استاد مربوطه: دکتر فاطمه منصوری — بهار 1403

رگرسیون لجستیک

یادگیری عمیق

گردآورنده: زهرا زارع

مقدمه

رگرسیون لجستیک، یک الگوریتم یادگیری ماشین با نظارت است که بیشتر برای طبقه‌بندی استفاده می‌شود. هدف این است که احتمال وابستگی یک داده به یک کلاس خاص را پیش‌بینی کند. این الگوریتم، که جنبه‌های آماری دارد، به بررسی ارتباط بین متغیرهای مستقل و متغیرهای وابسته دودویی می‌پردازد. این الگوریتم یک ابزار قدرتمند برای تصمیم‌گیری است، مثل تشخیص اینکه آیا یک ایمیل اسپم است یا خیر. این الگوریتم برای طبقه‌بندی به کار می‌رود و به نام رگرسیون لجستیک شناخته می‌شود. علت نامگذاری آن به رگرسیون این است که خروجی تابع رگرسیون خطی را به عنوان ورودی می‌گیرد ولی از تابع سیگموئید برای برآورد احتمال وابستگی به کلاس داده شده استفاده می‌کند. تفاوت بین رگرسیون خطی و رگرسیون لجستیک این است که خروجی رگرسیون خطی یک مقدار پیوسته است که می‌تواند هر چیزی باشد، در حالی که رگرسیون لجستیک احتمال وابستگی یک نمونه به کلاس مشخص یا عدم وابستگی به آن را پیش‌بینی می‌کند.

انواع رگرسیون لجستیک

بر اساس دسته‌بندی‌ها، رگرسیون لجستیک[1] را می‌توان به سه نوع تقسیم کرد:

- **دودویی**[2]: در رگرسیون لجستیک دودویی، تنها دو نوع ممکن برای متغیر وابسته وجود دارد، مانند 0 یا 1، قبولی یا ردی و غیره.
- **چندمتغیره**[3]:در رگرسیون لجستیک چندمتغیره، سه یا بیشتر نوع ممکن غیرمرتب برای متغیر وابسته وجود دارد، مانند گربه، سگ یا گوسفند.
- **ترتیبی**[4]:در رگرسیون لجستیک ترتیبی، سه یا بیشتر نوع ممکن مرتب برای متغیر وابسته وجود دارد، مانند کم، متوسط یا زیاد.

[1] Logistic Regression
[2] Binomial
[3] Multinomial
[4] Ordinal

مفاهیم مرتبط با رگرسیون لجستیک

در اینجا برخی از اصطلاحات رایج مرتبط با رگرسیون لجستیک آورده شده است:

- **متغیرهای مستقل**[5]: ویژگی‌های ورودی یا عوامل پیش‌بینی‌کننده‌ای که برای پیش‌بینی‌های متغیر وابسته به کار می‌روند.
- **متغیر وابسته**[6]: متغیر هدف در مدل رگرسیون لجستیک که تلاش داریم آن را پیش‌بینی کنیم.
- **تابع لجستیک**[7]: فرمولی که استفاده می‌شود تا نشان دهد چگونه متغیرهای مستقل و وابسته با یکدیگر ارتباط دارند. تابع لجستیک متغیرهای ورودی را به یک مقدار احتمال بین ۰ تا ۱ تبدیل می‌کند که احتمال اینکه متغیر وابسته ۱ یا ۰ باشد را نشان می‌دهد.
- **شانس یا احتمال موفقیت**[8]: نسبت وقوع یک رخداد به عدم وقوع آن است. این مفهوم با احتمال متفاوت است زیرا احتمال، نسبت وقوع یک رخداد به همه چیزهایی است که ممکن است رخ دهند.
- **لگاریتم موفقیت (لگاریتم شانس)**[9]: لگاریتم موفقیت که به عنوان تابع لوجیت نیز شناخته می‌شود، لگاریتم طبیعی موفقیت است. در رگرسیون لجستیک، لگاریتم موفقیت متغیر وابسته به عنوان یک ترکیب خطی از متغیرهای مستقل و عرض از مبداء مدل‌سازی می‌شود.
- **ضریب**[10]: پارامترهای تخمین زده شده مدل رگرسیون لجستیک، نشان می‌دهند که چگونه متغیرهای مستقل و وابسته به یکدیگر مرتبط هستند.
- **عرض از مبدا**[11]: یک عبارت ثابت در مدل رگرسیون لجستیک که لگاریتم احتمالات را زمانی که همه متغیرهای مستقل برابر با صفر هستند، نشان می‌دهد.
- **برآورد حداکثر احتمال**[12]: روشی که برای تخمین ضرایب مدل رگرسیون لجستیک استفاده می‌شود، که احتمال مشاهده داده‌ها را با توجه به مدل به حداکثر می‌رساند.

مدل رگرسیون لجستیک

مدل رگرسیون لجستیک، خروجی های مقدار پیوسته تابع رگرسیون خطی را با استفاده از تابع سیگموئید به خروجی‌های دسته‌ای تبدیل می‌کند. این تابع توانایی نگاشت هر مجموعه‌ای از متغیرهای مستقل به یک مقدار بین صفر و یک را دارد و به نام تابع لجستیک شناخته می‌شود.

---

[5] Independent variables
[6] The dependent variable
[7] Logistics function
[8] odds
[9] Logarithm of success
[10] Coefficient
[11] intercept
[12] Maximum likelihood estimation

لذا فرض کنیم ورودی‌های مستقل به صورت زیر باشند:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \cdots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

و متغیر وابسته Y فقط مقادیر دو حالتی یعنی صفر یا یک دارد:

$$Y = \begin{cases} 0 & if\ Class\ 1 \\ 1 & if\ Class\ 2 \end{cases}$$

سپس یک تابع خطی چندگانه را به متغیرهای ورودی X اعمال می‌کنیم:

$$Z = \left(\sum_{i=1}^{n} w_i x_i\right) + b$$

در اینجا x با اندیس i نمونه iام از X است و w در تصویر زیر وزن‌ها یا ضرایب هستند.

$$w_i = [w_1,\ w_2,\ w_3, \ldots\ldots, w_m]$$

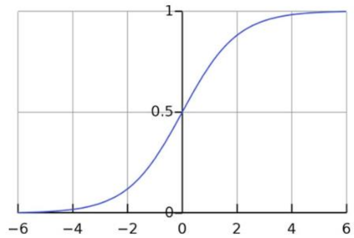
و مقدار b در واقع ترم بایاس است که به عنوان عرض از مبدا نیز شناخته می‌شود. پیش‌بینی خطی مدل قبل از اعمال تابع سیگموئید به سادگی می‌تواند به صورت حاصل‌ضرب نقطه‌ای وزن و بایاس، یعنی به صورت زیر بیان شود:

$$Z = w.X + b$$

آنچه که تا به حال بحث کردیم، رگرسیون خطی بود. برای رگرسیون لجستیک از تابع سیگموئید استفاده می‌کنیم که در آن ورودی Z خواهد بود و ما احتمال را بین 0 و 1 پیدا می‌کنیم، یعنی y پیش‌بینی شده.

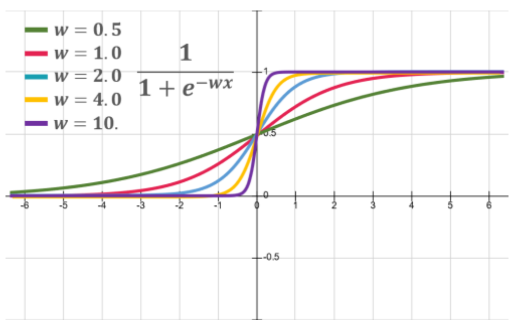
$$\sigma(Z) = \frac{1}{1 - e^{-Z}}$$

نمودار تابع سیگموئید در شکل 1 مشاهده کنید.



شکل 1: نمودار تابع سیگموئید

همانطور که در شکل 1 نشان داده شده است، تابع سیگموئید داده‌های متغیر پیوسته را به احتمال تبدیل می‌کند، یعنی بین 0 و 1. هنگامی که ورودی به سمت مثبت بی نهایت میل می کند، خروجی سیگما به سمت یک میل می‌کند و هنگامی که ورودی به سمت منفی بی نهایت میل می کند خروجی سیگما به سمت صفر میل می‌کند در نتیجه خروجی این تابع همیشه بین 0 و 1 است. در شکل 1، کاملا مشخص است که تابع سیگموید شبیه تابع پله هست. حال اگر بخواهیم حالت گذار در تابع سیگموید را به تابع پله نزدیک‌تر کنیم، کافی است ورژن‌های وزن‌دار آنها را استفاده کنیم (شکل 2 را مشاهده کنید).



شکل2: نسخه وزن دار تابع سیگموید

به ضریب w در رابطه سیگموید دقت کنید. با بزرگتر کردن آن، بیشتر شبیه پله می‌شود. پس خوب است که به مدل یادگیری ماشین‌مان آزادی عمل بدهیم که خودش این مقدار w را تعیین کند. راستی، اگر یک مقدار bهم کنار wx — می‌گذاشتیم، آزادی عمل بیشتر می‌شد و نمودار ما می‌توانست روی محور افقی جابجا شود .
احتمال وابستگی به یک کلاس می‌تواند به صورت زیر اندازه‌گیری شود:

$$P(y = 1) = \sigma(Z)$$

$$P(y = 0) = 1 - \sigma(Z)$$

شانس نشان‌دهنده نسبت احتمال وقوع یک رخداد به احتمال عدم وقوع آن است که متفاوت از احتمال مطلق است. در رگرسیون لجستیک، شانس به صورت زیر محاسبه می‌شود:

$$\frac{p(x)}{1-p(x)} = \sigma(Z)$$

زمانی که ما لگاریتم طبیعی این شانس را حساب می‌کنیم، خروجی به صورت زیرنشان داده می‌شود:

$$Log\left(\frac{p(x)}{1-p(x)}\right) = Z$$

و

$$Log\left(\frac{p(x)}{1-p(x)}\right) = w.X + b$$

با این حساب، معادله نهایی رگرسیون لجستیک که احتمال p(X) را نشان می‌دهد، به صورت زیر نشان داده می‌شود:

$$p(X; b, w) = \frac{e^{w.X+b}}{1+e^{w.X+b}} = \frac{1}{1+e^{-(w.X+b)}}$$

پس یک مدل جدید ساخته شد که الان می‌توانیم با گرادیان کاهشی و تابع اتلاف مانند رگرسیون خطی آنرا آموزش دهیم. یعنی، درواقع ما روی خرووجی رگرسیون خطی یک سیگموید اعمال می‌کنیم و بعد خروجی نهایی بدست می‌آید.

تابع احتمال برای رگرسیون لجستیک

احتمالات پیش‌بینی شده برای y=1 به این شکل p(X; b, w) = p(x) خواهد بود و تابع احتمال

L(b, w) به صورت زیر تعریف می‌شود:

$$\mathrm{L(b, w)} = \prod_{i=1} n p(x_i)^{y_i}(1-p(x_i))^{1-y_i}$$

با گرفتن لگاریتم طبیعی از هر دو طرف، به معادله لگاریتم احتمال می‌رسیم:

$$
\begin{aligned}
\mathrm{L(b,w)} = \log\big(\mathrm{L(b,w)}\big) &= \sum_{i=1}^{n} y_i \log p(x_i) + (1-y_i)\big(1-p(x_i)\big) \\
&= \sum_{i=1}^{n} y_i \log p(x_i) + (1)\big(1-p(x_i)\big) - y_i\big(1-p(x_i)\big) \\
&= \sum_{i=1}^{n} \log(1-p(x_i)) \;\; + \;\; \sum_{i=1}^{n} y_i \log\left(\frac{p(x_i)}{1-p(x_i)}\right) \\
&= \sum_{i=1}^{n} -\log 1 - e^{-(w.X+b)} \;\; + \;\; \sum_{i=1}^{n} y_i(w.x_i + b)
\end{aligned}
$$

برای یافتن برآوردهای حداکثر احتمال، نسبت به w مشتق می‌گیریم.

$$
\begin{aligned}
\frac{\partial J(\mathrm{L(b,w)})}{\partial w_j} &= -\sum_{i=1}^{n} \frac{1}{1+e^{(w.x_i+b)}} e^{(w.x_i+b)}\, x_{ij} + \;\; \sum_{i=1}^{n} y_i x_{ij} \\
&= -\sum_{i=1}^{n} p(\,x_i; b, w) x_{ij} + \sum_{i=1}^{n} y_i x_{ij} \\
&= \; -\sum_{i=1}^{n} (y_i - p(\,x_i; b, w) x_{ij}
\end{aligned}
$$

تابع هزینه در رگرسیون لجستیک

مدل ساخته شد، اما مساله دیگر این هست که ما تابع اتلافی داریم که برای مساله رگرسیون مناسب هست. چرا MSEبرای رگرسیون مناسب هست؟ چون در رگرسیون خطی، مدل می‌گفت با این داده x_i، من خروجی را 1.2 پیش‌بینی کرده‌ام. تابع اتلاف، اختلاف بین مقدار پیش‌بینی و واقعی را محاسبه می‌کند. اما تابع اتلاف در دسته بندی یا همین رگرسیون لجستیک متفاوت است. همان‌طور که پیش از این گفته شد ، در دسته بندی معمولا از تابع اتلاف کراس آنتروپی استفاده می‌شود. تابع اتلاف کراس آنتروپی به صورت زیر تعریف می‌شود:

$$
LOSS = \begin{cases} -log y_p & y_t = 1 \\ -log 1 - y_p & y_t = 0 \end{cases}
$$

یک تابع دو ضابطه‌ای داریم:

- اگر لیبل واقعی یک داده 1 هست (y_t=1) ، از ضابطه $-log y_p$ استفاده کنید.
- اگر هم صفر هست (y_t=0) از ضابطه پایینی $-log 1 - y_p$

به راحتی می‌توانیم این تابع دو ضابطه‌ای را یکپارچه کرده و به شکل زیر بنویسیم:

$$L = -(y_t \, log y_p + (1 - y_t)(log 1 - y_p)$$

طبیعتا N نمونه داریم. پس قاعدتا باید اتلاف تک‌تکشان را حساب کنیم و بعد میانگین بگیریم. خب، میانگین اتلاف L به شکل زیر محاسبه می‌شود:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(y_t \, log y_p + (1 - y_t)(log 1 - y_p)$$

به رابطه بالا تابع اتلاف رگرسیون لجستیک گفته می‌شود و همچنین توجه کنید که از آنجا که L یک تابع غیرخطی از x_i ورودی است، تابع هزینه دیگر یک تابع کوژ نخواهد بود و برای تعیین پارامترهای رگرسیون لجستیک از گرادیان کاهشی استفاده می‌کنیم.

رفرنس

Book_Deep Learning_Ian Goodfellow