



دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

درس یادگیری ماشین

تکلیف تئوری دوم

تاریخ تحویل: ۶ دی

سوال ۱ (۱۰ نمره)

فرض کنید یک مدل Logistic Regression با تابع فرضیه ^۱ H به صورت زیر داریم.

$$\hat{y} = h_{\theta}(x) = g(\theta_2 x_2 + \theta_1 x_1 + \theta_0)$$

الف) اگر $\theta_2 = -1, \theta_1 = 0, \theta_0 = 3$ باشند. مرز تصمیم به چه صورت خواهد بود؟ (مرز تصمیم قبل از اعمال تابع فعال سازی g) (۵ نمره)

ب) اگر $\theta_2 = 1, \theta_1 = 1, \theta_0 = -2$ باشند. مرز تصمیم به چه صورت خواهد بود؟ (مرز تصمیم قبل از اعمال تابع فعال سازی g) (۵ نمره)

ج) اگر $y = x_1 || x_2$ (Logical OR) باشد و x_1, x_2 مقادیر باینری باشند و تابع فعالسازی sigmoid باشد. پارامترهای مدل را به گونه‌ای مشخص کنید که این مدل توانایی تفکیک Logical OR را داشته باشد. (۵ نمره)

سوال ۲ (۲۰ نمره)

می‌دانیم در حالتی که تابع فعال سازی sigmoid باشد مشتق زنجیره‌ای BCE ^۲ نسبت به Z به صورت زیر است.

$$Z = WX, \hat{y} = \text{sigmoid}(Z)$$

$$J = \text{BCE} \rightarrow \frac{\partial J}{\partial Z} = \hat{y} - y$$

در حالتی که تابع فعال سازی softmax باشد مشتق زنجیره‌ای CCE ^۳ نسبت به Z را محاسبه کنید.

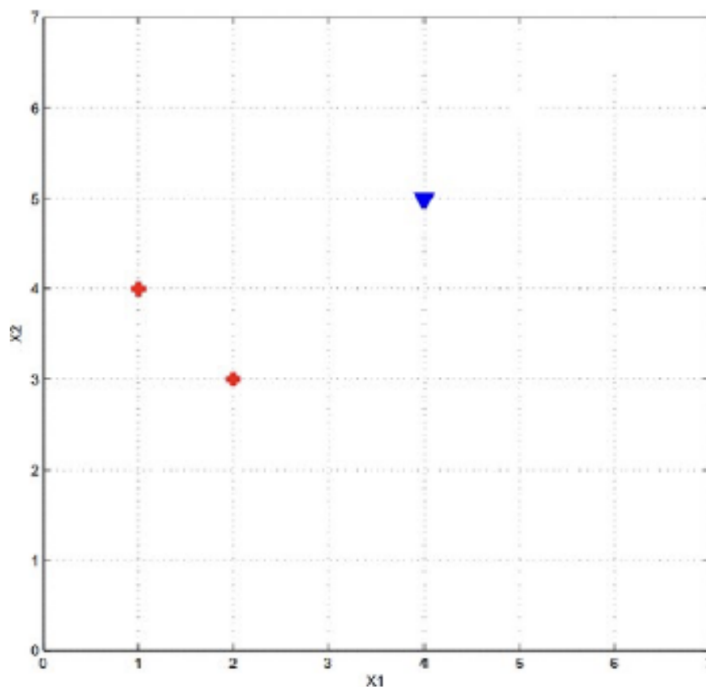
¹Hypothesis Function

²Binary Cross Entropy

³Categorical Cross Entropy

سؤال ۳ (۱۵ نمره)

می‌خواهیم یک طبقه بند ماشین بردار پشتیبان را روی داده‌های زیر آموزش دهیم. در این شکل ۲ داده با مقدار ۱- (مثبت‌های قرمز) و ۱+ (مثبت‌های آبی) نشان داده شده است. (سوال به صورت تحلیلی پاسخ داده شود)



الف) معادله خط تصمیم را به دست آورید. (مقادیر b, w, m را به دست آورید).

ب) نقاط بردار پشتیبان را روی تصویر مشخص کرده و خط تصمیم را رسم کنید.

سؤال ۴ (۲۰ نمره)

مثال‌های آموزشی نشان داده شده در جدول زیر را برای مسئله طبقه‌بندی باینری در نظر بگیرید:

Customer ID	Gender	Car Type	Shirt Size	Class
۱	M	Family	Small	C0
۲	M	Sports	Medium	C0
۳	M	Sports	Medium	C0
۴	M	Sports	Large	C0
۵	M	Sports	Extra Large	C0
۶	M	Sports	Extra Large	C0
۷	F	Sports	Small	C0
۸	F	Sports	Small	C0
۹	F	Sports	Medium	C0
۱۰	F	Luxury	Large	C0
۱۱	M	Family	Large	C1
۱۲	M	Family	Extra Large	C1
۱۳	M	Family	Medium	C1
۱۴	M	Luxury	Extra Large	C1
۱۵	F	Luxury	Small	C1
۱۶	F	Luxury	Small	C1
۱۷	F	Luxury	Medium	C1
۱۸	F	Luxury	Medium	C1
۱۹	F	Luxury	Medium	C1
۲۰	F	Luxury	Large	C1

الف) شاخص Gini را برای کل مجموعه نمونه‌های آموزشی محاسبه کنید.

ب) شاخص Gini را برای ویژگی Customer ID محاسبه کنید.

ج) شاخص Gini را برای ویژگی Gender محاسبه کنید.

د) شاخص Gini را برای ویژگی Car Type با استفاده از multiway split محاسبه کنید.

و) شاخص Gini را برای ویژگی Shirt Size با استفاده از multiway split محاسبه کنید.

ه) کدام یک از ویژگی‌های Gender ، Car Type و Shirt Size بهتر است؟

ی) توضیح دهید چرا Customer ID علیرغم داشتن کمترین شاخص Gini نباید به عنوان شرایط تست استفاده شود؟

سؤال ۵ (۱۵ نمره)

مثال‌های آموزشی نشان داده شده در جدول زیر را برای مسئله طبقه‌بندی باینری در نظر بگیرید:

Instance	a_1	a_2	a_3	Target class
۱	T	T	۱	+
۲	T	T	۶	+
۳	T	F	۵	-
۴	F	F	۴	+
۵	F	T	۷	-
۶	F	T	۳	-
۷	F	F	۸	-
۸	T	F	۷	+
۹	F	T	۵	-

الف) آنتروپی این مجموعه نمونه‌های آموزشی با توجه به کلاس مثبت چیست؟

ب) information gain متغیرهای a_1 و a_2 را محاسبه کنید.

ج) information gain متغیرهای a_3 را به ازای هر بخش‌بندی ممکن محاسبه کنید.

د) بهترین تقسیم‌بندی بین متغیرهای a_1 ، a_2 و a_3 با توجه به information gain کدام است؟

و) بهترین تقسیم‌بندی بین متغیرهای a_1 و a_2 با توجه به پارامتر میزان خطای طبقه‌بندی^۴ کدام است؟

ه) بهترین تقسیم‌بندی بین متغیرهای a_1 و a_2 با توجه به شاخص Gini کدام است؟

^۴classification error rate

سؤال ۶ (اختیاری) (۱۰ نمره)

نشان دهید که آنتروپی یک گره پس از تقسیم آن به گره های جانشین کوچکتر هرگز افزایش نمی یابد.

سؤال ۷ (۱۰ نمره)

نشان دهید که راه حل مسئله بهینه سازی زیر:

$$\min_{\beta_0, \beta} \sum_{i=1}^N [1 - y_i f(x_i)]_+ + \frac{\lambda}{2} \|\beta\|^2$$

با $\lambda = \frac{1}{C}$ مانند معادله زیر است:

$$\min_{\beta_0, \beta} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i$$

تحت شرایط:

$$\forall i : \xi_i \geq 0, y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i$$

نکات تکمیلی

۱. لزومی به تایپ کردن سوالات تئوری نیست؛ ولی در صورتیکه پاسخ آنها به صورت تایپ شده تحویل داده شود، ۱۰ درصد نمره اضافه به شما تعلق میگیرد. در صورتیکه پاسخهای شما تایپ شده نیست، باید پاسخها خوانا و باکیفیت در قالب فایل pdf ارسال شوند.
۲. فرمت نامگذاری تکلیف ارسالی باید به صورت زیر باشد: HWX_Theory_LastName_StudentID که X شماره تکلیف LastName نام خانوادگی شما و StudentID شماره دانشجویی شما است.
۳. انجام این تکلیف به صورت تک نفره است. در صورت مشاهده تقلب، نمرات هم مبدا کپی و هم مقصد آن صفر لحاظ می شود.
۴. شما می توانید تا یک هفته پس از پایان مهلت تکلیف آن را در یکتا بارگزاری کنید. در این صورت به ازای هر روز تاخیر ۷ درصد از نمره تکلیف کسر می شود. پس از اتمام این یک هفته امکان ارسال با تاخیر وجود ندارد.
۵. در صورت وجود هر گونه ابهام و یا سوال می توانید سوالات خود را در گروه سروش بپرسید. هم چنین می توانید برای رفع ابهامات با دستیاران آموزشی از طریق تلگرام در تماس باشید.
آیدی ها:

@Fatemeh2114P

@amir7d0

@mastaraan

@alireza20010226