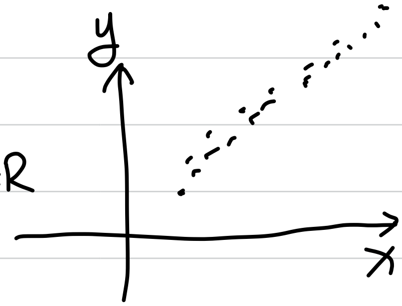
第七课. 线性问题

1. 回归

1.1 $S = \{(x_i, y_i)\}_{i=1}^{n}$ $x_i \in R$ $y_i \in R$



$f_\theta(x) = ax + b$

$L_S = \frac{1}{2}\sum_{i=1}^{n}(y_i - f_\theta(x_i))^2$

GD.   SGD.

$a^{t+1} = a^t - \eta \frac{\partial L_S}{\partial a}$ $b^{t+1} = b^t - \eta \frac{\partial L_S}{\partial b}$

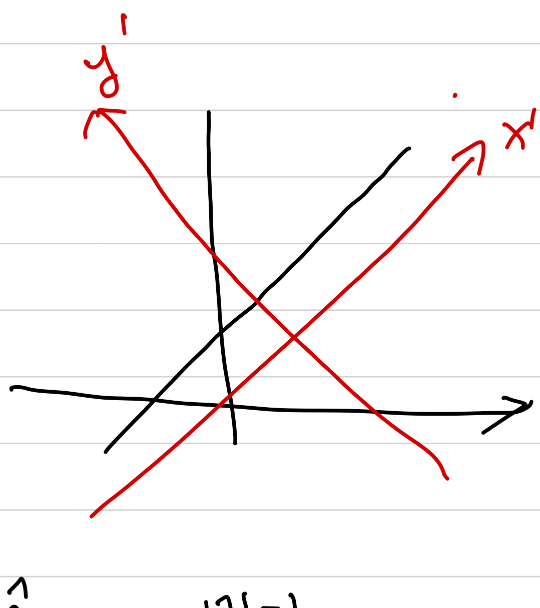
1.2. $x_i \in R^d$ $y_i \in R$

$L_S = \frac{1}{2}\sum_{i=1}^{n}(y_i - f_\theta(x_i))^2$

$f_\theta(x_i) = \sum_{j=1}^{m} a_j x_{ij} + b$

$= a^T x_i + b$

$\theta_i^{t+1} = \theta_i^t - \frac{\partial L_S}{\partial \theta_i}$



$f_\theta(x) = a^T x + b$ $a = A\hat{a}$ $|\hat{a}| = 1$

$= A\hat{a}^T x + b$

original data $\{(x_i, y_i)\}_{i=1}^{n}$ $x_i \in R^d$

$$x' = \hat{a}^T x \qquad S' = \{(x_i', y_i)\}_{i=1}^{n}$$

$$f_\theta(x') = Ax' + b \qquad A, b \in R \quad x' \in R$$

1.2. 分类问题

$$S = \{(x_i, y_i)\}_{i=1}^{n} \qquad x_i \in R^d \qquad y_i \in \{0,1\}$$

$$f_\theta(z) = \frac{1}{1+e^{-z}}$$

$f_\theta(z) = 1$ $\quad z = +\infty$

$f_\theta(z) = 0$ $\quad z = -\infty$

$f(z) = \frac{1}{2}$ $\quad z = 0$

$$\frac{1}{f_\theta(z)} = 1 + e^{-z}$$

$$\frac{1 - f_\theta(z)}{f_\theta(z)} = e^{-z}$$

$$z = w^T x + b = (w^T, b)\begin{pmatrix} x \\ 1 \end{pmatrix} = \beta^T \hat{x} = \beta^T x$$

$$f_\theta(x) = \frac{1}{1+e^{-\beta^T x}}$$

$$\ln \frac{1 - f_\theta(z)}{f_\theta(z)} = -z \qquad \ln \frac{f_\theta(x)}{1 - f_\theta(x)} = \beta^T x$$

~~线性回归~~

Assumption: Data are independent.

$$P(x_i) = f_\theta(x_i)^{y_i} \, (1 - f_\theta(x))^{1-y_i}$$

Likelihood:

$$\hat{L} = \prod_{i=1}^{n} P(x_i)$$

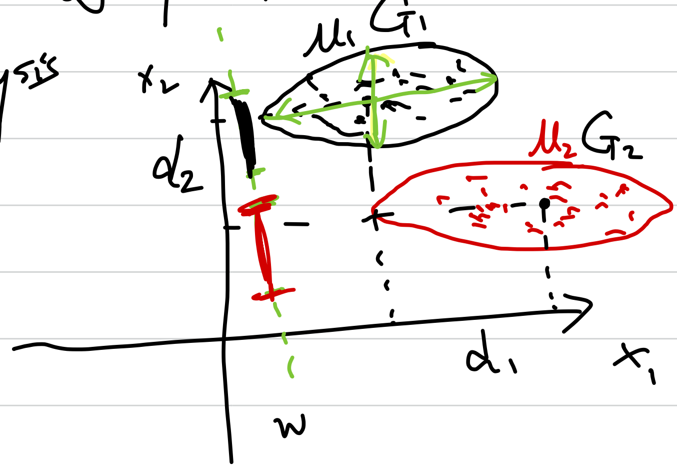
$$L_s = \log \hat{L} = \sum_{i=1}^{n} \log P(x_i)$$ Cross-Entropy

$$= \sum_{i=1}^{n} y_i \log f(x_i) + (1-y_i)(1 - f_\theta(x_i))$$

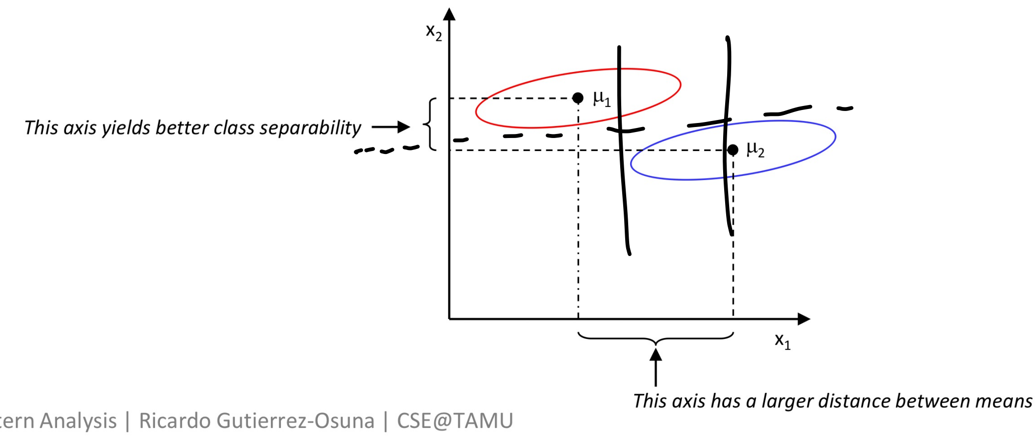
$$\theta_i^{t+1} = \theta_i^{t} - \eta \frac{\partial L_s}{\partial \theta_i}$$ 25:34

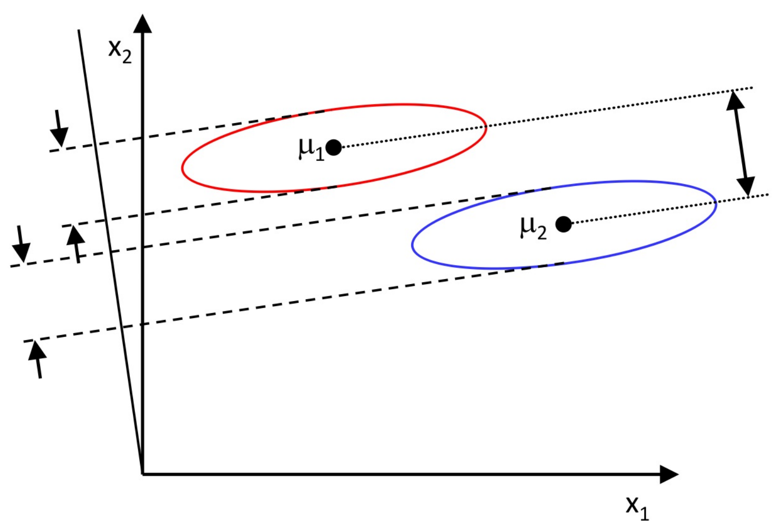
2 LDA 线性判别分析.

Linear discriminant analysis



2. 线性判别式





Dataset: Ω_1, Ω_2 Find w s.t distance max

$\mu_1 = \sum_{x \in \Omega_1} x$ $\mu_2 = \sum_{x \in \Omega_2} x$

$\tilde{\mu}_1 = w^T \mu_1$ $\tilde{\mu}_2 = w^T \mu_2$

First define a measure of scatter in feature space x

$$S_i = \sum_{x \in \Omega_i} (x-\mu_i)(x-\mu_i)^T$$

$X \in R^{d\times n}$ $x_i \in R^{1\times n}$

$$C_{ij} = \sum_{k=1}^{n} (X_{ik}-\mu_i)(X_{jk}-\mu_j)$$
$$= (X_i-\mu_i)(X_j-\mu_j)^T$$

$$C = (X-\bar{X})(X-\bar{X})^T$$

$$S_W = S_1 + S_2$$

within-class scatter matrix

$$(\tilde{\mu}_1-\tilde{\mu}_2)^2$$
$$= (w^T(\mu_1-\mu_2))^2$$
$$= w^T(\mu_1-\mu_2)(\mu_1-\mu_2)^T w$$
$$\triangleq w^T S_B w$$

S_B = between-class scatter

$$J = \frac{w^T S_B w}{w^T S_W w}$$. Fisher criterion (1936)

$S_W^{-1} S_B$

$$\max_{w} J$$

$$\frac{dJ}{dw} = \frac{S_B w (w^T S_W w) - S_W w (w^T S_B w)}{(\quad)^2} = 0$$

$$S_B w = S_W w \cdot \frac{w^T S_B w}{w^T S_W w}$$

$$S_W^{-1} S_B w = J w$$ 特征值问题

$u=(u_1-u_2)$

$$S_w^{-1}uu^Tw = Jw$$

$$\boxed{J = u^Tw} \qquad \boxed{w = S_w^{-1}u}$$

$$S_w^{-1}uu^Tw = \frac{w^Tuu^Tw}{w^TS_ww}w$$

$$= \frac{w^Tuu^Tw}{u^TS_w^{-1}S_wS_w^{-1}u}w$$

$$= \frac{u^TS_w^{-1}u\,u^Tw}{u^TS_w^{-1}u}w$$

$$w\cdot u^Tw = u^Tw\cdot w \quad ✓$$

$$\boxed{ab^Tc = b^Tca}$$

3. 多分类问题

① one vs one OvO ② one vs Rest OvR ③ Many vs Many MvM

N : C_N^2

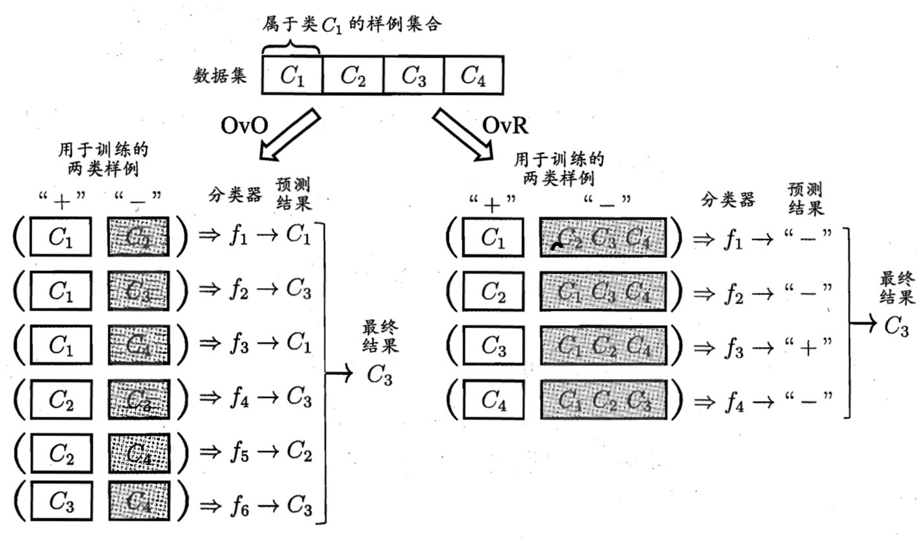
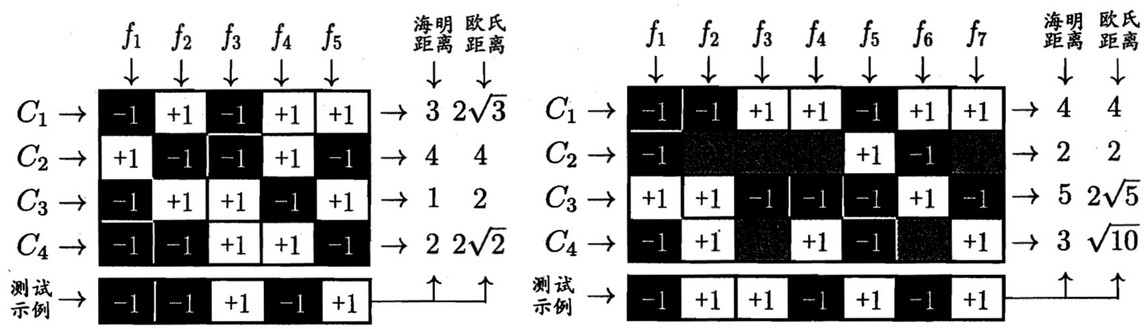


图 3.4 OvO 与 OvR 示意图

纠错输出编码 Error Correcting output codes

ECOC



(a) 二元 ECOC 码　　(b) 三元 ECOC 码

图 3.5 ECOC 编码示意图. “+1”、“−1”分别表示学习器 f_i 将该类样本作为正、反例; 三元码中“0”表示 f_i 不使用该类样本

There are three kinds of lies:

Lies, damned lies, and statistics

本科生课程：统计计算与机器学习

许志钦

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Lecture 7: **Yule-Simpson’s paradox**

2020年春

招生录取中的性别歧视

- “校长，不好了，有很多男生在校门口抗议，他们说今年研究所女生录取率42%是男生21%的两倍，我们学校遴选学生有性别歧视”，校长满脸疑惑的问秘书：“我不是特别交代，今年要尽量提升男生录取率以免落人口实吗？”
- 秘书赶紧回答说：“确实有交代下去，我刚刚也查过，的确是有注意到，今年商学院录取率是男性75%，女性只有49%；而法学院录取率是男性10%，女性为5%。二个学院都是男生录取率比较高，校长这是我作的调查报告。”
- “秘书，你知道为什么个别录取率男皆大于女，但是总体录取率男却远小于女吗？”

百度百科

招生录取中的性别歧视

学院	女生 申请	女生 录取	女生 录取率	男生 申请	男生 录取	男生 录取率	合计 申请	合计 录取	合计 录取率
商学院	100	49	49%	20	15	75%	120	64	53.3%
法学院	20	1	5%	100	10	10%	120	11	9.2%
总计	120	50	42%	120	25	21%	240	75	31.3%

百度百科

治疗效果是好是坏?

Table 1: Yule-Simpson's Paradox

Population	Survive	Die	Survive Rate
Treatment	20	20	50%
Control	16	24	40%
Male	Survive	Die	Survive Rate
Treatment	18	12	60%
Control	7	3	70%
Female	Survive	Die	Survive Rate
Treatment	2	8	20%
Control	9	21	30%

WHY?

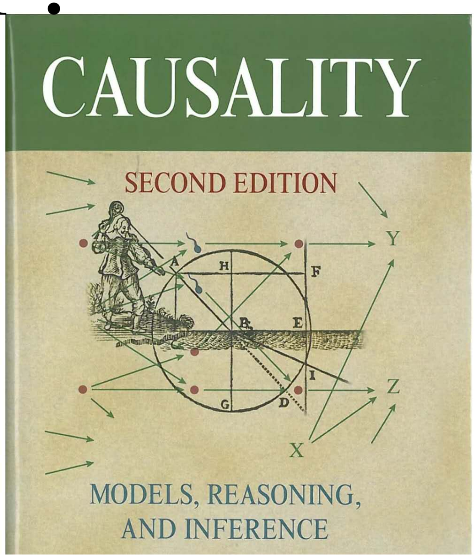
Perl 2000

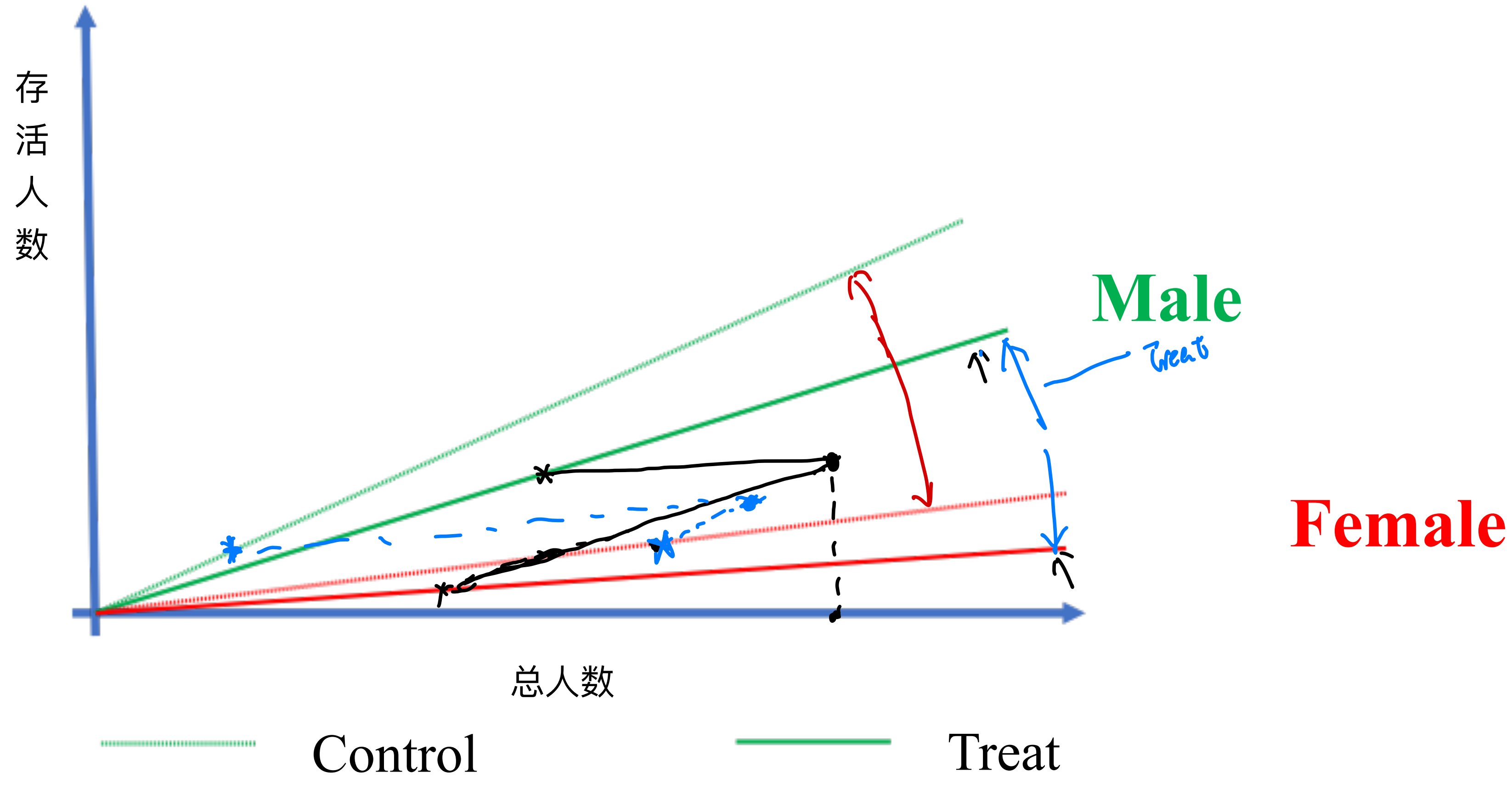
7:48

因果推断简介之一：从 Yule-Simpson's Paradox 讲起

丁鹏 统计之都 2019-12-05

编辑部于2019年10月在微信端开启《朝花夕拾》栏目，目的是推送2013年（含）之前主站发表的优秀文章，微信端与主站的同步始于2013年年初，然而初期用户量有限，故优质文章可能被埋没。



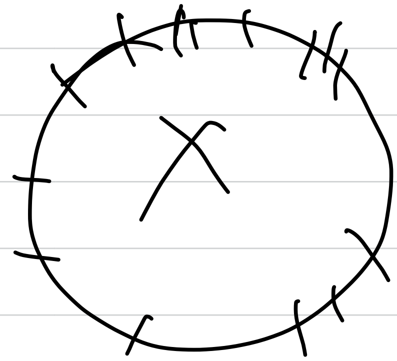


存活人数
Male
Treat
Female
总人数
Control
Treat

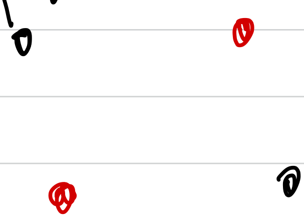
m^+ m^-

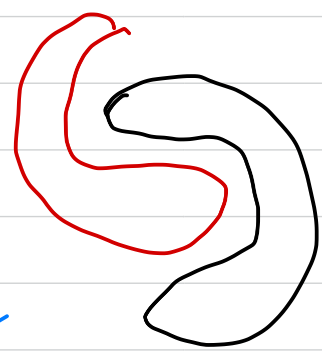
$$\frac{y}{1-y} > 1$$

$$\boxed{\frac{y}{1-y} > \frac{m^+}{m^-}}$$

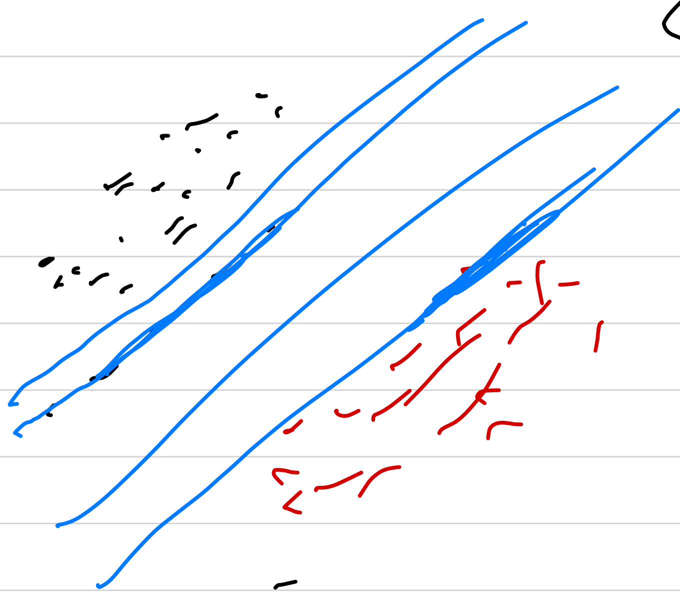


① 线性可分





②



③