دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف اول درس مبانی هوش محاسباتی

نام و نام خانوادگی: علیرضا ابره فروش

شماره دانشجویی: ۹۸۱۶۶۰۳

نیم سال تحصیلی: بهار ۱۴۰۱/۱۴۰۲

مدرّس: دکتر مهران صفایانی

دستیاران آموزشی: فاطمه پیری-علیرضا حبیبی-علیرضا صالحی

۱

۱.۱ نتیجه (خروجی)

یک مدل رگرسیون خطی مبتنی بر یک متغیر وابسته پیوسته است. این به معنای این است که متغیر وابسته مقادیر عددی را به جای دسته‌بندی زیرمجموعه‌ها یا گروه‌ها پذیرفته می‌کند. در مقابل، مدل‌های رگرسیون لجستیک مبتنی بر متغیرهای وابسته دودویی هستند. متغیر وابسته فقط می‌تواند دو مقدار ۰ یا ۱ را بپذیرد. همچنین، خروجی رگرسیون خطی مقدار پیوسته دارد (محدوده‌ای از مقادیر را برمی‌گرداند). برای مثال،

- طول سقف (۲۵ اینچ، ۱۹ اینچ، ۵ فوت)
- ارتفاع (۵ فوت و ۸ اینچ، ۶ فوت و ۲ اینچ، ۵ فوت و ۱۰ اینچ)
- سرعت فرار (۲۶۰۰۰ مایل در ساعت، ۲۱۵۰۰ مایل در ساعت، ۲۹۵۰۰ مایل در ساعت)

در عین حال، مدل رگرسیون لجستیک بیان احتمالاتی دارد. به عنوان مثال:

- احتمال شکست در یک بازی تنیس ۳.۸۴ درصد است
- احتمال تصویب یک قانون در کنگره ۱.۲۳ درصد است
- احتمال اعمال محدودیت شبانه‌روزی در زمان شیوع کروناویروس ۱.۶۵ درصد است

علاوه بر این، رگرسیون خطی توزیع نرمال یا گاوسی را مشاهده می‌کند و رگرسیون لجستیک توزیع دوجمله‌ای را نشان می‌دهد.

۲.۱ ارتباط بین متغیرها

فهم رابطه بین متغیرها در هنگام تصمیم‌گیری درباره نوع مدل رگرسیون برای اهداف مختلف بسیار حائز اهمیت است.

رگرسیون خطی یک رابطه خطی بین متغیرها را با رسم یک خط راست بر روی نمودار توصیف می‌کند. این امکان را به حرفه‌ای‌ها می‌دهد تا روابط خطی را بررسی کرده و حرکت آن‌ها را در طول زمان پیگیری کنند. در مقابل، رگرسیون لجستیک برای مطالعه و بررسی احتمال وقوع یک رویداد شناخته شده است. از آنجا که ساختار خطی رابطه متغیری را نشان نمی‌دهد، پیگیری رگرسیون لجستیک با استفاده از ساختارهای خطی لازم نیست.

۳.۱ خطا

هدف رگرسیون خطی یافتن بهترین خط تطبیق داده شده است در حالی که رگرسیون لجستیک مقادیر خط را به منحنی سیگموید متناسب می‌کند. روش محاسبه‌ی تابع خطای رگرسیون خطی، خطای میانگین مربعات است، در حالی که برای رگرسیون لجستیک، این مقدار با استفاده از تخمین بیشینه درست‌نمایی محاسبه می‌شود. رگرسیون خطی از خطای میانگین مربعات (RMSE) برای محاسبه مقدار وزن بعدی نقاط داده‌ای (یا مشاهدات) از طول خط رگرسیون استفاده می‌کند. به عکس رگرسیون لجستیک از روش دقت برای پیش‌بینی مقدار وزن بعدی استفاده می‌کند.

۴.۱ روش‌های تخمین (برآورد)

یک مدل رگرسیون خطی از روش "ordinary least squares" استفاده می‌کند تا بهترین رابطهٔ رگرسیون را پیدا کند. با توجه به این روش، ضرایب رگرسیون برای کاهش جمع مربعات فاصلهٔ هر متغیر پاسخ به مقدار سازگاری انتخاب می‌شوند.

در مقابل، رگرسیون لجستیک از روش "maximum likelihood estimation" استفاده می‌کند، جایی که ضرایب رگرسیون برای بیشینه کردن احتمال y به ازای یک مقدار x داده شده (احتمال مطلوب) انتخاب می‌شوند. در زمینه یادگیری ماشین، سیستم با چند بار تکرار، مقادیر بیشینه‌ی احتمال را به دست می‌آورد.

یک مدل رگرسیون خطی با در نظر گرفتن مجموع مقادیر تمام ویژگی های ورودی (متغیرها)، یک خروجی y (مقدار واقعی) را برآورد می‌کند.

مدل مقادیر ضرایب $z, p_1, p_2, p_3, \cdots, p_n$ را تعیین کرده و در نتیجه با کمترین خطا به داده‌های آموزشی سازگاری داده شده است تا خروجی واقعی (y) را پیش‌بینی کند.

از طرفی دیگر، مدل رگرسیون لجستیک با محاسبه‌ی مجموع مقادیر متغیرهای ورودی، یک تابع لجستیک یا سیگموید را در نتیجه اعمال می‌کند. به این ترتیب، تابع غیر خطی یک خروجی دودویی به شکل ۰ یا ۱ (یا حتی "صحیح یا غلط") در اختیار ما قرار می‌دهد.

۲

چندین دلیل وجود دارد که تابع هزینه‌ی میانگین مربعات خطا استفاده شده در رگرسیون خطی، برای رگرسیون لجستیک مناسب نیست:

۱. غیرخطی بودن تابع لجستیک: تابع لجستیک مورد استفاده در رگرسیون لجستیک، غیرخطی است؛ به این معنی که رابطه‌ی بین متغیرهای ورودی و احتمال خروجی یک خط نیست. تابع هزینه‌ی میانگین مربعات خطا رابطه‌ی خطی بین ورودی و خروجی را فرض می‌کند که برای رگرسیون لجستیک مناسب نیست.

۲. مقادیر خروجی احتمالات هستند: در رگرسیون لجستیک، مقادیر خروجی احتمالاتی هستند که بین ۰ و ۱ واقع می‌شوند. تابع هزینه‌ی میانگین مربعات خطا، مقادیر منفی را می‌تواند تولید کند، که احتمالات معنی داری ندارند.

۳. حساسیت به داده های پرت: تابع هزینه‌ی میانگین مربعات خطا، حساس به داده های پرت است که می تواند باعث ایجاد خطاهای بزرگ در تابع هزینه شود. در رگرسیون لجستیک، داده‌های پرت می‌توانند تأثیر قابل توجهی بر احتمالات حاصل شده داشته باشند، لذا استفاده از یک تابع هزینه‌ی کمتر حساس به داده‌های پرت مهم است.

۴. داده‌های نامتعادل: تابع هزینه‌ی میانگین مربعات خطا فرض می‌کند که داده‌ها متعادل هستند. به این معنی که تعداد مثبت و منفی مثال‌ها برابر است. با این حال، در کاربردهای واقعی، داده‌ها اغلب نامتعادل هستند که می‌تواند به پیش‌بینی‌های biasشده منجر شود.

۵. محدب‌نبودن: تابع هزینه‌ی میانگین مربعات خطا، در رگرسیون خطی محدب است. اما در رگرسیون لجستیک محدب نیست. به این معنی که الگوریتم‌های بهینه‌سازی ممکن است در مینیمم‌های محلی گیر کنند و پیدا کردن مینیمم کلی تابع هزینه مشکل باشد.

به همین دلیل، استفاده از تابع هزینه‌ای که به طور خاص برای رگرسیون لجستیک طراحی شده است، مانند تابع هزینه‌ی cross-entropy، بهتر است. تابع هزینه‌ی cross-entropy، قادر به تشخیص رابطه‌ی غیرخطی بین متغیرهای ورودی و احتمالات خروجی

است و کمتر حساس به داده‌های پرت است.

۳

از معادله‌ی اول شروع می‌کنیم:

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \Rightarrow$$
$$\frac{p(X)}{1-p(X)} = \frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{1-\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}} = \frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1}{1+e^{\beta_0+\beta_1 X}}} = e^{\beta_0+\beta_1 X} \Rightarrow$$
$$\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

بنابراین، این دو معادله معادل هستند و رابطه‌ای مشابه بین متغیر وابسته و مستقل در رگرسیون لجستیک را توصیف می‌کنند. به عبارت دیگر، نمایش تابع لجستیک و نمایش لاجیت برای مدل رگرسیون لجستیک معادل هستند.

۴

رگرسیون لجستیک در اصل برای مسائل دسته‌بندی دو کلاسه طراحی شده است. با این حال، روش‌های متعددی برای گسترش رگرسیون لجستیک برای مسائل دسته‌بندی چند کلاسه وجود دارد که در آن متغیر وابسته می‌تواند بیش از دو مقدار ممکن داشته باشد.

۱.۴

روش "one-vs-all" (یا "one-vs-rest") است. در این روش، یک مدل رگرسیون لجستیک برای هر کلاس آموزش داده می‌شود. جایی که متغیر وابسته نشان می‌دهد که آیا یک نمونه به آن کلاس تعلق دارد یا خیر (یعنی برای کلاس مربوطه مقدار ۱ و برای تمام کلاس‌های دیگر مقدار صفر را به خود می‌گیرد). در فاز تست، ما با استفاده از هر مدل، یک مشاهده جدید را دسته‌بندی می‌کنیم و کلاس با بیشترین احتمال پیش‌بینی شده را انتخاب می‌کنیم.

۲.۴

روش دیگر "softmax regression" است. در این روش، ما مدل رگرسیون لجستیک را گسترش می‌دهیم تا به طور مستقیم توزیع احتمال شرطی متغیر وابسته با در نظر گرفتن متغیرهای مستقل را مدل کنیم. به طور خاص، ما از تابع softmax برای تبدیل ترکیب خطی از متغیرهای مستقل به یک توزیع احتمال برای تمام کلاس‌های ممکن استفاده می‌کنیم. در فاز تست، نمونه‌ی جدید را با انتخاب کلاس با بالاترین احتمال پیش‌بینی شده، طبقه‌بندی می‌کنیم.

۵

۱.۵

$$\theta = \begin{bmatrix} 6 \\ -3 \\ 0 \end{bmatrix},$$

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix},$$

$$g(z) = \sigma(z),$$

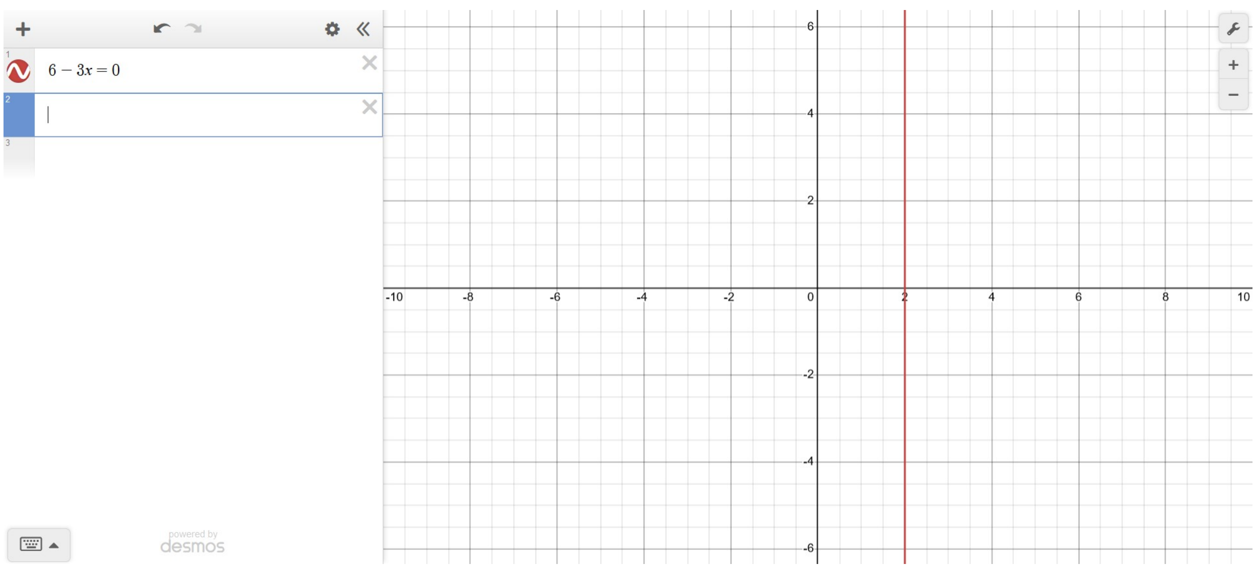
$$h(\theta) = g(\theta^t X),$$

$$\theta^t X = 0 \Rightarrow$$

$$6 - 3x_1 + 0x_2 = 0 \Rightarrow$$

$$x_1 = 2$$

تصویر زیر مرز تصمیم را نشان می‌دهد.



شکل ۱

۲.۵

$$\theta = \begin{bmatrix} 18 \\ 0 \\ 6 \end{bmatrix},$$

$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix},$$

$g(z) = \sigma(z)\,,$

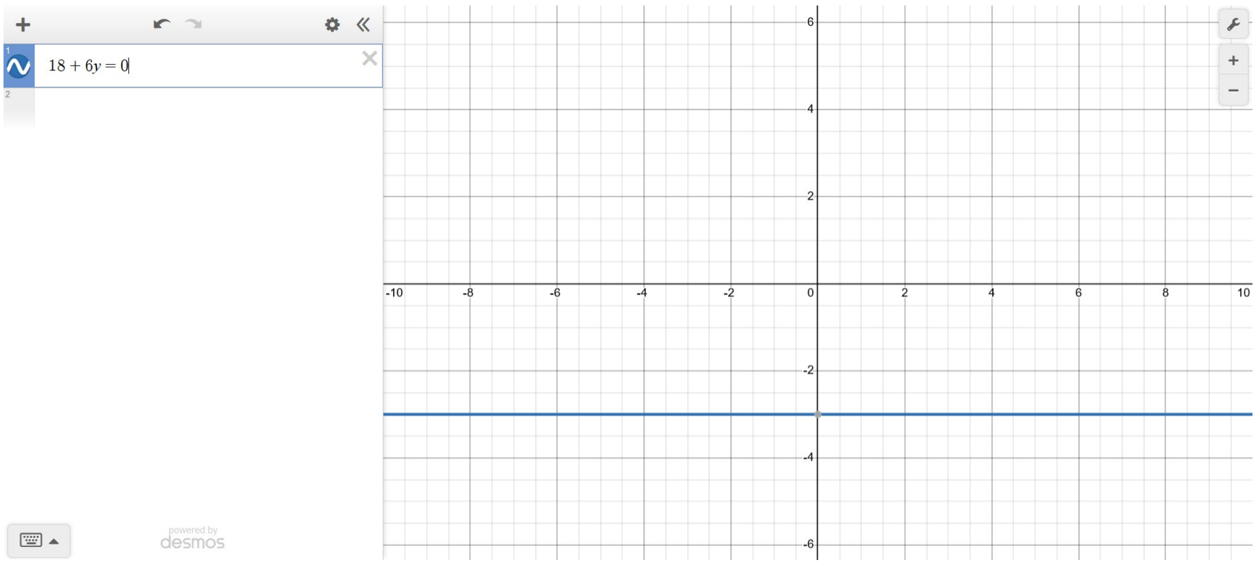
$h(\theta) = g(\theta^t X)\,,$

$\theta^t X = 0 \Rightarrow$

$18 + 0x_1 + 6x_2 = 0 \Rightarrow$

$x_2 = -3$

تصویر زیر مرز تصمیم را نشان می‌دهد.



شکل ۲

۶

$X = 0 \Rightarrow \sigma(\beta_0 + \beta_1 X) = \sigma(\beta_0) = 0.5$

$\Rightarrow \frac{1}{1+e^{\beta_0}} = 0.5$

$\Rightarrow \beta_0 = 0$

$z \geq 0 \Rightarrow \sigma(z) \geq 0.5$

$z \leq 0 \Rightarrow \sigma(z) \leq 0.5$

در نتیجه در مدل سبز رنگ ضریبِ β_1 منفی و در مدل سیاه رنگ ضریبِ β_1 مثبت است.

۷

۱.۷ الف

۱.۱.۷

$z_{1_{D_{a_1}\times 1}} = W_{1_{D_{a_1}\times D_x}} \times x^{(i)}_{D_x\times 1} + b_{1_{D_{a_1}\times 1}}$
$z_{3_{D_{a_1}\times 1}} = W_{2_{D_{a_1}\times D_{a_1}}} \times a_{D_{a_1}\times 1} + b_{2_{D_{a_1}\times 1}}$

بنابراین ابعاد پارامترها به ترتیب برابر $D_{a_1}\times D_x$، $D_{a_1}\times 1$، $D_{a_1}\times D_{a_1}$، و $D_{a_1}\times 1$ است.

۲.۱.۷

$x^{(i)}$ها و $x'^{(i)}$ها در ستون‌های X قرار می‌گیرند. پس ابعاد X برابر است با $2D_x \times m$. همچنین $y^{(i)}$ها در ستون‌های Y قرار دارند. پس ابعاد Y برابر است با $1 \times m$.

۲.۷ ب

۱.۲.۷

$\frac{\partial J}{\partial z_3} = \frac{\partial J}{\partial L^{(i)}}\times\frac{\partial L^{(i)}}{\partial \hat{y}^{(i)}}\times\frac{\partial \hat{y}^{(i)}}{\partial z_3} = -\frac{1}{m}\times\left(\frac{y^{(i)}}{\hat{y}^{(i)}} + \frac{1-y^{(i)}}{1}\times\frac{-1}{1-\hat{y}^{(i)}}\right)\times\sigma\left(z_3\right)\left(1-\sigma\left(z_3\right)\right) = -\frac{1}{m}\times\left(\frac{y^{(i)}}{\hat{y}^{(i)}} + \frac{1-y^{(i)}}{1}\times\frac{-1}{1-\hat{y}^{(i)}}\right)\times$
$\hat{y}^{(i)}\left(1-\hat{y}^{(i)}\right) = \frac{\hat{y}^{(i)}-y^{(i)}}{m}$

۲.۲.۷

$\frac{\partial a}{\partial z_2} = \frac{\partial(a_1-a_2)}{\partial z_2} = \frac{\partial a_1}{\partial z_2} - \frac{\partial a_2}{\partial z_2} = -\frac{\partial a_2}{\partial z_2} = -\frac{\partial(\tanh(z_2))}{\partial z_2} = -\frac{1}{1-z_2^2}$

۳.۲.۷

$\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial L^{(i)}}\times\frac{\partial L^{(i)}}{\partial \hat{y}^{(i)}}\times\frac{\partial \hat{y}^{(i)}}{\partial z_3}\times\frac{\partial z_3}{\partial a}\times\frac{\partial(a_1-a_2)}{\partial W_1} = \frac{\partial J}{\partial L^{(i)}}\times\frac{\partial L^{(i)}}{\partial \hat{y}^{(i)}}\times\frac{\partial \hat{y}^{(i)}}{\partial z_3}\times\frac{\partial z_3}{\partial a}\times\left(\frac{\partial a_1}{\partial z_1}\times\frac{\partial z_1}{\partial W_1} - \frac{\partial a_2}{\partial z_2}\times\frac{\partial z_2}{\partial W_1}\right) =$
$\frac{\hat{y}^{(i)}-y^{(i)}}{m}\times W_2\times\left(\frac{1}{1-z_1^2}\times x^{(i)} - \frac{1}{1-z_2^2}\times x'^{(i)}\right)$

۳.۷ ج

۱.۳.۷

$W_1 := W_1 - \alpha\times\frac{\partial J}{\partial W_1} = W_1 - \alpha\times\frac{\hat{y}^{(i)}-y^{(i)}}{m}\times W_2\times\left(\frac{1}{1-z_1^2}\times x^{(i)} - \frac{1}{1-z_2^2}\times x'^{(i)}\right)$

۲.۳.۷

$b_1 := b_1 - \alpha\times\frac{\partial J}{\partial b_1} = b_1 - \alpha\times\frac{\hat{y}^{(i)}-y^{(i)}}{m}\times W_2\times\left(\frac{1}{1-z_1^2} - \frac{1}{1-z_2^2}\right)$

۳.۳.۷

$W_2 := W_2 - \alpha\times\frac{\partial J}{\partial W_2} = W_2 - \alpha\times\frac{\hat{y}^{(i)}-y^{(i)}}{m}\times a$

۴.۳.۷

$$b_2 := b_2 - \alpha \times \frac{\partial J}{\partial b_2} = b_2 - \alpha \times \frac{\hat{y}^{(i)} - y^{(i)}}{m}$$

۴.۷ د

منابع

□□□□□□□□□□