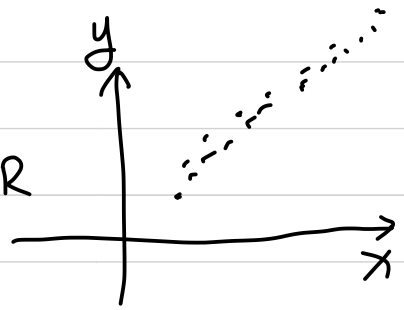


第七课. 线性问题

1. (1) 3

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



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

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

GD. SGD.

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

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

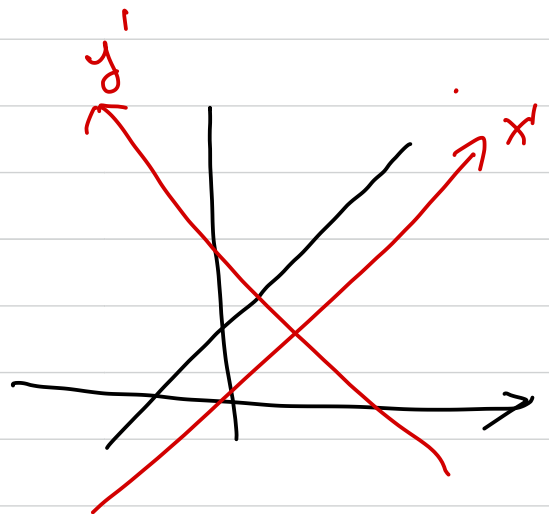
1.2. $x_i \in \mathbb{R}^d \quad y_i \in \mathbb{R}$

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

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

$$= a^T x_i + b$$

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



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

$$a = A \hat{a}$$

$$|\hat{a}| = 1$$

original data

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

$$x' = \hat{a}^T x$$

$$S' = \{(x'_i, y_i)\}_{i=1}^n$$

$$f_{\theta}(x') = Ax' + b$$

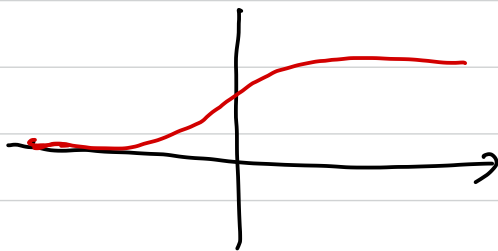
$$A, b \in \mathbb{R} \quad x' \in \mathbb{R}$$

1.2.

分类问题

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

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



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

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

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

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

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

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

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

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

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

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

Assumption: Data are independent.

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

Likelihood:

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

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

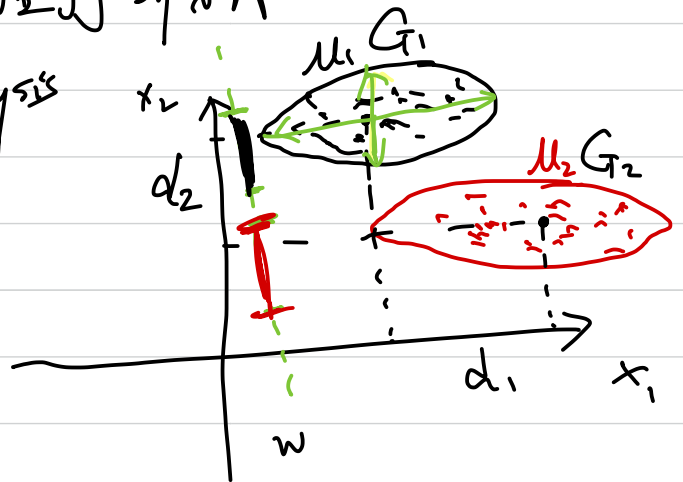
$$= \sum_{i=1}^n y_i \log f_{\theta}(x_i) + (1-y_i) \log (1 - f_{\theta}(x_i))$$

$$\theta_i^{\text{opt}} = \theta_i^{\text{tr}} - \eta \frac{\partial L_s}{\partial \theta_i}$$

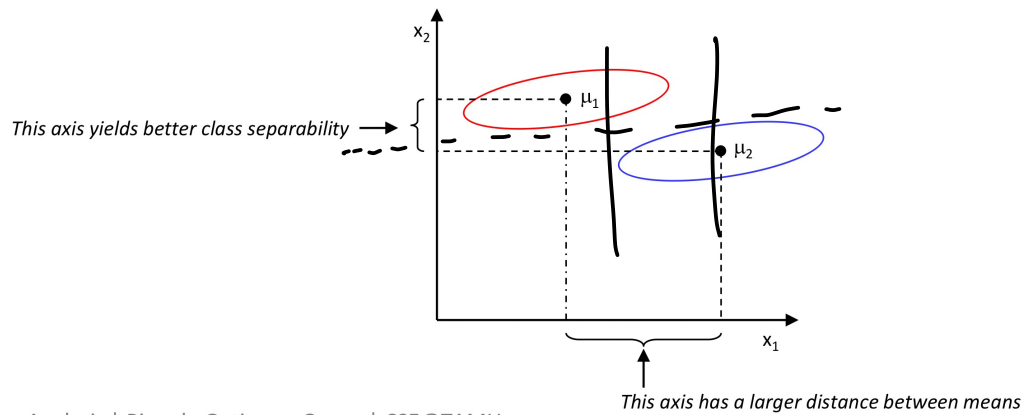
25 = 34

2 LDA 线性判别分析

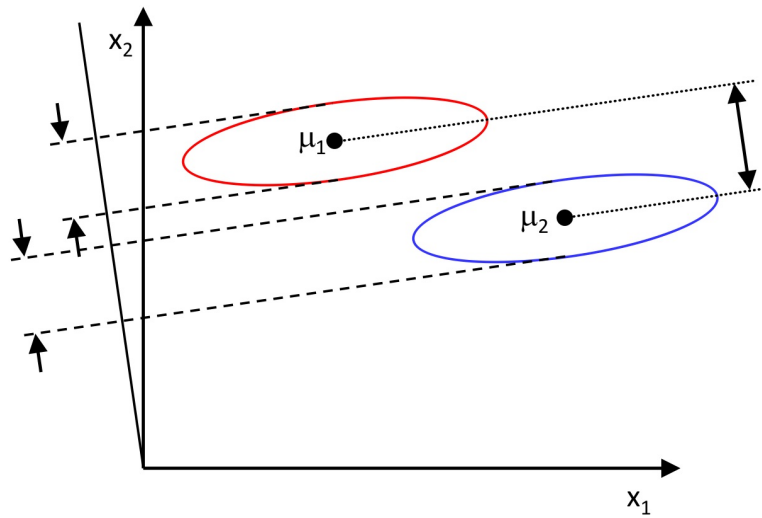
Linear discriminant analysis



2. 线性判别分析



CSC 666 Pattern Analysis | Ricardo Gutierrez-Osuna | CSE@TAMU



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

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

$$\left. \begin{aligned} X \in \mathbb{R}^{d \times n} \quad x_i \in \mathbb{R}^{1 \times n} \\ C_{ij} &= \sum_{k=1}^n (X_{ik} - \mu_i)(X_{jk} - \mu_j) \\ &= (X_i - \mu_i)(X_j - \mu_j)^T \\ C &= (X - \bar{X})(X - \bar{X})^T \end{aligned} \right\}$$

$$S_w = S_1 + S_2$$

within-class scatter matrix

$$\begin{aligned} &(\tilde{\mu}_1 - \tilde{\mu}_2)^2 \\ &= (W^T(\mu_1 - \mu_2))^2 \\ &= W^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T W \\ &\triangleq W^T S_B W \end{aligned}$$

S_B : between-class scatter

$$J = \frac{W^T S_B W}{W^T S_W W} \quad \text{Fisher criterion (1936)}$$

$$\max_W J \quad S_W^{-1} S_B$$

$$\frac{dJ}{dW} = \frac{S_B W (W^T S_W W) - S_W W (W^T S_B W)}{(W^T S_W W)^2} = 0$$

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

$$S_W^{-1} S_B W = J W \quad \text{特征值问题}$$

$$u = (\mu_1 - \mu_2)$$

$$S_w^{-1} u u^T w = J w$$

$$\boxed{J = u^T w} \quad \boxed{w = S_w^{-1} u}$$

$$\begin{aligned} S_w^{-1} u u^T w &= \frac{w^T u u^T w}{w^T S_w w} w \\ &= \frac{w^T u u^T w}{u^T S_w^{-1} S_w S_w^{-1} u} w \\ &= \frac{u^T S_w^{-1} u u^T w}{u^T S_w^{-1} u} w \end{aligned}$$

$$w \cdot u^T w = u^T w \cdot w \quad \checkmark$$

$$\boxed{a b^T c = b^T c a}$$

3. 多分类问题

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

$N: C_N^2$

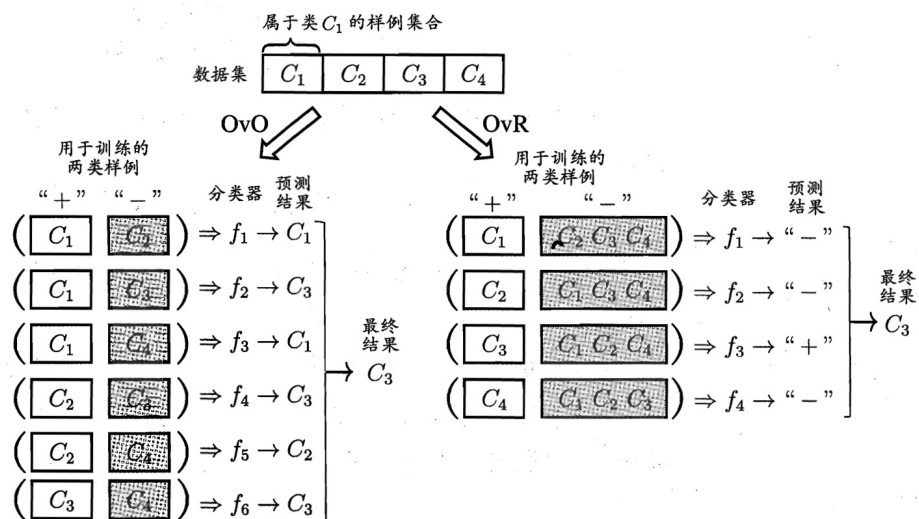


图 3.4 OvO 与 OvR 示意图

纠错输出编码 Error Correcting output codes ECOC

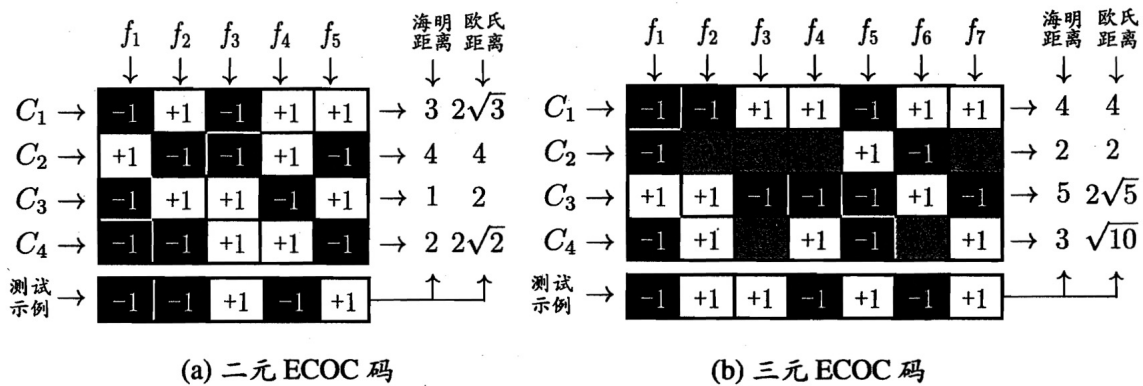


图 3.5 ECOC 编码示意图。“+1”、“-1”分别表示学习器 f_i 将该类样本作为正、反例；三元码中“0”表示 f_i 不使用该类样本

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



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

Lecture 7: Yule-Simpson's paradox

2020年春季

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？”

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%

治疗效果是好是坏？

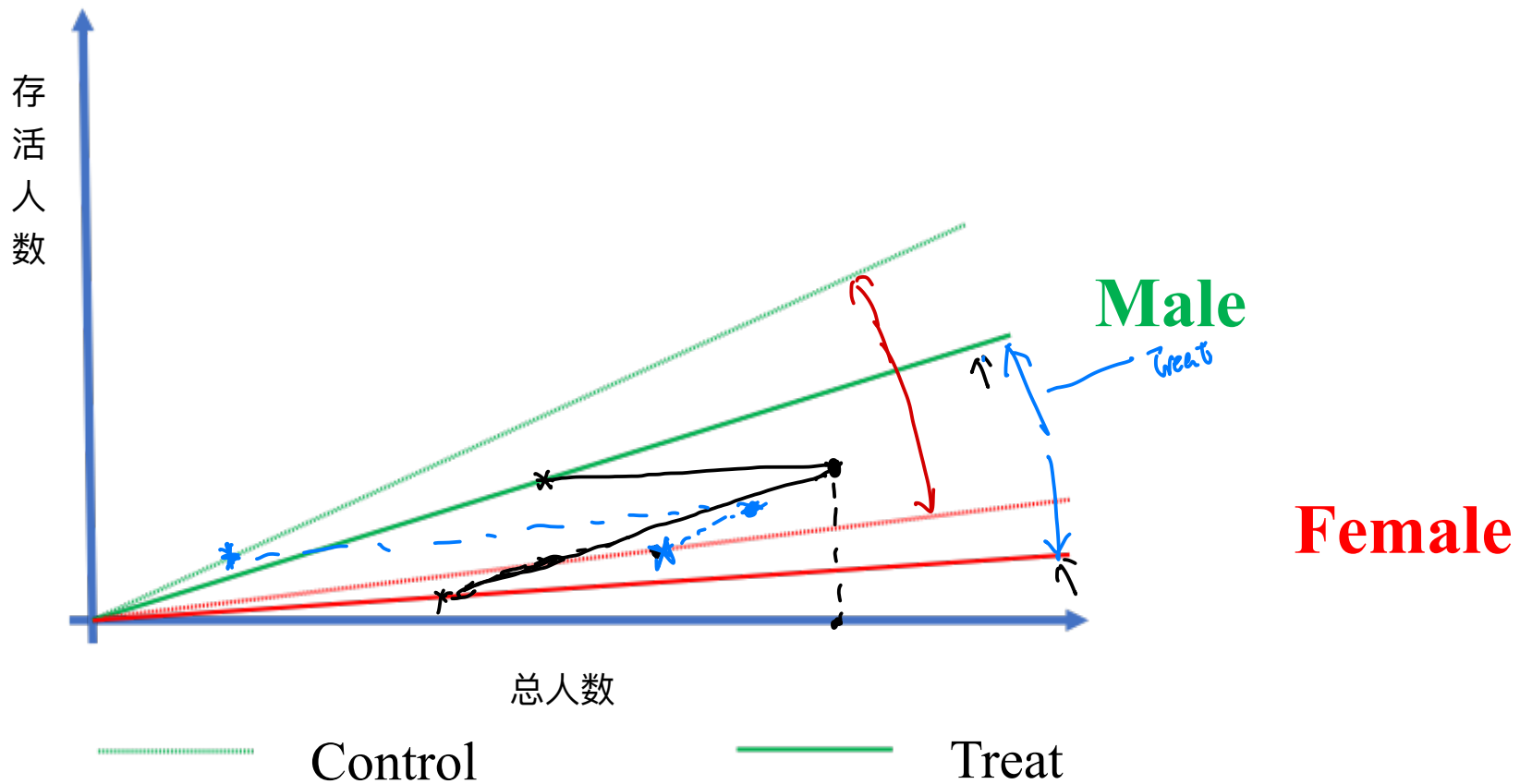
Table 1: Yule-Simpson's Paradox

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%

Perl 2000

WHY?

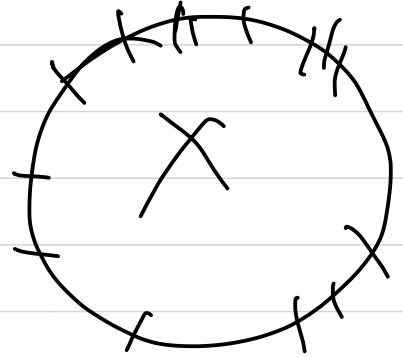
The screenshot shows a mobile browser interface with a dark status bar at the top containing signal strength, Wi-Fi, and battery icons. The browser address bar shows a close button (X) and a menu button (three dots). The main content area displays the title '因果推断简介之一：从 Yule-Simpson's Paradox 讲起' (Introduction to Causal Inference I: From Yule-Simpson's Paradox) by '丁鹏 统计之都' (Ding Peng, StatCity) dated '2019-12-05'. A paragraph of text explains that the 'Morning and Evening' column was launched in October 2019 to promote high-quality articles from before 2013, but some quality articles might be buried. At the bottom right, there is a book cover for 'CAUSALITY: MODELS, REASONING, AND INFERENCE, SECOND EDITION' by Judea Pearl. The cover features a diagram of a person in a hat standing next to a large circular diagram with various labeled points (A, B, C, D, E, F, G, H, I, X, Y, Z) and arrows indicating causal relationships.



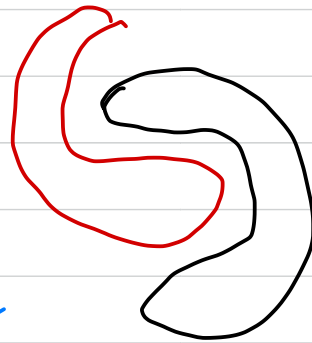
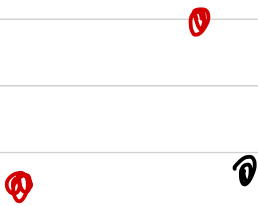
m^+ m^-

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

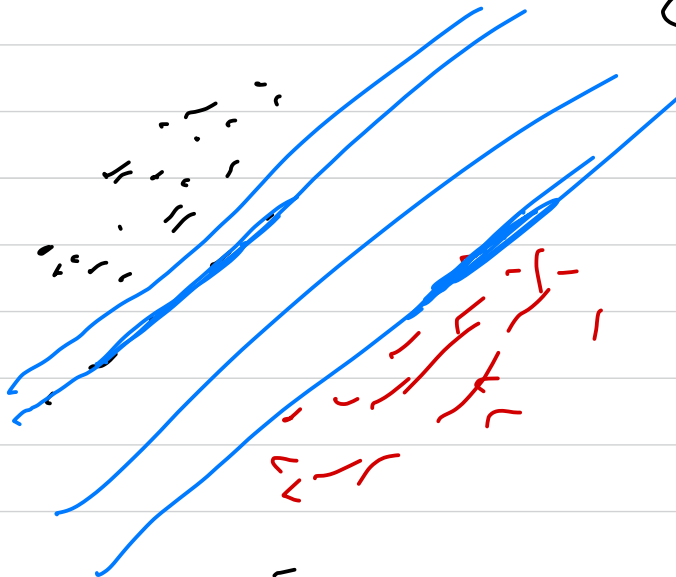
$\frac{y_D}{1-y_D} > \frac{m^+}{m^-}$



① ~~2/2~~ 性能の分



②



③

