05 Logistic Regression

1) 极大似然法

$$c = \underset{c \in C}{\mathrm{argmax}}\ p(d|c)\,p(c)$$

目标分类 = 使概率取最大时的目标分类

$p(d|c)p(c)$ 是事件在 c 条件下达到 d 结果的概率.

2) Logistic 函数.

$$\begin{cases} p(y=1) = \dfrac{1}{1+e^{-(wx+b)}} \\ p(y=0) = \dfrac{e^{-(wx+b)}}{1+e^{-(wx+b)}} \end{cases}$$

3) 例1:

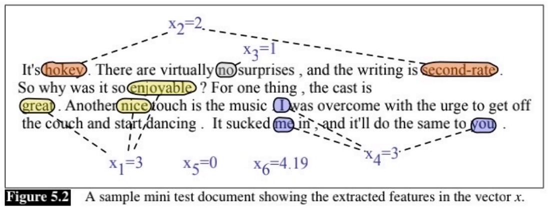

Figure 5.2 A sample mini test document showing the extracted features in the vector x.

- 从文章中摘取特征得到 $\vec{x}$.
- 然后利用 $\vec{w}\vec{x}+b$ 标准化
- 输入到 Sigmoid 函数中. 进行判定

所以, 从文本中抽取得到 x 是很麻烦的, 也是关键.

4) 信息与信息熵

信息是用来描述消除不确定性能力的量,
信息越大, 获得信息后, 不确定性消除的就越大

$$I(x) = -\log p(x)$$

事件的发生, 并非完全是无法捉摸的纯随机事件.
而是携带着某种信息/动向

信息熵是对整个事件(而非某某一次观察)的描述.
它是对事件的评价, 不是对某次事件结果的评价.

5) 相对熵(KL散度)

对于同一个随机变量 X 的两个单独的分布 $P(X)$ 和 $Q(X)$.
用此描述两个概率分布的差异

$$D_{KL}(p\|q) = \sum_{i=1}^{n} p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right)$$

KL散度越小, 两分布越接近
可用作目标函数训练两个相似的分布

6) 交叉熵

$$D_{KL}(p\|q)=\sum_{i=1}^{n}p(x_i)\log p(x_i)-\sum_{i=1}^{n}p(x_i)\log q(x_i)$$

$$=-H(p(x_i))+\underbrace{\left[-\sum_{i=1}^{n}p(x_i)\log q(x_i)\right]}_{\text{交叉熵}}$$

7) 多元 logistic regression

$$softmax(z_i)=\frac{e^{z_i}}{\sum_{j=1}^{k}e^{z_j}}\qquad 1\le i\le k$$

对于输入向量 $z=[z_1, z_2\cdots z_k]$

$$softmax(z)=\left[\frac{e^{z_1}}{\sum_{j=1}^{k}e^{z_j}},\frac{e^{z_2}}{\sum_{j=1}^{k}e^{z_j}},\cdots,\frac{}{\sum_{j=1}^{k}e^{z_j}}\right]$$