درس‌نامه هفتم

در مساله دسته بندی بیز مشاهده کردیم که بهترین تصمیم گیری از رابطه

$$\hat{y} = \arg\max_{k} \mathbb{P}[y = k \mid x]$$

به دست می آید. برای حل مساله فوق به توزیع مشترک (y, x) نیازمند بودیم که به طور عموم در اختیار نمی باشد. در راه حل LDA فرض کردیم که توزیع مشترک گوسی است و از روی داده‌ها پارامترها گوسی را تخمین زدیم. یک راه سر راست‌تر آن است که برای $\mathbb{P}[y \mid x]$ یک مدل در نظر بگیریم و از روی داده پارامترها را تخمین زده و سپس از رابطه بیز استفاده کنیم. در روش رگرسیون لوجستیک $\mathbb{P}[y \mid x]$ به صورت زیر مدل می‌گردد:

$$\mathbb{P}[y = k \mid x] = \frac{e^{\beta_k^T x}}{\mathcal{Z}}$$

در رابطه فوق درایه اول x مقدار ۱ می‌باشد و مقدار $\mathcal{Z}$ برای نرمالایز کردن می‌باشد. به راحتی می‌بینیم که

$$\mathcal{Z} = \sum_{k=1}^{K} e^{\beta_k^T x}$$

برای یافتن پارامترهای مساله می‌توانیم از روش بیشینه احتمال (ML) استفاده نماییم. بدین منظور برای داده‌ها

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

رابطه زیر را در نظر می گیریم:

$$\hat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} log \mathbb{P}_\beta[y_i \mid x_i]$$

در رابطه فوق $\beta = (\beta_1, \dots, \beta_k)$ محتوی تمامی پارامترهای مساله می‌باشد.

حالت ساده

فرض می‌نماییم دو دسته وجود دارد و ورودی یک بعدی است. در این حالت

$$\begin{aligned}
\mathbb{P}[y = ۱ \mid x] &= \frac{e^{\beta_۰ + \beta_۱ x}}{\mathcal{Z}} \\
\mathbb{P}[y = ۲ \mid x] &= ۱ - \mathbb{P}[y = ۱ \mid x] \\
&= \frac{\mathcal{Z} - e^{\beta_۰ + \beta_۱ x}}{\mathcal{Z}} \\
&= \frac{۱}{\mathcal{Z}}
\end{aligned}$$

در رابطه فوق $\mathbb{Z} = ۱ + e^{\beta_۰ + \beta_۱ x}$ در نظر گرفتیم تا محاسبات ساده گردند. در این شرایط

$$\mathbb{P}[y = ۱ \mid x] = \frac{e^{\beta_۰ + \beta_۱ x}}{۱ + e^{\beta_۰ + \beta_۱ x}}^{۱}$$

و برای تخمین $\beta_۰, \beta_۱$ باید رابطه زیر را حل نماییم:

$$(\hat{\beta}_۰, \hat{\beta}_۱) = \arg\max_{(\beta_۰, \beta_۱)} \sum_{i: y_i = ۱} log \frac{e^{\beta_۰ + \beta_۱ x_i}}{۱ + e^{\beta_۰ + \beta_۱ x_i}} + \sum_{i: y_i = ۲} log \frac{۱}{۱ + e^{\beta_۰ + \beta_۱ x_i}}$$

اگر برچسب‌ها را با صفر و یک نمایش دهیم: یعنی $y \in \{۰, ۱\}$ آنگاه می‌توان رابطه فوق را به صورت زیر نوشت:

$$\begin{aligned}
(\hat{\beta}_۰, \hat{\beta}_۱) &= \arg\min_{(\beta_۰, \beta_۱)} \sum y_i log \frac{۱}{f_i} + (۱ - y_i) log \frac{۱}{۱ - f_i} \\
&= \arg\min_{(\beta_۰, \beta_۱)} D(y \mid\mid f)
\end{aligned}$$

عبارت بالا تابع هزینه ما را تشکیل می‌دهد که در آن $f_i = \frac{e^{\beta_۰ + \beta_۱ x_i}}{۱ + e^{\beta_۰ + \beta_۱ x_i}}$
با فرض حل مساله فوق کافیست برای تصمیم‌گیری رابطه زیر را محاسبه کنیم و با مقدار صفر مقایسه نماییم:

$$log \frac{\mathbb{P}[y = ۱ \mid x]}{\mathbb{P}[y = ۲ \mid x]} = \hat{\beta}_۰ + \hat{\beta}_۱ x$$

تابع سمت چپ را تابع log odds یا logit می‌نامیم.

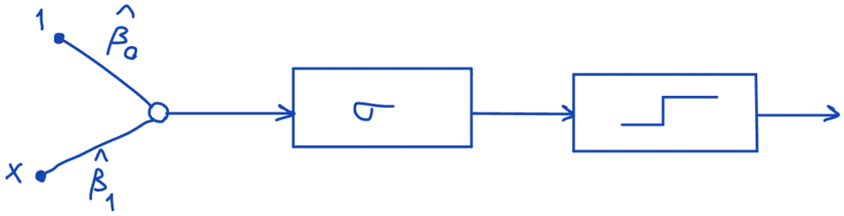
$$\operatorname{logit} p = log \frac{p}{۱ - p}$$

تابع معکوس logit تابع لوجستیک می‌باشد که به صورت زیر بیان می‌شود:

$$\sigma(x) = \frac{e^x}{۱ + e^x}$$

[۱] تابع لجستیک $\frac{e^x}{۱ + e^x} = \frac{۱}{۱ + e^{-x}}$

می‌توانیم روش رگرسیون لوجستیک را با شکل زیر نمایش دهیم:



که یک شبکه عصبی می‌باشد.

تعمیم مساله دو دسته‌بندی به حالتی که ورودی چند بعدی است کار آسانی می‌باشد.

حالت چند دسته‌ای $K > ۲$

در این حالت تنها نکته آن است که همچون حالت دوتایی برای آخرین دسته پارامتر در نظر نمی‌گیریم. به طور خاص

$$\mathbb{P}[y = k \mid x] = \frac{e^{\beta_k^T x}}{\mathcal{Z}} \quad k = ۱, \ldots, K - ۱$$
$$\mathbb{P}[y = K \mid x] = \frac{۱}{\mathcal{Z}}$$

در این شرایط

$$\mathcal{Z} = ۱ + \sum_{k=۱}^{K-۱} e^{\beta_k^T x}$$

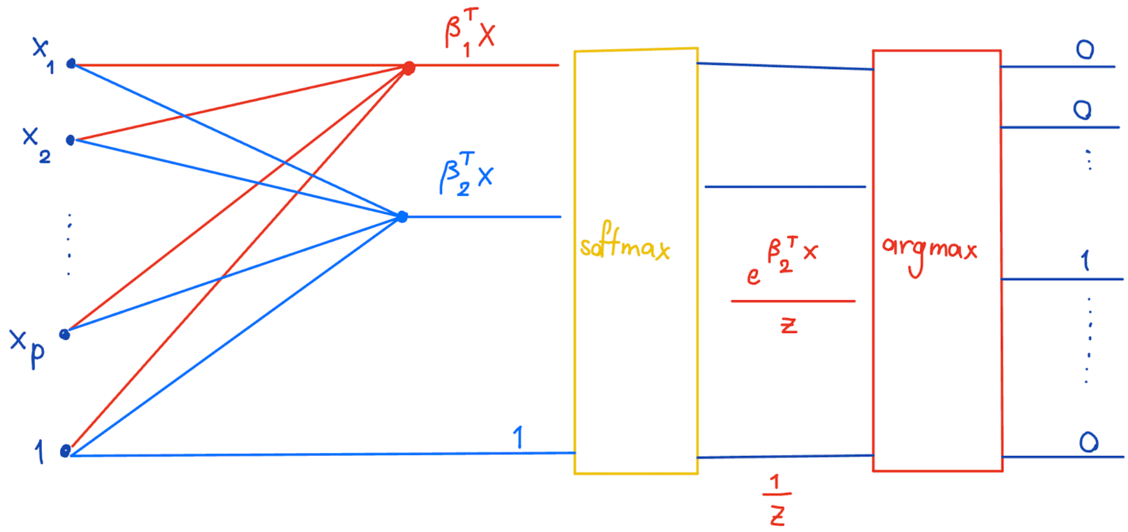
همچنین داریم:

$$log \frac{\mathbb{P}[y = k \mid x]}{\mathbb{P}[y = K \mid x]} = \beta_k^T x$$

و در نتیجه:

$$log \frac{\mathbb{P}[y = k \mid x]}{\mathbb{P}[y = k' \mid x]} = (\beta_k - \beta_{k'})^T x$$

می‌توانیم شبکه چند دسته‌ای را به صورت زیر ترسیم کنیم:



نکته: اگر می‌خواستیم که جداکننده خطی داشته باشیم و از قانون بیز هم استفاده کنیم آنگاه به رگرسیون لجستیک می‌رسیدیم. در یک مساله دوکلاسه بایستی تابع logit را با آستانه صفر مقایسه کنیم و اگر بخواهیم رابطه خطی باشد آنگاه به رابطه زیر می‌رسیم:

$$log\frac{\mathbb{P}(y=۱\mid x)}{۱-\mathbb{P}(y=۲\mid x)}=\beta^T x$$

و در نتیجه:

$$\mathbb{P}(y=۱\mid x)=\frac{e^{\beta^T x}}{۱+e^{\beta^T x}}$$