第三章 线性模型

$$f(x) = w_1x_1 + w_2x_2 + \cdots w_d x_d + b = w^T x + b$$

3.2 线性回归

离散属性：存在"序"关系 → 连续值

身高 → $\{0.0, 1.0\}$

无"序"关系 → k维向量

(x_i, y_i)：$f(x_i) = w x_i + b$ 使 $f(x_i) \simeq y_i$

均方误差：
$$(w^*, b^*) = \arg\min \sum_{i=1}^{m} (f(x_i) - y_i)^2 = \arg\min \sum_{i=1}^{m} (y_i - w x_i - b)^2$$

多元线性回归：$f(x) = \hat{w} x$

$$\hat{w}^* = \arg\min (y - X\hat{w})^T (y - X\hat{w})$$

$$\frac{\partial E}{\partial \hat{w}} = 2X^T (X\hat{w} - y)$$

广义线性模型：$y = g^{-1}(w^T x + b)$　$g(\cdot)$ 单调可微 → "联系函数"

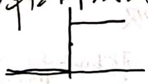
对数线性回归：$g(\cdot) = \ln(\cdot)$

采用加权最小二乘法或极大似然估计

3.3 对数几率回归

用于分类任务

① 单位阶跃函数：$y = \begin{cases} 0 & z<0 \\ 0.5 & z=0 \\ 1 & z>0 \end{cases}$　不连续



② 对数几率函数 (logistic function)

一种 sigmoid 函数

$$y = \frac{1}{1 + e^{-(w^T x + b)}}$$

$$\Rightarrow \ln \frac{y}{1-y} = w^T x + b$$

y → 正例，$1-y$ → 反例　了几率：作为正例的相对可能性

对数几率，用线性回归逼近真实标记的对数几率，直接对分类可能性建模

$$p(y=1|x) = \frac{e^{w^T x + b}}{1 + e^{w^T x + b}}$$

$$p(y=0|x) = \frac{1}{1 + e^{w^T x + b}}$$

极大似然估计 求 w, b　对数似然

给定 $\{(x_i, y_i)\}_{i=1}^{m}$　原始：$l(w;b) = \prod_{i=1}^{m} p(y_i|x_i; w, b)$

对数似然：
$$l(w,b) = \sum_{i=1}^{m} \ln p(y_i|x_i; w, b)$$

$\beta = (w; b)$　$\hat{x} = (x; 1) \Rightarrow w^T x + b = \beta^T \hat{x}$

$p_1(\hat{x}; \beta) = p(y=1|\hat{x}; \beta)$

$p_0(\hat{x}; \beta) = p(y=0|\hat{x}; \beta)$

$l(w,b) = \sum_{i=1}^{m} \ln p(y_i|x_i; w, b)$　最大化

$$\begin{aligned}
&= \sum_{i=1}^{m} \ln[y_i p(y_i=1|\hat{x};\beta) + (1-y_i) p(y_i=0|\hat{x};\beta)] \\
&= \sum_{i=1}^{m} \ln\left[\frac{y_i e^{w^T x_i + b}}{1 + e^{w^T x_i + b}} + \frac{(1-y_i)}{1 + e^{w^T x_i + b}}\right] \\
&= \sum_{i=1}^{m} \ln\left(\frac{y_i e^{w^T x_i + b} + (1-y_i)}{1 + e^{w^T x_i + b}}\right) \\
&= \sum_{i=1}^{m} \left[\ln[y_i e^{w^T x_i + b} + (1-y_i)] - \ln(1 + e^{w^T x_i + b})\right] \\
&= \sum_{i=1}^{m} \left(y_i (w^T x_i + b) - \ln(1 + e^{w^T x_i + b})\right)
\end{aligned}$$

高阶连续可导凸函数

统计学习(李航)：

$$l(w;b) = \prod_{i=1}^{m} p(y=1|\hat{x};\beta)^{y_i} p(y=0|\hat{x};\beta)^{(1-y_i)}$$

对数似然：

$$\begin{aligned}
&= \sum_{i=1}^{m} [y_i \ln p(y=1|\hat{x};\beta) + (1-y_i) \ln p(y=0|\hat{x};\beta)] \\
&= \sum_{i=1}^{m} \left[y_i \ln \frac{e^{w\hat{x}}}{1 + e^{w\hat{x}}} + (1-y_i) \ln \frac{1}{1 + e^{w\hat{x}}}\right] \\
&= \sum_{i=1}^{m} [y_i w \hat{x} - \ln(1 + e^{w\hat{x}})]
\end{aligned}$$

3.4 线性判别分析

找一条直线，使训练集合中的两类点投影到该直线上时，同类尽可能近，异类尽可能远

X_i：$i\in\{0,1\}$ 类的集合

μ_i：$i\in\{0,1\}$ 类的均值向量

Σ_i：　　〃 协方差矩阵

投影到直线上 $\begin{cases} w^T\mu_0,\ w^T\mu_1 \\ w^T\Sigma_0 w,\ w^T\Sigma_1 w \end{cases}$

同样例投影之协方差尽可能小

$$w^T\Sigma_0 w+w^T\Sigma_1 w \downarrow$$

异类远离

$$\|w^T\mu_0-w^T\mu_1\|_2^2 \uparrow$$

$$\max J=\frac{\|w^T\mu_0-w^T\mu_1\|_2^2}{w^T\Sigma_0 w+w^T\Sigma_1 w} \uparrow$$

$$=\frac{w^T(\mu_0-\mu_1)(\mu_0-\mu_1)^T w}{w^T(\Sigma_0+\Sigma_1)w}$$

类内散度：$S_w=\Sigma_0+\Sigma_1$

$$=\sum_{x\in X_0}(x-\mu_0)(x-\mu_0)^T+\sum_{x\in X_1}(x-\mu_1)(x-\mu_1)^T$$

类间散度　$S_b=(\mu_0-\mu_1)(\mu_0-\mu_1)^T$

$$J=\frac{w^TS_bw}{w^TS_ww}$$

S_b与S_w的广义瑞利商

$$\min J=-w^TS_bw$$
$$s.t.\ w^TS_ww=1$$

拉格朗日乘子

$$J(w,\lambda)=-w^TS_bw+\lambda(w^TS_ww-1)$$

$$\frac{\partial J}{\partial w}=-2S_bw+2\lambda S_ww=0 \Rightarrow S_bw=\lambda S_ww$$

? S_bw的方向恒为 $\mu_0-\mu_1$，不妨令 ?

$$S_bw=\lambda(\mu_0-\mu_1)$$

$$w=S_w^{-1}(\mu_0-\mu_1)$$

3.5 多分类学习

拆 $\begin{cases}\text{一对一 OvO}\\ \text{一对余 OvR}\\ \text{多对多 MvM}\end{cases}$

OvO：两两构成一个分类器 $\frac{N(N-1)}{2}$ 个

预测时投票；存储、预测开销大，训练小

OvR：某一类作为正例，其余为负例，N个

预测时只有一个正则，多个则排序，置信度最大的为最终类别　训练大

MvM：若干类为正、若干类为负

纠错输出码(ECOC)

编码：N类M次划分，M个分类器

每次划分：一部分为正、一部分为负

二分类

解码：M个分类器分别进行预测，标记组成一个编码，与每个类别编码进行比较

返回距离最小的

3.6 类别不平衡问题

$y=w^Tx+b$ 来预测

阈值为0.5 $\Leftrightarrow \begin{cases}\frac{y}{1-y}>1 \Rightarrow 正\\ \frac{y}{1-y}<1 \Rightarrow 负\end{cases}$

分类器默认真实正反例可能性相同

真实几率

观测几率 $\frac{m^+}{m^-}$　假设是无偏采样

真实分布在训练集采样

令 $\frac{y'}{1-y'}=\frac{y}{1-y}\times\frac{m^-}{m^+}$（再缩放）即 $\frac{y}{1-y}>\frac{m^+}{m^-}$

① 欠采样

② 过采样

③ 阈值移动：照常训练，预测用再缩放