Linear Regression

$w_1x_1^2 + w_2x_2^2 \cdots$

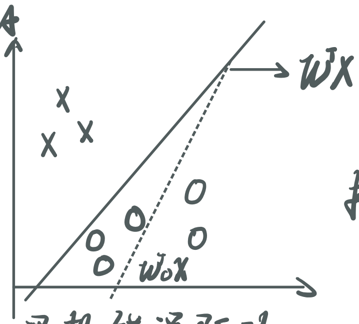
sign, relu

LR
- ① 线性 $\xrightarrow{\times}$
  - 属性非线性：特征转换（多项式回归）
  - 全局非线性：线性分类（activation f 是非线性）
  - 系数非线性：NN. PLA
- ② 全局性 $\xrightarrow{\times}$ 线性样条回归，决策树
- ③ 数据未加工 $\xrightarrow{\times}$ PCA

线性回归 $\xRightarrow[\text{降维}]{\text{activation}}$ 线性分类 (0, 1) $\Rightarrow$
- $y = f(W^TX+b)$, $y \in \{0,1\}$ / $[0,1]$
  - activation function
- f^{-1}: link function
- $f: W^TX+b \longmapsto \{0,1\}$
- $f^{-1}: \{0,1\} \longmapsto W^TX+b$

线性分类
- 硬分类：$y \in \{0,1\}$
  - 线性判别分析 (fisher)
  - 感知机
- 软分类：$y \in [0,1]$
  - 生成式
    - Naive Bayes Classifier
    - Gaussian Discriminant Analysis
  - 判别式：Logistic Regression

PLA



D: {被错误分类的样本}

样本集: $\{(x_i, y_i)\}_{i=1}^{N}$

思想: 错误驱动

模型: $f(x) = sign(w^T x)$, $x \in \mathbb{R}^p$, $w \in \mathbb{R}^p$

$$sign(a) = \begin{cases} +1, & a \geq 0 \\ -1, & a < 0 \end{cases}$$

策略: ① loss function: $\{0, 1\}$

$$L(w) = \sum_{i=1}^{N} I\{y_i w^T x_i < 0\}$$

无法求解

② $L(w) = \sum_{x_i \in D} -y_i w^T x_i$

$$\nabla_w L = -y_i x_i$$

Algo: SGD:

$$w^{(t+1)} \leftarrow w^{(t)} - \lambda \nabla_w L$$

$$w^{(t)} + \lambda y_i x_i$$

Fisher

$$X=(x_1, x_2 \cdots x_N)^T=\begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix}_{N\times p}$$

$$Y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_{N\times 1}$$

$$\{(x_i, y_i)\}_{i=1}^N,\ x_i\in\mathbb{R}^p,\ y_i\in\{\overset{C_1}{+1}, \overset{C_2}{-1}\}$$

$$x_{c1}=\{x_i \mid y_i=+1\},\ x_{c2}=\{x_i \mid y_i=-1\}$$

$$|x_{c1}|=N_1,\ |x_{c2}|=N_2,\ N_1+N_2=N$$

思想：类内小，类间大

目标函数：$J(w)=\dfrac{(\bar{z}_1-\bar{z}_2)^2}{S_1+S_2}$

$$z_i=w^Tx_i$$

$$\bar{z}=\frac{1}{N}\sum_{i=1}^N z_i=\frac{1}{N}\sum_{i=1}^N w^Tx_i$$

$$S_z=\frac{1}{N}\sum_{i=1}^N (z_i-\bar{z})(z_i-\bar{z})^T$$

$$=\frac{1}{N}\sum_{i=1}^N (w^Tx_i-\bar{z})(w^Tx_i-\bar{z})^T$$

$$C_1:\ \bar{z}_1=\frac{1}{N_1}\sum_{i=1}^{N_1} w^Tx_i$$

$$S_1=\frac{1}{N_1}\sum_{i=1}^{N_1}(w^Tx_i-\bar{z}_1)(w^Tx_i-\bar{z}_1)^T$$

$$C_2:\ \bar{z}_2=\frac{1}{N_2}\sum_{i=1}^{N_2} w^Tx_i$$

$$S_2=\frac{1}{N_2}\sum_{i=1}^{N_2}(w^Tx_i-\bar{z}_2)(w^Tx_i-\bar{z}_2)^T$$

类间：$(\bar{z}_1-\bar{z}_2)^2$

类内：S_1+S_2

$$\hat{w}=\arg\max J(w)$$

$$分子=\left(\frac{1}{N_1}\sum_{i=1}^{N_1} w^Tx_i-\frac{1}{N_2}\sum_{i=1}^{N_2} w^Tx_i\right)^2$$

$$=\left[w^T\left(\frac{1}{N_1}\sum_{i=1}^{N_1} x_i-\frac{1}{N_2}\sum_{i=1}^{N_2} x_i\right)\right]^2$$

$$=w^T(\bar{x}_{c1}-\bar{x}_{c2})(\bar{x}_{c1}-\bar{x}_{c2})^Tw$$

分母：$=S_1+S_2$

$$S_1=\frac{1}{N_1}\sum_{i=1}^{N_1}\left(w^Tx_i-\frac{1}{N_1}\sum_{j=1}^{N_1}w^Tx_j\right)\left(w^Tx_i-\frac{1}{N_1}\sum_{j=1}^{N_1}w^Tx_j\right)^T$$

$$=\frac{1}{N_1}\sum_{i=1}^{N_1}w^T(x_i-\bar{x}_{c1})(x_i-\bar{x}_{c1})^Tw$$

$$=w^T\underbrace{\left[\frac{1}{N_1}\sum_{i=1}^{N_1}(x_i-\bar{x}_{c1})(x_i-\bar{x}_{c1})^T\right]}_{S_{c1}}w$$

$$=w^T\cdot S_{c1}\cdot w$$

$$分母=w^TS_{c1}w+w^TS_{c2}w$$

$$=w^T(S_{c1}+S_{c2})w$$

$$\therefore J(w) = \frac{w^T(\bar{x}_{c1}-\bar{x}_{c2})(\bar{x}_{c1}-\bar{x}_{c2})^T w}{w^T(S_{c1}+S_{c2})w}$$

S_b: between-class 类间方差 $=(\bar{x}_{c1}-\bar{x}_{c2})(\bar{x}_{c1}-\bar{x}_{c2})^T$

S_w: within-class 类内方差 $=(S_{c1}+S_{c2})$

$$= \frac{w^T S_b w}{w^T S_w w} =$$

$$= (w^T S_b w)(w^T S_w w)^{-1}$$

$$\frac{\partial J(w)}{\partial w} = 2 S_b w \cdot (w^T S_w w)^{-1} + (w^T S_b w)\cdot(-1)(w^T S_w w)^{-2}\cdot(2 S_w w)$$

$$= 0$$

$$S_b w (w^T S_w w) - w^T S_b w S_w w = 0$$

$w: p\times 1 \quad S_w = p\times p$

$w^T: 1\times p \quad S_b = p\times p$

$$\underbrace{\underset{1\times p}{w^T}\ \underset{p\times p}{S_b}\ \underset{p\times 1}{w}}_{1\times 1\ \mathbb{R}} S_w w = S_b w (\underbrace{\underset{1\times p}{w^T}\ \underset{p\times p}{S_w}\ \underset{p\times 1}{w}}_{1\times 1}) \ \mathbb{R}$$

$$S_w w = \frac{\overset{>0}{w^T S_w w}}{w^T S_b w}\cdot S_b w$$

$$w = \frac{\overset{>0}{w^T S_w w}}{w^T S_b w}\cdot S_w^{-1}\cdot S_b\cdot w$$

$$\propto S_w^{-1}\cdot \underbrace{S_b\cdot w}_{p\times 1}$$

$$\boxed{(\bar{x}_{c1}-\bar{x}_{c2})}\underbrace{\underset{1\times p}{(\bar{x}_{c1}-\bar{x}_{c2})^T}\ \underset{p\times 1}{w}}_{1\times 1}$$

↖不影响方向

$$\propto \boxed{S_w^{-1}\cdot(\bar{x}_{c1}-\bar{x}_{c2})}$$

w { ~~大小~~ / 方向

→ $w^T x$ scaling

Logistic Regression

$$P_1 = P(y=1|x) = \sigma(W^T x) = \frac{1}{1+e^{-W^T x}} = \varphi(x_i, w), \quad y=1$$

$$P_0 = 1 - P(y=1|x) = \frac{e^{-W^T x}}{1+e^{-W^T x}}, \quad y=0$$

$$= 1 - \varphi(x_i, w)$$

$$P(y|x) = P_1^{y} P_0^{(1-y)}$$

$$MLE: \hat{w} = \underset{w}{\operatorname{argmax}} \log(\underbrace{P(y|x)}_{\text{likelihood}})$$

$$= \operatorname{argmax} \log\left(\prod_{i=1}^{n} P(y|x)\right)$$

$$= \operatorname{argmax} \sum_{i=1}^{n} \log(P(y|x))$$

$$= \operatorname{argmax} \sum_{i=1}^{n} y_i \log(P_1) + (1-y_i)\log(P_0)$$

$$= \operatorname{argmax} \underbrace{\sum_{i=1}^{n}\left[y_i \log(\varphi(x_i, w)) + (1-y_i)\log(1-\varphi(x_i, w))\right]}_{-\text{Cross Entropy}}$$

$$\underset{(max)}{MLE} \Rightarrow \underset{(min)}{\text{loss}} \text{ function (min Cross Entropy)}$$

Gaussian Discriminant Analysis

Data $\{x_i, y_i\}_{i=1}^{N}$

$x_i \in \mathbb{R}^p,\ y_i \in \{0,1\}$

$\hat{y} = \underset{y\in\{0,1\}}{\arg\max}\ P(y|x) = \underset{y}{\arg\max}\ P(y)P(x|y)$

$P(y|x) = \dfrac{P(x|y)\cdot P(y)}{P(x)}$

$\underset{\text{posterior}}{P(y|x)} \propto \underset{\text{likelihood}}{P(x|y)}\ \overset{\text{prior}}{P(y)}$

$y \sim \text{Bernoulli}(\phi)$

y	1	0
p	ϕ	$1-\phi$

$\begin{cases} \phi^{y} & y=1 \\ (1-\phi)^{1-y}, & y=0 \end{cases} \to \phi^{y}\cdot(1-\phi)^{1-y}$

$$\left.\begin{aligned} x|y=1 &\sim N(\mu_1,\Sigma) \\ x|y=0 &\sim N(\mu_2,\Sigma) \end{aligned}\right\} \Rightarrow N(\mu_1,\Sigma)^{y}\cdot N(\mu_2,\Sigma)^{1-y}$$

log-likelihood : $l(\theta) = \log \prod_{i=1}^{N} P(x_i, y_i) = \sum_{i=1}^{N} \log(P(x_i|y_i)P(y_i))$

$\theta = (\mu_1, \mu_2, \Sigma, \phi)$ $\quad = \sum_{i=1}^{N}[\log P(x_i|y_i) + \log P(y_i)]$

$\hat{\theta} = \underset{\theta}{\arg\max}\ l(\theta)$ $\quad = \sum_{i=1}^{N}[\log N(\mu_1,\Sigma)^{y_i}\cdot N(\mu_2,\Sigma)^{1-y_i} + \log \phi^{y_i}\cdot(1-\phi)^{1-y_i}]$

$$= \sum_{i=1}^{N}[\underbrace{\log N(\mu_1,\Sigma)^{y_i}}_{①} + \underbrace{\log N(\mu_2,\Sigma)^{1-y_i}}_{②} + \underbrace{\log \phi^{y_i}(1-\phi)^{1-y_i}}_{③}]$$

求ϕ: $③ = \sum_{i=1}^{N} \log \phi^{y_i}(1-\phi)^{1-y_i}$

$$\frac{\partial ③}{\partial \phi} = \sum_{i=1}^{N} y_i\frac{1}{\phi} + (1-y_i)\frac{1}{1-\phi}(-1)$$

$$= \sum_{i=1}^{N} y_i\frac{1}{\phi} - (1-y_i)\frac{1}{1-\phi} = 0$$

$\begin{cases} y=1: N_1 \\ y=0: N_2 \\ N = N_1 + N_2 \end{cases}$

$$\Rightarrow \sum_{i=1}^{N} y_i(1-\phi) - (1-y_i)\phi = 0$$

$$\sum_{i=1}^{N} y_i - \cancel{y_i\phi} - \phi + \cancel{y_i\phi} = 0$$

$$\sum_{i=1}^{N} y_i - N\phi = 0$$

$$\boxed{\therefore\ \hat{\phi} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{N_1}{N}}$$

求 μ_1

$$N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$① = \sum_{i=1}^{N} \log N(\mu_1, \Sigma)^{y_i}$$

$$= \sum_{i=1}^{N} y_i \log \underbrace{\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}}_{\text{Constant}} \exp\left(-\frac{1}{2}(x_i-\mu_1)^T \Sigma^{-1}(x_i-\mu_1)\right)$$

$$\mu_1 = \underset{\mu_1}{\operatorname{argmax}}\, ① = \operatorname{argmax} \sum_{i=1}^{N} y_i \left(-\frac{1}{2}(x_i-\mu_1)^T \Sigma^{-1}(x_i-\mu_1)\right)$$

$$\Delta = \sum_{i=1}^{N} y_i \left(-\frac{1}{2}(x_i-\mu_1)^T \Sigma^{-1}(x_i-\mu_1)\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{N} y_i (x_i^T\Sigma^{-1} - \mu_1^T\Sigma^{-1})(x_i-\mu_1) \quad 1\times1$$

$$= -\frac{1}{2}\sum_{i=1}^{N} y_i \{\underbrace{x_i^T\Sigma^{-1}x_i}_{\text{Constant}} - \underbrace{x_i^T\Sigma^{-1}\mu_1}_{1\times p\ \ p\times p\ \ p\times 1} - \mu_1^T\Sigma^{-1}x_i + \mu_1^T\Sigma^{-1}\mu_1\}$$

$$= -\frac{1}{2}\sum_{i=1}^{N} y_i \{\cancel{x_i^T\Sigma^{-1}x_i} - 2\mu_1^T\Sigma^{-1}x_i + \mu_1^T\Sigma^{-1}\mu_1\}$$

$$\frac{\partial\Delta}{\partial\mu_1} = -\frac{1}{2}\sum_{i=1}^{N} y_i \{-2\Sigma^{-1}x_i + 2\Sigma^{-1}\mu_1\} = 0$$

$$\sum_{i=1}^{N} y_i \{\Sigma^{-1}\mu_1 - \Sigma^{-1}x_i\} = 0$$

$$\sum_{i=1}^{N} y_i(\mu_1 - x_i) = 0$$

$$\sum_{i=1}^{N} y_i\mu_1 = \sum_{i=1}^{N} y_1 x_i$$

$$\boxed{\hat{\mu}_1 = \frac{\sum_{i=1}^{N} y_1 x_i}{\underbrace{\sum_{i=1}^{N} y_i}_{N_1}} = \frac{\sum_{i=1}^{N} y_i x_i}{N_1}}$$

$$= \sum_{i=1}^{N}\{\underbrace{\log N(\mu_1,\Sigma)^{y_i}}_{①} + \underbrace{\log N(\mu_2,\Sigma)^{1-y_i}}_{②} + \underbrace{\log \phi^{y_i}(1-\phi)^{1-y_i}}_{③}\}$$

求 $\hat{\Sigma} = \underset{\Sigma}{\operatorname{argmax}}\, ① + ②$

$$① + ② = \sum_{x_i \in C_1} \log N(\mu_1, \Sigma) + \sum_{x_i\in C_2} \log N(\mu_2, \Sigma)$$

$$\sum_{i=1}^{N} \log N(\mu, \Sigma) \quad \sum_{i=1}^{N} \log \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

$$\sum_{i=1}^{N} \log \frac{1}{(2\pi)^{\frac{p}{2}}} + \log|\Sigma|^{-\frac{1}{2}} + \left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

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

$$tr(AB) = tr(BA)$$

$$tr(ABC) = tr(CAB) = tr(BCA)$$

$$= \sum_{i=1}^{N} \overbrace{\left(-\frac{1}{2}\log|\Sigma| + \left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \right] \right)}$$

$$= C - \frac{1}{2}\underbrace{\cancel{\sum_{i=1}^{N}}}_{N\cdot}\log|\Sigma| - \frac{1}{2}\underbrace{\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$$

$$(1\times p \quad p\times p \quad p\times 1) = 1\times 1$$

$$= tr((x_i-\mu)^T \Sigma^{-1}(x_i-\mu))$$

$$= tr((x_i-\mu)\cdot(x_i-\mu)^T \Sigma^{-1}))$$

$$= tr(\underbrace{\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T}_{NS} \Sigma^{-1})$$

$$= N tr(S\Sigma^{-1})$$

$$= C - \frac{1}{2}N\cdot\log|\Sigma|$$

$$= -\frac{1}{2}N\cdot\log|\Sigma| - \frac{1}{2}N tr(S\Sigma^{-1}) + C$$

$$① + ② = -\frac{1}{2}N_1\cdot\log|\Sigma| - \frac{1}{2}N_1 tr(S_1\Sigma^{-1}) - \frac{1}{2}N_2\log|\Sigma| - \frac{1}{2}N_2 tr(S_2\Sigma^{-1}) + C$$

$$= -\frac{1}{2}(N_1+N_2)\log|\Sigma| - \frac{1}{2}N_1 tr(S_1\Sigma^{-1}) - \frac{1}{2}N_2 tr(S_2\Sigma^{-1}) + C$$

$$= -\frac{1}{2}(N\log|\Sigma| + N_1 tr(S_1\Sigma^{-1}) + N_2 tr(S_2\Sigma^{-1})) + C$$

$$\frac{\partial(①+②)}{\partial\Sigma} = -\frac{1}{2}\left(N\cdot\frac{1}{|\Sigma|}\cdot|\Sigma|\,\Sigma^{-1} + S_1^T\cdot(-1)\Sigma^{-2} + S_2^T\cdot(-1)\Sigma^{-2}\right)$$

$$= \cancel{-\frac{1}{2}}(N\Sigma^{-1} - N_1S_1\Sigma^{-2} - N_2S_2\Sigma^{-2}) = 0$$

$$N\Sigma - N_1S_1 - N_2S_2 = 0$$

$$\boxed{\hat{\Sigma} = \frac{1}{N}(N_1S_1 + N_2S_2)}$$

Naive Bayes Classifier

思想：朴素贝叶斯假设

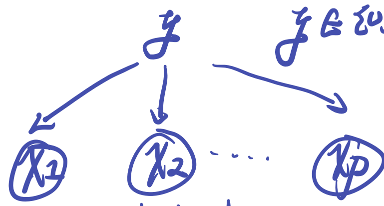
↳ 条件独立性假设

↳ 最简单的概率图模型

有向图

动机：简化运算

$y \in \{0,1\}$



$x_i \perp x_j \mid y \quad (i \neq j)$

$$P(x|y)=\prod_{j=1}^{p} P(x_j|y)$$

Data: $\{(x_i, y_i)\}_{i=1}^{N}$

$x_i \in \mathbb{R}^p$, $y_i \in \{0,1\}$

给定 x，y? 1/0

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

$$\hat{y} = \arg\max_y P(y|x) = \arg\max_y \frac{P(x,y)}{P(x)}$$

$$= \arg\max_y P(y)\cdot P(x|y) \rightarrow \text{MLE}$$

二分类 0/1：$y \sim$ Bernoulli → Bin

多分类：$y \sim$ Categorical → Multinomial

$$P(x|y)=\prod_{j=1}^{p} P(x_j|y)$$

x 离散：$x_j \sim$ Categorical

x 连续：$x_j \sim N(\mu_j, \sigma_j^2)$