداشته باشد. در این صورت T را پیدا کنید. (منظور از گره تکراری یعنی یک شرط چند بار در درخت تکرار شده باشد. مثلا $x_1 \geqslant 2$ یا معادل آن دو بار یا بیشتر در درخت تکرار شده باشد. اما دو شرط $x_1 \geqslant 1$ و $x_1 \geqslant 2$ تکراری محسوب نمی‌شوند. همچنین منظور از گره پوچ آن است که خروجی آن همیشه مثبت یا منفی باشد و در نتیجه بتوان آن را از درخت حذف کرد. مثلا اگر ذیل شاخه مثبت $x_1 \geqslant 2$ شرط $x_1 \geqslant 1$ آمده باشد، نتیجه آن همیشه مثبت بوده و می‌توان آن را حذف کرد.)

۲. مریم می‌خواهد به جای استفاده از تابع فعالسازی σ از تابع فعالسازی

$$\text{ReLU}(x) = \max(0, x) \qquad (۶)$$

در شبکه عصبی خود استفاده کند. به مریم کمک کنید تا شبکه عصبی معادل درخت تصمیم‌گیری T پیدا کند. (برای سادگی مساله فرض کنید پارامترهای ورودی، یعنی x_iها، اعداد صحیح هستند.)

۴ در شهر نوسترادامیا، سه پیشگو با قدرت‌های خارق‌العاده زندگی می‌کنند که ساکنان شهر برای تصمیم‌گیری‌های مهم به پیش‌بینی‌های آن‌ها تکیه می‌کنند. هر پیشگو برای هر پرسش از اتفاقاتی که در آینده ممکن است رخ دهد یکی از دو پاسخ زیر را ارائه می‌کند.

- "اتفاق می‌افتد" (۱)
- "اتفاق نمی‌افتد" (۰)

اما مشکل اینجاست که هر پیشگو در بازه‌های خاصی از زمان دچار اشتباه می‌شود و پیش‌بینی نادرستی ارائه می‌دهد! برای کاهش خطاها، شورای شهر تصمیم گرفته است که از رأی اکثریت استفاده کند؛ یعنی اگر حداقل دو پیشگو پیش‌بینی یکسانی داشته باشند، همان را به‌عنوان پیش‌بینی نهایی اعلام می‌کنند.

تعریف دقیق مسئله سه پیشگو به نام‌های M_1، M_2 و M_3 داریم که هرکدام برای هر لحظه‌ی آینده، یک پیش‌بینی ارائه می‌دهند. این پیشگوها می‌خواهند رخداد یا عدم رخداد اتفاق موردنظر را در زمان‌های بازه‌ی $[0, 1]$ پیش‌بینی کنند. پیشگویی‌ها قطعی هستند، یعنی اگر لحظه‌ای از آینده دوباره پیش‌بینی شود، همان پاسخ قبلی توسط پیشگو

@sampaadia amoozz.ir برای دسترسی به دوره های متنوع به وبسایت آموزشگاه ذهن زیبا مراجعه کنید

ارائه می‌شود. هر پیشگو در یک بازه‌ی خاص از زمان دچار خطا می‌شود. در این بازه‌ها، پیشگوی موردنظر همیشه پیش‌بینی اشتباه ارائه می‌دهد. این بازه‌های خطا به‌صورت زیر تعریف شده‌اند:

- پیشگو $M_۱$ در بازه‌ی $[I_۱, I_۲]$ دچار خطا می‌شود که طول این بازه برابر $I_۲ - I_۱ = ۰/۳۰$ است.
- پیشگو $M_۲$ در بازه‌ی $[I_۳, I_۴]$ دچار خطا می‌شود که طول آن $I_۴ - I_۳ = ۰/۳۵$ است.
- پیشگو $M_۳$ در بازه‌ی $[I_۵, I_۶]$ دچار خطا می‌شود که طول آن $I_۶ - I_۵ = ۰/۴۰$ است.

۱. **سناریوی کابوس: بیشترین میزان خطا**
مقادیر $I_۱$ تا $I_۶$ را به‌گونه‌ای انتخاب کنید که بیشترین میزان خطا در خروجی نهایی پیش‌بینی رخ دهد. یعنی حالتی را بیابید که در آن، تصمیم‌گیری اکثریت در بیشترین لحظات دچار اشتباه شود. سپس ثابت کنید که این انتخاب شما بیشترین میزان خطا را ایجاد می‌کند.

۲. **سناریوی طلایی: کمترین میزان خطا** مقادیر $I_۱$ تا $I_۶$ را به‌گونه‌ای انتخاب کنید که کمترین میزان خطا در خروجی نهایی پیش‌بینی رخ دهد. یعنی حالتی را بیابید که در آن، رأی اکثریت کمترین میزان اشتباه را داشته باشد. سپس ثابت کنید که این انتخاب شما کمترین میزان خطا را به همراه دارد.