线性回归

$$Y = f(X) + \epsilon$$

input

x_1^m
x_2^m
x_3^m

output

y^m
$Y = M\beta \Rightarrow M^TY = M^TM\beta \Rightarrow \beta = (M^TM)^{-1}M^{-1}Y$

$$y = ax_1 + bx_2 + cx_3 + d + \epsilon$$

频率学派--最大似然

> **分号与逗号的区别：**
> ;x -- x是个未知的常数
> ,x -- x是个变量

$$L(\theta) = \prod_{i=1}^{m} P(x_i;\theta)$$

$$\begin{aligned} \theta_{MLE} &= arg\max_\theta L(\theta) \\ &= arg\max_\theta \log L(\theta) \\ &= arg\max_\theta \log P(x_i;\theta) \end{aligned}$$

θ没有统计上的解释

条件概率

$$P(\theta|X) = \frac{P(\theta, X)}{P(X)} = \frac{P(\theta)P(X|\theta)}{P(X)}$$

$$\begin{aligned} \theta_{MAP} &= arg\max_\theta P(\theta|X) \\ &= arg\max_\theta P(\theta)P(X|\theta) \\ &= arg\max_\theta \{\log P(\theta) + \log P(X|\theta)\} \\ &= arg\max_\theta \{\log P(\theta) + \sum_{i=1}^{m} \log P(x_i|\theta)\} \end{aligned}$$

$$y = \theta^T X$$

$$y^{(i)} = \theta^T X^{(i)} + \epsilon^{(i)}$$

$$\epsilon^{(i)} = y^{(i)} - \theta^T X^{(i)} \sim N(0, \sigma^2)$$

$$\sum_{i=1}^{m} \log P(x_i|\theta) = m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{\sum_{i=1}^{m}(y^{(i)} - \theta^T X^{(i)})^2}{2\sigma^2}$$

$$\log P(\theta) = \log \frac{1}{(2\pi)^{\frac{k}{2}}} - \lambda\theta^T\theta$$

$$\begin{aligned}\theta_{MAP} &= arg\max_{\theta}\{-\sum_{i=1}^{m}(y^{(i)} - \theta^T X^{(i)})^2 - \lambda\theta^T\theta\} \\ &= arg\min_{\theta}\{\sum_{i=1}^{m}(y^{(i)} - \theta^T X^{(i)})^2 + \lambda\theta^T\theta\}\end{aligned}$$

$\lambda\theta^T\theta$：正则项，这与θ的分布有关

$$LASSO$$

非线性回归

$$y = ax_1^2 + b\log x_2 + ce^{x_3} + d$$

input

$$\begin{pmatrix} x_1^2 \\ \log x_2 \\ e^{x_3} \\ 1 \end{pmatrix}$$

$$\min_{\beta} \quad \|X\beta - Y\|^2 + \lambda\|\beta\|^2$$

凸函数

$$f(\lambda\beta_1 + (1-\lambda)\beta_2) \geq \lambda f(\beta_1) + (1-\lambda)f(\beta_2)$$

伪梯度

$y = |x_1| + |x_2| + |x_3|$
$g = \operatorname{sgn}(x_1, x_2, x_3)$

求解

$$\min_{\beta} \quad \|X\beta - Y\|_2^2 + \lambda\|\beta\|_2^2$$

- 求导

$$\frac{\partial(X\beta - Y)^T(X\beta - Y) + \lambda\beta^T\beta}{\partial\beta}$$

$$(X\beta - Y)^T X + \lambda\beta^T = 0$$

$$(X^T X + \lambda I)^T\beta = X^T Y$$

矩阵乘法

$$C = A \cdot B \Rightarrow C_{ij} = \sum_k A_{ik}B_{kj}$$

$$D = A \cdot B \cdot C \Rightarrow D_{ij} = \sum_{k_1}\sum_{k_2} A_{ik_1}B_{k_1k_2}C_{k_2j}$$

矩阵求导

$$f(X)=AX+b$$

- 变量是向量，函数是向量

$$\frac{\partial(f_1,\dots,f_m)}{\partial(x_1,\dots,x_n)}$$

$$D_{ij}=\frac{\partial f_i}{\partial x_j}=A_{ij}$$

则

$$\frac{\partial f}{\partial X}=A$$

A是矩阵，x是实数$A=f(x)$
则

$$\frac{\partial A^{-1}}{\partial x}=-A^{-1}\frac{\partial A}{\partial x}A^{-1}$$

- 变量是矩阵，函数是标量

$$\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & \ddots & \vdots\\ \frac{\partial f}{\partial x_{n1}} & \cdots & \frac{\partial f}{\partial x_{nn}}\end{bmatrix}$$

- 迹
  迹
  `tr` 是矩阵对角线元素之和

$$\frac{\partial tr[AXB]}{\partial x}$$

$$[AXB]_{hk}=\sum_i\sum_j A_{hi}X_{ij}B_{jk}$$

A是个矩阵
A的迹 $tr[A]=\sum_{i=1}^n a_{ii}$

$$tr[AXB]=\sum_h\sum_i\sum_j A_{hi}X_{ij}B_{jh}$$

$$\because\quad \frac{\partial tr[AXB]}{\partial x_{ij}}=\sum_h A_{hi}B_{jh}=\sum_h B_{jh}A_{hi}=[BA]_{ji}=[(BA)^T]_{ij}$$

$$\therefore\quad \frac{\partial tr[AXB]}{\partial x}=A^TB^T$$

$$\frac{\partial tr[AX^TB]}{\partial x}=\frac{\partial tr[B^TXA^T]}{\partial x}=BA$$

$$\frac{\partial tr[AXBXC]}{\partial x}=A^TC^TX^TB^T+B^TX^TA^TC^T$$

- 范式

$$\|A\|_F^2 = \sum_i \sum_j a_{ij}^2$$

$$[AA^T]_{ii} = \sum_k A_{ik}^2$$

$$tr[AA^T] = \sum_i \sum_j A_{ij}^2 = \|A\|_F^2$$

$$\begin{aligned} \frac{\partial \|B-XA\|_F^2}{\partial x} &= \frac{\partial tr[(B-XA)^T(B-XA)]}{\partial x} \\ &= \frac{\partial tr[B^TB + A^TX^TXA - A^TX^TB - B^TXA]}{\partial x} \\ &= 2XAA^T - 2BA^T \end{aligned}$$

令$2XAA^T - 2BA^T = 0$得:

$$X = BA^T(AA^T)^{-1}$$

logistic regression⇒分类

$$q(X=x) = P(Y=1|X=x)$$

$$\frac{q}{1-q} \in (0,+\infty) \quad \log\frac{q}{1-q} \in (-\infty,+\infty)$$

$$\log\frac{q}{1-q} = ax_1 + bx_2 + cx_3 + d = \beta x$$

$$q = \frac{e^{\beta x}}{1+e^{\beta x}}$$

$$P(Y_1 = y_1, \ldots, Y_n = y_n | X_1 = x_1, \ldots, X_n = x_n)$$

$$\begin{aligned} &= \prod_{i:y_i=1} q(x_i,\beta) \cdot \prod_{i:y_i=0} (1-q(x_i,\beta)) \\ &= \sum_{i=1}^n y_i \log q(x_i;\beta) + (1-y_i)\log(1-q(x_i;\beta)) \\ &= \sum \{y_i \beta^T x_i - \log(1+e^{\beta^T x_i})\} \end{aligned}$$

GDA（高斯判别分析）⇒二分类

> **基本概念**
>
> - 概率密度
>
> $$(X,Y) \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$
>
> - 边缘密度函数

> $$f_X(x)\int_{-\infty}^{\infty}f(x,y)dy$$
>
> $$f_Y(y)\int_{-\infty}^{\infty}f(x,y)dx$$
>
> - 独立
>   if $f(x,y)=f_X(x)\cdot f_Y(y)\quad\forall x,y$
>   then X和Y相互独立
> - 条件概率
>
> $$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}$$
>
> $$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$
>
> - 贝叶斯定理
>
> $$\begin{aligned}P(A|B)&=\frac{P(A,B)}{P(B)}=\frac{P(B|A)P(A)}{P(B)}\\&=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\bar{A})P(\bar{A})}\end{aligned}$$

$$P(y|x)=\frac{P(x|y)P(y)}{P(x)}$$

$$arg\max_y P(y|x)=arg\max_y P(x|y)P(y)$$

$$\begin{cases}P_0=P(Y=0|X=x)\sim P(X=x|Y=0)P(Y=0)\\P_1=P(Y=1|X=x)\sim P(X=x|Y=1)P(Y=1)\end{cases}$$

$$(x^{(i)},y^{(i)})$$

$$\prod_{i=1}^{m}P(x^{(i)},y^{(i)})=\prod_{i=1}^{m}P(x^{(i)}|y^{(i)})P(y^{(i)})$$

目标函数⇒

$$\sum_{i=1}^{m}\log P(x^{(i)}|y^{(i)})+\sum_{i=1}^{m}\log P(y^{(i)})$$

$$P(X=x|Y=0)\sim N(\mu_0,\Sigma)$$

$$P(X=x|Y=1)\sim N(\mu_1,\Sigma)$$

$$P(y^{(i)})=\Phi^{y^{(i)}}(1-\Phi)^{y^{(i)}}$$

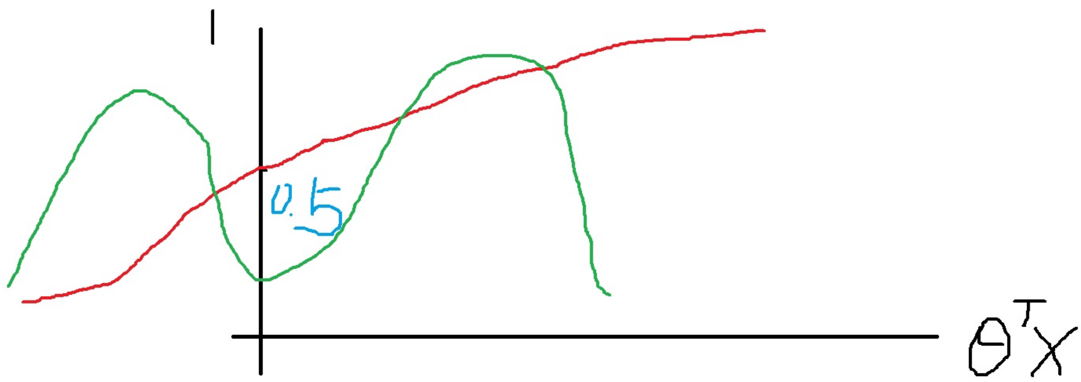
$$\Phi=\frac{1}{m}\sum_{i=1}^{m}\mathbb{S}\{y^{(i)}=1\}$$

$$\mu_0 = \frac{\sum_{i=1}^{m} \mathbb{S}\{y^{(i)} = 0\} x^{(i)}}{\sum_{i=1}^{m} \mathbb{S}\{y^{(i)} = 0\}}$$

$$\mu_1 = \frac{\sum_{i=1}^{m} \mathbb{S}\{y^{(i)} = 1\} x^{(i)}}{\sum_{i=1}^{m} \mathbb{S}\{y^{(i)} = 1\}}$$

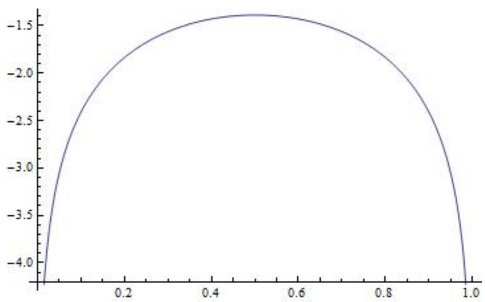
$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$$

$$\mathbb{S}(\text{表达式}) = \begin{cases} 1 & \text{表达式为真} \\ 0 & \text{表达式为假} \end{cases}$$



Naive Bayes

$$f(x) = \log x + \log(1 - x) \quad x \in (0, 1)$$



$$f(x) = a \log x + b \log(1 - x) \quad a > 0, b > 0$$

令

$$f'(x) = \frac{a}{x} - \frac{b}{1 - x} = 0 \Rightarrow x = \frac{a}{a + b}$$

$$f(x_1,x_2)=a\log x_1+b\log x_2 \quad a>0,b>0,x_1,x_2>0,x_1+x_2=1$$

则

$$x_1=\frac{a}{a+b},x_2=\frac{b}{a+b}$$

$$f(x_1,\cdots,x_n)=\sum_{i=1}^{n}a_i\log x_i \quad x_i>0,\sum x_i=1$$

则

$$x_i=\frac{a_i}{\sum_j a_j}$$

Input:

$$(x^{(i)},y^{(i)})_{i=1,2,\cdots,m}$$

$$x^{(i)}=(x_1^{(i)},\cdots,x_n^{(n)}) \quad x_j^{(i)}=\{0,1\}$$

$$y^{(i)}\in\{1,\cdots,k\}$$

$$\begin{aligned}&\log\prod_{i=1}^{m}P(x^{(i)},y^{(i)})\\&=\sum_{i=1}^{m}\log P(x^{(i)},y^{(i)})\\&=\sum_{i=1}^{m}\log P(y^{(i)})+\sum_{i=1}^{m}\log P(x^{(i)}|y^{(i)})\\&=\sum_{i=1}^{m}\log P(y^{(i)})+\sum_{i=1}^{m}\sum_{j=1}^{n}\log P(x_j^{(i)}|y^{(i)})\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{m}\log P(y^{(i)})&=\sum_{j=1}^{k}\sum_{i:y^{(i)}=j}^{m}\log P(y^{(i)}=j)\\&=\sum_{j=1}^{k}\alpha_j\log P_j\\ \sum P_j=1 \quad P_j\sim\alpha_j \quad P_j&=\frac{\alpha_j}{\sum\alpha_i}\end{aligned}$$

令α_j代表那些满足$y^{(i)}=j$的样本个数
拉普拉斯平滑

$$P_j=\frac{\alpha_i+1}{m+k}>0$$

$$P(x_j|y)=\frac{\text{那些个}y^{(i)}=y\text{且}x_j^{(i)}=x\text{的样本个数}}{\text{那些个}y^{(i)}=y\text{的样本个数}}$$

拉普拉斯平滑

$$P(x_j|y) = \frac{1 + 那些个y^{(i)} = y且x_j^{(i)} = x的样本个数}{2 + 那些个y^{(i)} = y的样本个数}$$

$$\begin{aligned} & P(Y=1|(x_1, x_2, \cdots, x_n)) \\ = & \frac{P(Y=1) \cdot P(x_1, \cdots, x_n|Y=1)}{P(x_1, \cdots, x_n)} \end{aligned}$$

EM

在凸区域上，满足$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$则为凸函数

$$f(y) \ge f(x) + Df(x)^T(y-x)$$

$$f(\sum_{i=1}^{n} \lambda_i x_i) \le \sum \lambda_i f(x_i) \quad \sum \lambda_i = 1$$

设$p(x)$是概率密度函数，则

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

$$E(x) = \int_{-\infty}^{\infty} x \cdot p(x)dx$$

令$y = f(x)$是凸函数

$$f(\int x \cdot p(x)dx) \le \int_{-\infty}^{\infty} f(x)p(x)dx$$

$$f(E(X)) \le E(f(X))$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

则

$$\int_0^{+\infty} f(x) \log x dx \le \log \int_0^{+\infty} f(x) x dx$$

GMM-高斯混合模型

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{m} \log(P(x_i;\theta)) \\ &= \sum_{i=1}^{m} \log \sum_{z_i \in \Omega} P(x_i, z_i; \theta) \quad 边缘概率 \\ &= \sum_{i=1}^{m} \log \sum_{z_i} Q(z^i) \frac{P(x_i, z_i; \theta)}{Q(z^i)} \end{aligned}$$

令$\Phi(z^i) = \frac{P(x_i, z_i; \theta)}{Q(z^i)}$,$\sum_{z^i \in \Omega} Q(z^i) = 1$，则:

$$\sum_{i=1}^{m} \log \sum_{z_i} Q(z^i) \frac{P(x_i, z_i; \theta)}{Q(z^i)} \ge \sum_{i=1}^{m} \sum_{z_i \in \Omega} Q(z^i) \log \frac{P(x_i, z_i; \theta)}{Q(z^i)}$$

取等条件:

$$\frac{P(x_i,z_i;\theta)}{Q(z^i)}=C \quad \text{C是常数}$$

即

$$P(x_i,z_i;\theta)\sim Q(z^i) \quad \sum_{z_i}Q(z_i)=1$$

$$Q(z_i)=\frac{P(x_i,z_i;\theta)}{\sum_{z_i\in\Omega}P(x_i,z_i;\theta)}==\frac{P(x_i,z_i;\theta)}{P(x_i;\theta)}=P(z_i|x_i;\theta)$$

E-step: $\theta(z_i)$如何选取
M-step: 寻找θ，优化目标函数

$$z(x)=1\Rightarrow\mu_1,\Sigma_1$$
$$z(x)=2\Rightarrow\mu_2,\Sigma_2$$

$$\begin{aligned}Q_i(z^{(i)}=j)=P(z^{(i)}=j|x^{(i)};\mu,\Sigma)\\=\frac{P(z^{(i)}=j;x^{(i)};\mu,\Sigma)}{P(x^{(i)};\mu,\Sigma)}\\=\frac{P(z^{(i)}=j)\cdot P(x^{(i)}|z^{(i)}=j)}{\sum_{z^{(i)}}P(x^{(i)},z^{(i)};\mu,\Sigma)}\end{aligned}$$