

به نام خدا

## **Logistic regression**

Logistic regression در اوایل قرن بیستم در علوم بیولوژیکی و سپس در بسیاری از کاربردهای علوم اجتماعی مورد استفاده قرار گرفت. Logistic regression وقتی استفاده میشود که متغیر وابسته یا هدف categorical باشد. مثلاً پیش بینی اینکه یک ایمیل اسپم است (1) یا نه (0). یا یک تومور بدخیم است (1) یا نه (0). یک سناریو را در نظر بگیرید که در آن نیاز داریم که دسته بندی کنیم که یک ایمیل اسپم است یا نه. اگر از رگرسیون خطی (linear regression) استفاده کنیم باید یک آستانه برای آن تعیین کنیم تا براساس آن دسته بندی انجام شود. اما رگرسیون خطی unbounded میباشد (یک خط میکشد که حتی میتواند مقدار منفی هم پیدا کند) و به همین دلیل رگرسیون لجستیک وارد صحنه میشود که مقادیر آن بین 0 تا 1 میباشد.

### **Simple logistic regression:**

#### **Model**

Output = 0 or 1

Hypothesis =>  $Z = WX + B$

$h_{\Theta}(x) = \text{sigmoid}(Z)$

#### **Sigmoid Function**

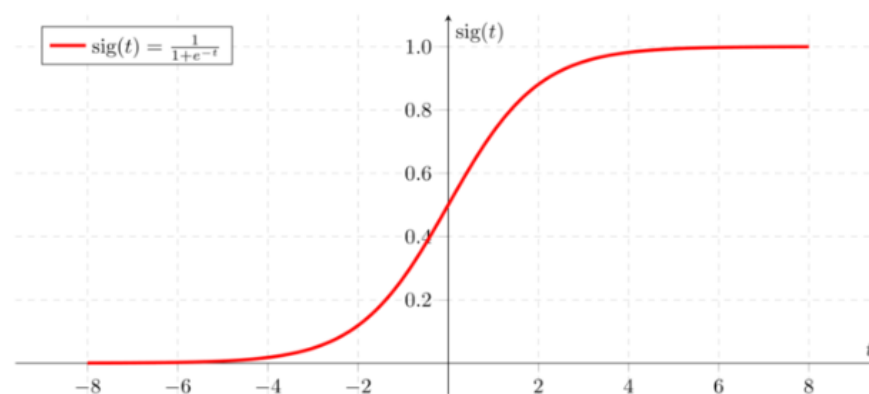


Figure 2: Sigmoid Activation Function

اگر  $Z$  به سمت بی نهایت برود مقدار  $Y$  (مقدار پیش بینی شده) 1 میشود و اگر  $Z$  به سمت منفی بی نهایت برود مقدار  $Y$  برابر با 0 میشود. مقدار خروجی این hypothesis احتمال تخمین زده شده میباشد که از آن استفاده میشود تا استنتاج کنیم با چه اطمینانی مقدار پیش بینی شده میتواند مقدار واقعی باشد وقتی ورودی  $X$  داده شده است.

برای مثال:

$$X = [x_0 \ x_1] = [1 \ \text{IP-Address}]$$

بر مبنای  $x_1$  اجازه بدهید بگوییم که احتمال تخمین زده شده 0.8 را به دست آورده ایم. این یعنی 80 درصد شانس برای اینکه ایمیل اسپم باشد. به صورت ریاضی میتوان این ها را به صورت زیر نوشت:

$$h_{\theta}(x) = P(Y=1|X; \theta)$$

Probability that  $Y=1$  given  $X$  which is parameterized by 'theta'.

$$P(Y=1|X; \theta) + P(Y=0|X; \theta) = 1$$

$$P(Y=0|X; \theta) = 1 - P(Y=1|X; \theta)$$

انواع لجستیک رگرسیون ها:

## 1. Binary Logistic Regression

پاسخ دسته دو مقدار ممکن دارد. مثلا اسپم هست یا نیست.

## 2. Multinomial Logistic Regression

سه دسته با بیشتر بدون ترتیب. مثلا اینکه کدام غذا بیشتر ترجیح داده میشود. (Veg, Non-Veg, Vegan)

## 3. Ordinal Logistic Regression

سه دسته یا بیشتر با ترتیب. مثلا رتبه یک فیلم از 1 تا 5.

برای اینکه پیش بینی کنیم داده به چه کلاسی تعلق دارد میتوانیم یک آستانه تنظیم کنیم. بر مبنای این آستانه احتمال تخمین زده شده به دست آمده به کلاسش دسته بندی میشود. مثلاً اگر مقدار پیش بینی شده از 0.5 بیشتر مساوی باشد ایمیل اسپم میباشد در غیر این صورت اسپم نیست.

**Cost function:**

$$\text{Cost}(h_{\theta}(x), Y(\text{actual})) = -\log(h_{\theta}(x)) \text{ if } y=1$$

$$-\log(1 - h_{\Theta}(x)) \text{ if } y=0$$

### *Deriving the formula for Gradient Descent Algorithm:*

## Gradient

$$\begin{aligned}
 z &= w_1 x_1 + w_2 x_2 + b \rightarrow \hat{y} = a = \sigma(z) \rightarrow L(\hat{y}, y) \\
 &\Leftrightarrow a = \hat{y} \\
 w_1 &\Rightarrow \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w_1} \\
 \frac{\partial L}{\partial a} &= \frac{\partial}{\partial a} (-y \log a - (1-y) \log(1-a)) \\
 &= -y \left( \frac{1}{a} \right) - (-1) \left( \frac{1-y}{1-a} \right) \\
 \frac{\partial L}{\partial a} &= \left( \frac{-y}{a} \right) + \left( \frac{1-y}{1-a} \right) \\
 \frac{\partial a}{\partial z} &= a(1-a) \\
 \frac{\partial z}{\partial w_1} &= x_1
 \end{aligned}$$

$$\frac{\partial L}{\partial w_1} = \left( \left( \frac{-y}{a} + \frac{(1-y)}{1-a} \right) \cdot (a)(1-a) \right) \cdot x_1$$

$$= (a-y) \cdot x_1$$

Update for  $w_1$ ,

$$\frac{\partial L}{\partial w_1} = (a-y) \cdot x_1$$

$$\text{Here, } (a-y) = \frac{\partial L}{\partial z}$$

$$\boxed{w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1}}$$

Similarly, for all parameters

$$\boxed{w_i = w_i - \alpha \frac{\partial L}{\partial w_i}}$$

$$i = 1, 2, \dots, m$$

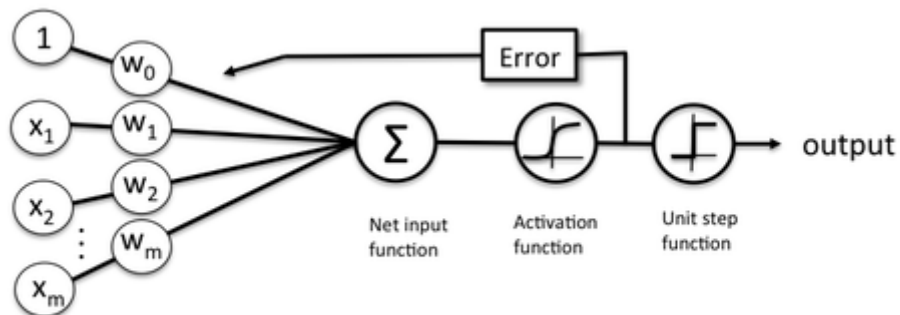
$m$  = no. of parameters

$$\boxed{b = b - \alpha \frac{\partial L}{\partial b}}$$

$$\text{where, } \frac{\partial L}{\partial b} = (a-y)$$

## ارتباط لجستیک رگرسیون و شبکه عصبی:

اساسا به لجستیک رگرسیون میتوان به عنوان یک شبکه عصبی یک لایه نگاه کرد.



Schematic of a logistic regression classifier.

در واقع خیلی معمول است که از تابع سیگموئید لجستیک به عنوان یک تابع فعال سازی در لایه مخفی شبکه عصبی استفاده کرد. در شبکه های عصبی میتوان از توابع فعال سازی دیگری مثل  $\tanh$  استفاده نمود که در عمل بهتر است زیرا در لایه های مخفی به به خروجی مثبت محدود نمیباشد.

یکی از ویژگی های خوب رگرسیون لجستیک این است که تابع هزینه اش محدب است پس ما تضمین میکنیم که میتوانیم مینیمم سراسری هزینه را پیدا کنیم در حالی که ما تابع فعال سازی لجستیک را در شبکه عصبی چند لایه انباشته میکنیم و محدب بودن را از دست میدهیم و باعث میشود الگوریتم بتواند در مینیمم های محلی گیر کند. با این حال ما همچنان با یک مدل پیش بینی کننده قوی روبرو هستیم.

## چه زمانی از کدام الگوریتم استفاده کنیم:

برای مسائل دسته بندی پیشنهاد میشود در ابتدا از مدل های ساده مثل لجستیک رگرسیون استفاده شود. این الگوریتم در بعضی مواقع مسئله شما را به قدر کافی خوب حل میکنند ولی اگر از عملکرد آن راضی نبودید و به قدر کافی داده آموزشی داشتید (همانطور که در مقاله اشاره شده در تمرین گفته شده شبکه های عصبی منعطف تر از لجستیک رگرسیون ها هستند بنابراین مستعد **overffiting** میباشند پس باید به قدر کافی داده در اختیار داشته باشیم) از شبکه عصبی استفاده کنید که مزیت یادگیری تابع های پیچیده تر و غیرخطی تر را دارند.

منابع: لینک های داده شده در سوال 2 تمرین