logistic 回归

判别模型，拟合的是 $P(Y|X)$，X 没有随机性。

二项 logistic 回归

利用一个线性模型去拟合 logit 变换

输出 $Y \in \{0,1\}$

$$P(Y=1|X) = \pi(X)$$

$$\text{logit}\,\pi(X) = \log\frac{P(Y=1|X)}{P(Y=0|X)} = \log\frac{\pi(X)}{1-\pi(X)} = w\cdot x$$

$$\pi(X) = \frac{\exp(w\cdot x)}{1+\exp(w\cdot x)}$$

过拟合：当 $w\cdot x$ 趋于 $+\infty$ 时，$\pi(x)$ 趋近于 1，导致模型"过分自信"

多项 logistic 回归

同样是利用一个线性模型去拟合 logit 变换

$Y \in \{1,2,3,\cdots K\}$

$$P(Y=k|X) = \pi_k(X)$$

$$\text{logit}\,\pi_k(X) = \log\frac{P(Y=k|X)}{P(Y=K|X)} = w_k\cdot x$$

$$P(Y=k|X) = e^{w_k\cdot x}P(Y=K|X)$$

$$\sum_{k=1}^{K}P(Y=k|X) = \sum_{k=1}^{K-1}e^{w_k\cdot x}P(Y=K|X) + P(Y=K|X)$$

$$= P(Y=K|X)\left(\sum_{k=1}^{K-1}e^{w_k\cdot x}+1\right)$$

$$P(Y=k|X) = \frac{1}{1+\sum_{k=1}^{K-1}e^{w_k\cdot x}}$$

$$P(Y=k|X) = \frac{e^{w_k\cdot x}}{1+\sum_{k=1}^{K-1}e^{w_k\cdot x}}$$

学习算法　极大似然估计 → 估计 W_k(多分类)/W(二分类)

共有N个样本，以二分类为例　$P(Y=1|x)=\pi(x)$

$$L(w)=\log\prod_{i=1}^{N}P(Y=y_i|x_i)$$
$$=\log\prod_{i=1}^{N}[\pi(x_i)]^{y_i}[1-\pi(x_i)]^{1-y_i}$$
$$=\sum_{i=1}^{N}y_i\log\pi(x_i)+(1-y_i)\log(1-\pi(x_i))$$
$$=\sum_{i=1}^{N}y_i\log\frac{\pi(x_i)}{1-\pi(x_i)}+\log(1-\pi(x_i))$$
$$=\sum_{i=1}^{N}y_i(w\cdot x_i)-\log(1+e^{w\cdot x_i})$$ ← 梯度下降法求解该函数的极值

最大熵模型

最大熵原理：在满足约束条件的模型集合中选择熵最大的模型

{有输入 条件熵
无输入 熵

最大熵模型：

给定训练集 $T=\{(x_1,y_1),\dots(x_N,y_N)\}$，以及特征函数 $f_i(x,y)$，$i=1,2,\dots n$.

学习最大熵模型等价于求解如下约束最优化问题：

$$\max_{P\in C}H(P)=-\sum_{x,y}\tilde{P}(x)P(y|x)\log P(y|x)$$ ← 条件熵

← 等于 $P(x,y)$　$\tilde{P}(x)$ 为 x 的经验分布

$$s.t.\quad E_P(f_i)=E_{\tilde{P}}(f_i)$$
$$\sum_y P(y|x)=1$$

$$E_P(f_i)=\sum_{x,y}\tilde{P}(x)P(y|x)f_i(x,y)\qquad E_{\tilde{P}}(f_i)=\sum_{x,y}\tilde{P}(x,y)f_i(x,y)$$

$\tilde{P}(x,y)$ 为 x 与 y 的联合经验分布

上述问题可通过拉格朗日对偶性转化

有约束的最优化问题利用拉格朗日函数进行转换

$$\min L(P,W)=-H(P)+w_0\left(1-\sum_y P(y|x)\right)+\sum_{i=1}^{n}w_i(E_P(f_i)-E_{\tilde{P}}(f_i))$$

（极小极大问题）

等价于 $\min_{P\in C}\max_{w}-H(P)+w_0\left(1-\sum_y P(y|x)\right)+\sum_{i=1}^{n}w_i(E_P(f_i)-E_{\tilde{P}}(f_i))$

对偶问题

例子：$\max_{w} \min_{p \in C} -H(p) + w_0(1 - \sum_x p(y|x)) + \sum_{i=1}^{n} w_i(E_p(f_i) - E_{\tilde{p}}(f_i))$

$$\Psi(w) = \min_{p \in C} -H(p) + w_0(1 - \sum_x p(y|x)) + \sum_{i=1}^{n} w_i(E_p(f_i) - E_{\tilde{p}}(f_i))$$

$$= \min_{p \in C} \sum_{x,y} \tilde{p}(x) \log p(y|x) \; p(y|x) + w_0(1 - \sum_y p(y|x)) +$$

$$\sum_{i=1}^{n} w_i \left[\sum_{x,y} \tilde{p}(x,y) f_i(x,y) - \sum_{x,y} \tilde{p}(x) p(y|x) f_i(x,y) \right]$$

$$= \min_{p \in C} L(p, w)$$

$$\frac{\partial L(p,w)}{\partial p(y|x)} = \sum_{x,y} \tilde{p}(x)(1 + \log p(y|x)) - \sum_y w_0 - \sum_{i=1}^{n} w_i \sum_{x,y} \tilde{p}(x) f_i(x,y)$$

$$= \sum_{x,y} \tilde{p}(x)(1 + \log p(y|x)) - \sum_{x,y} \tilde{p}(x) w_0 - \sum_{x,y} \tilde{p}(x) \sum_{i=1}^{n} w_i f_i(x,y)$$

$$= \sum_{x,y} \tilde{p}(x) \left[1 + \log p(y|x) - w_0 - \sum_{i=1}^{n} w_i f_i(x,y) \right]$$

$= 0$

$$1 + \log p(y|x) - w_0 - \sum_{i=1}^{n} w_i f_i(x,y) = 0$$

$$\log p(y|x) = w_0 + \sum_{i=1}^{n} w_i f_i(x,y) - 1$$

$$p(y|x) = \frac{\exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right)}{\exp(1 - w_0)}$$

$\sum_y p(y|x) = 1 \Rightarrow \sum_y \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right) = \exp(1 - w_0)$

$$P_w(y|x) = \frac{\exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right)}{\sum_y \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right)} = \frac{\exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right)}{Z_w(x)}$$

上述模型就是"要求的最大熵模型"，可以看到还有要优化的参数 w_i.

这些参数就是我们要通过最优化算法进一步求解的

↳ 梯度下降，迭代尺度法，拟牛顿法

首先，我们需要定义这个最优化问题，两种方法：极大似然估计或者是上述的对偶问题

① 极大似然估计

$$L_{\tilde{P}}(P) = \log \prod_{x,y} P(y|x)^{\tilde{P}(x,y)}$$

$$= \sum_{x,y} \tilde{P}(x,y) \log P(y|x)$$

$$= \sum_{x,y} \tilde{P}(x,y) \left[\log \exp \sum_{i=1}^{n} w_i f_i - \log Z_w(x)\right]$$

$$= \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i - \sum_{x,y} \tilde{P}(x,y) \log Z_w(x)$$

$$= \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i - \sum_{x} \log Z_w(x) \sum_{y} \tilde{P}(x,y)$$

$$L(w) = \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i - \sum_{x} \log Z_w(x) \tilde{P}(x)$$

$$w_i^* = \max_{w_i} \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i - \sum_{x} \log Z_w(x) \tilde{P}(x)$$

迭代尺度法推导

$w \to w + \delta$

$$L(w+\delta) - L(w) = \sum_{x,y} \tilde{P}(x,y) \left[\sum_{i=1}^{n} (w_i + \delta_i) f_i - w_i f_i\right]$$

$$- \sum_{x} \tilde{P}(x) \left[\log Z_{w+\delta}(x) - \log Z_w(x)\right]$$

$$= \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} \delta_i f_i - \sum_{x} \tilde{P}(x) \log \frac{Z_{w+\delta}(x)}{Z_w(x)}$$

$-\log \alpha \geq 1 - \alpha$

$$-\sum_{x} \tilde{P}(x) \log \frac{Z_{w+\delta}(x)}{Z_w(x)} \geq \sum_{x} \tilde{P}(x) \left[1 - \frac{Z_{w+\delta}(x)}{Z_w(x)}\right]$$

$$\geq \sum_{x} \tilde{P}(x) - \sum_{x} \tilde{P}(x) \frac{Z_{w+\delta}(x)}{Z_w(x)}$$

$$\geq 1 - \sum_{x} \tilde{P}(x) \frac{Z_{w+\delta}(x)}{Z_w(x)}$$

$$\frac{Z_{w+\delta}(x)}{Z_w(x)} = \frac{\sum_y \exp\left(\sum_{i=1}^n (w_i+\delta_i) f_i\right)}{Z_w(x)}$$

$$= \frac{\sum_y \exp \sum_{i=1}^n w_i f_i \cdot \exp \sum_{i=1}^n \delta_i f_i}{Z_w(x)}$$

$P_w(y|x)$

$$= \sum_y P_w(y|x) \exp \sum_{i=1}^n \delta_i f_i$$

$$L(w+\delta) - L(w) \geq \sum_{x,y} \tilde{p}(x,y) \sum_{i=1}^n \delta_i f_i + 1 - \sum_x \tilde{p}(x) \sum_y P_w(y|x) \exp \sum_{i=1}^n \delta_i f_i$$

Jeson不等式：φ为凸函数，$\sum \alpha_i = 1$.

对有 $\varphi(\sum \alpha_i x_i) \leq \sum \alpha_i \varphi(x_i)$

$$\exp \sum_{i=1}^n \delta_i f_i = \exp \sum_{i=1}^n \frac{f_i}{f^\#(x,y)} f^\#(x,y) \delta_i \quad (f^\#(x,y) = \sum_{i=1}^n f_i)$$

$$\leq \sum_{i=1}^n \frac{f_i}{f^\#(x,y)} e^{f^\#(x,y)\delta_i}$$

$$L(w+\delta) - L(w) \geq \sum_{x,y} \tilde{p}(x,y) \sum_{i=1}^n \delta_i f_i + 1 - \sum_x \tilde{p}(x) \sum_y P_w(y|x) \sum_{i=1}^n \frac{f_i}{f^\#(x,y)} e^{f^\#(x,y)\delta_i}$$

下界可最大化对 δ_i 求偏导

$$\frac{\partial (L(w+\delta) - L(w))}{\partial \delta_i} = \sum_{x,y} \tilde{p}(x,y) f_i - \sum_x \tilde{p}(x) \sum_y P_w(y|x) f_i e^{f^\#(x,y)\delta_i}$$

$$= 0.$$