线性分类

频率派 ⟶ 统计机器学习.

贝叶斯派 ⟶ 概率图模型.

线性回归

$f(w,b) = w^T x + b$

$x \in \mathbb{R}^p$

- ① 线性
  - 属性: x 是 p 维的, f 关于 x 是线性的.
  - 全局: $w^T x + b$ 是线性组合, 直接输出.
  - 系数: f 关于 w 是线性的.
- ② 全局性: 在整个特征空间拟合一条直线, 非分段
- ③ 数据未加工: 直接使用数据.

线性回归是统计机器学习的基本模型, 其它模型通过打破其一个或多个特点形成了统计机器学习的架构.

若打破属性线性, 如多项式回归, 其属性可能是多次的, 如 x^2

全局线性: 如线性分类, 将 f 作为激活函数的输入, 感知机.

系数线性: 神经网络, 感知机

若打破全局性: 对样本空间分割, 如线性样条回归, 决策树.

若打破数据未加工: PCA, 流形.

线性回归 $\xrightarrow[\text{降维}]{\text{激活函数}}$ 线性分类.

- 线性分类
  - 硬分类 ($y \in \{0,1\}$)
    - LDA
    - 感知机.
  - 软分类 ($y \in [0,1]$)
    - 生成式: 高斯判别分析.
    - 判别式: LR

线性判别分析 (LDA)

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

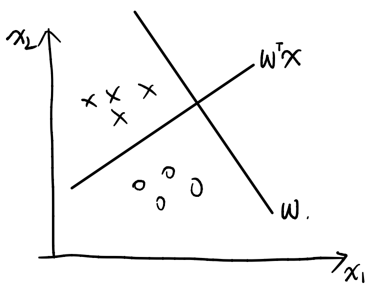
$Y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_N\end{pmatrix}_{N\times 1}$

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

$x_{c_1}=\{x_i\mid y_i=+1\}$, $x_{c_2}=\{x_i\mid y_i=-1\}$

$|x_{c_1}|=N_1$ $|x_{c_2}|=N_2$ $N_1+N_2=N$.

LDA 的思想：类内小，类间大.



限定 $\|w\|=1$.

样本点在 w 上的投影：$z_i=w^Tx_i$

样本投影的均值：$\bar{z}=\frac{1}{N}\sum_{i=1}^N z_i=\frac{1}{N}\sum_{i=1}^N w^Tx_i$

z 的协方差矩阵：$S_z=\frac{1}{N}\sum_{i=1}^N(z_i-\bar{z})(z_i-\bar{z})^T$

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

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

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

类间距离：$(z_1-z_2)^2$

类内距离：S_1+S_2

目标函数：$J(w)=\frac{(z_1-z_2)^2}{S_1+S_2}$，则 $\hat{w}=\arg\max\limits_{w}J(w)$

$$J(w)=\frac{(\bar{z}_1-\bar{z}_2)^2}{S_1+S_2}$$

分子 $=\left(\frac{1}{N_1}\sum_{i=1}^{N_1}w^Tx_i-\frac{1}{N_2}\sum_{i=1}^{N_2}w^Tx_i\right)^2=\left(w^T\left(\frac{1}{N_1}\sum_{i=1}^{N}x_i-\frac{1}{N_2}\sum_{i=1}^{N_2}x_i\right)\right)^2=\left(w^T(\bar{x}_{c_1}-\bar{x}_{c_2})\right)^2$

分母 $= S_1 + S_2$

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

$$= \frac{1}{N_1}\sum_{i=1}^{N_1} w^T(x_i-\bar{x}_{c_1})(x_i-\bar{x}_{c_1})^T w$$

$$= w^T \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N_1}(x_i-\bar x_{c_1})(x_i-\bar x_{c_1})^T\right]}_{S_{c_1}} w$$

$$= w^T S_{c_1} w$$

故 分母 $= w^T S_{c_1} w + w^T S_{c_2} w$

$$= w^T(S_{c_1}+S_{c_2}) w$$

故 $J(w) = \dfrac{w^T(\bar x_{c_1}-\bar x_{c_2})(\bar x_{c_1}-\bar x_{c_2})^T w}{w^T(S_{c_1}-S_{c_2})w}$

模型求解：$J(w) = \dfrac{w^T S_b w}{w^T S_w w}$，其中 S_b：类间方差，S_w：类内方差.

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

令 $\dfrac{\partial J(w)}{\partial w} = 2S_b w (w^T S_w w)^{-1} + w^T S_b w \cdot (-1)\cdot (w^T S_w w)^{-2}\cdot 2 S_w \cdot w = 0$.

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

$$\underbrace{w^T S_b w}_{1\times1 \in R} S_w w = S_b w \underbrace{(w^T S_w w)}_{1\times p\cdot p\times p\cdot p\times 1 \Rightarrow 1\times 1 \in R}$$

则 $S_w \cdot w = \dfrac{w^T S_w w}{w^T S_b w}\cdot S_b w$

$$w = \frac{w^TS_ww}{w^TS_bw}\cdot S_w^{-1}\cdot S_b\cdot w \propto S_w^{-1}\cdot \underbrace{S_b\cdot w}_{(\bar x_{c_1}-\bar x_{c_2})\,\overset{1\times p\,\cdot\, p\times 1}{\underbrace{(\bar x_{c_1}-\bar x_{c_2})^T w}_{1\times 1}}}$$

$$\propto S_w^{-1}\cdot(\bar x_{c_1}-\bar x_{c_2})$$

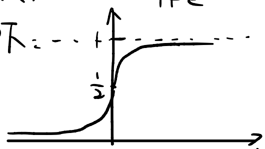
当类别有K个时 最多降维到 K-1维.

Logistic Regression

判别模型直接对 $P(Y|X)$ 进行建模，应用MLE.

sigmoid 函数：$\sigma(z)=\frac{1}{1+e^{-z}}$

其图象如下：



$z\to+\infty,\ \sigma(z)\to 1$

$z=0,\ \sigma(z)=\frac{1}{2}$

$z\to-\infty,\ \sigma(z)\to 0$

$$\begin{cases} p_1=P(y=1|x)=\sigma(w^Tx)=\dfrac{1}{1+e^{-w^Tx}}, & y=1 \\ p_0=P(y=0|x)=1-\dfrac{1}{1+e^{-w^Tx}}=\dfrac{e^{-w^Tx}}{1+e^{-w^Tx}}, & y=0. \end{cases}$$

$$\Rightarrow P(y|x)=p_1^{y}p_0^{1-y}$$

$$\begin{aligned} \text{MLE}:\ \hat{w} &= \arg\max_w \log P(Y|X) \\ &= \arg\max_w \log \prod_{i=1}^{N} P(y_i|x_i) \\ &= \arg\max_w \sum_{i=1}^{N} \log P(y_i|x_i) \\ &= \arg\max_w \underbrace{\sum_{i=1}^{N}\left(y_i\log p_1+(1-y_i)\log p_0\right)}_{-\text{cross entropy}} \end{aligned}$$

$$\underset{(\max)}{\text{MLE}} \Longrightarrow \underset{(\min)}{\text{loss function}}\ (\min\ \text{cross entropy})$$