逻辑斯蒂回归

赵海涛

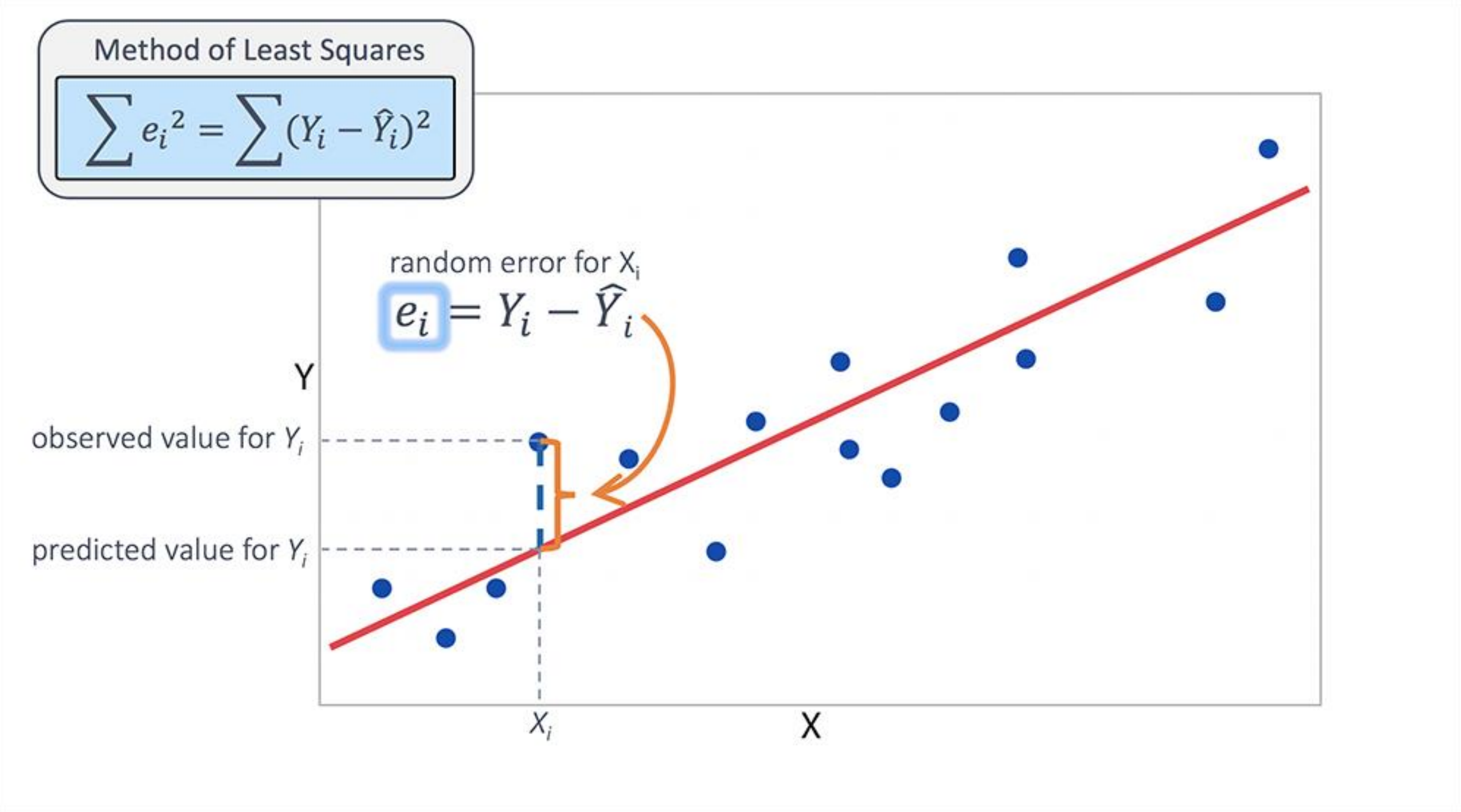
haitaozhao@ecust.edu.cn

大纲

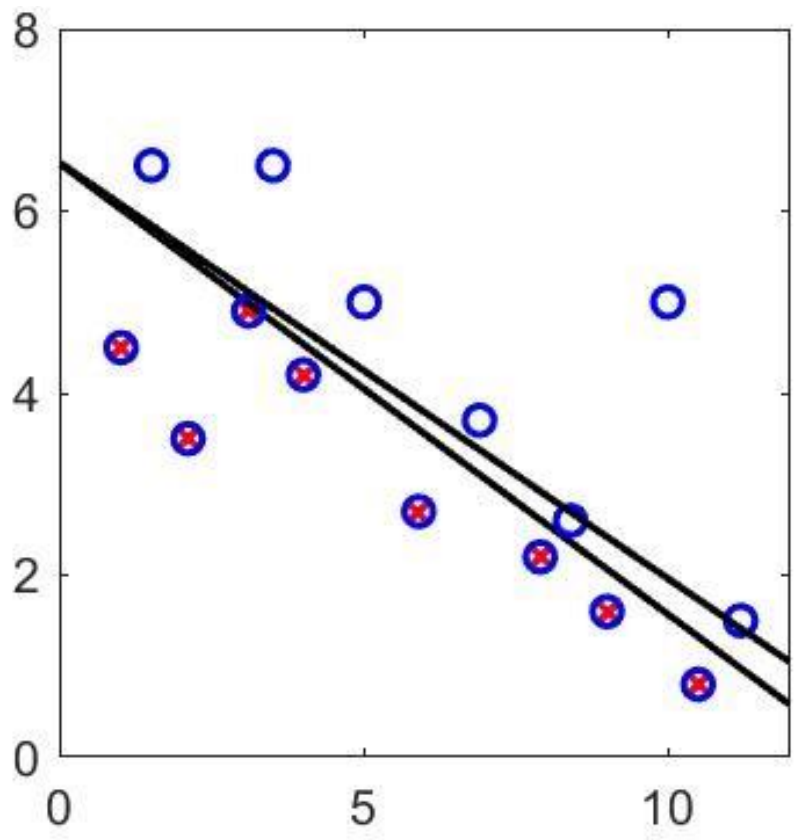
- Logistic回归的模型
- Logistic回归的策略
- Logistic回归的算法
- 优化算法

线性回归 （Linear Regression）

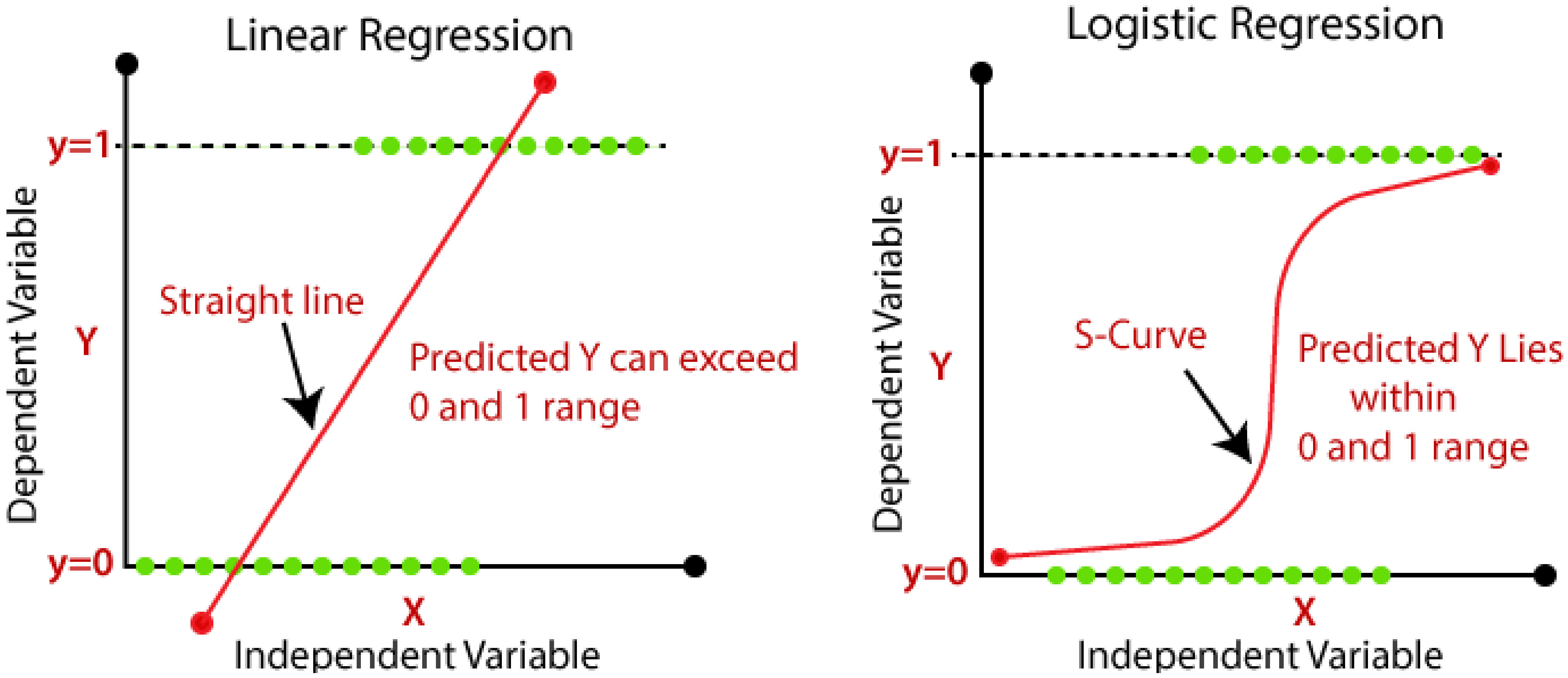
面积 (m^2)	销售价钱（万元）
123	250
150	320
87	160
102	220
…	…



线性回归 （Linear Regression）



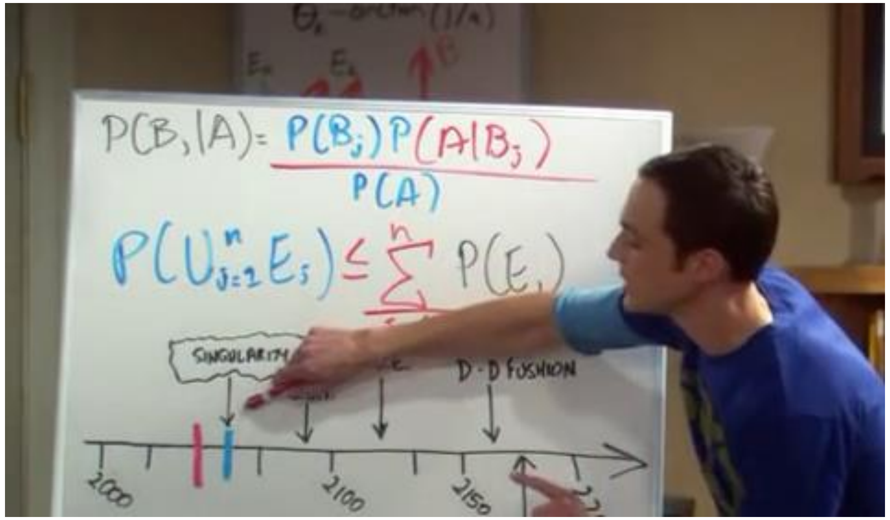
线性回归与逻辑斯蒂回归



Logistic 回归的模型

针对二分类问题（$y_i \in \{-1,1\}$），当判别函数$g_i(x) > 0.5$, 即可认为$y_i = 1$, $g_i\,(x) < 0.5$, 即可认为$y_i = -1$。取后验概率作为判别函数$g_i(x)$:

$$P(\omega_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p\left(\boldsymbol{x}|\omega_j\right)P\left(\omega_j\right)}$$



Bayes' rule

Logistic 回归的模型

针对二分类问题，只有ω_1和ω_2两类，$P(\omega_1|\boldsymbol{x})+P(\omega_2|\boldsymbol{x})=1$，假设两类样本个数相同， 即$P(\omega_1)=P(\omega_2)$，可得

$$P(\omega_i|\boldsymbol{x})=\frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{p(\boldsymbol{x}|\omega_1)P(\omega_1)+p(\boldsymbol{x}|\omega_2)P(\omega_2)} \quad （i=1,2）$$

$$=\frac{p(\boldsymbol{x}|\omega_i)}{p(\boldsymbol{x}|\omega_1)+p(\boldsymbol{x}|\omega_2)} \quad （i=1,2）$$

Logistic 回归的模型

取d维多元正态密度函数作为概率密度函数：

$$p(\boldsymbol{x}|\omega_i) = \mathcal{N}(\boldsymbol{x};\ \mu_i, \sigma)$$

$$p(\boldsymbol{x}|\omega_i) = \frac{1}{\left(\sqrt{2\pi\sigma}\right)^d} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_i)^T(\boldsymbol{x}-\boldsymbol{\mu}_i)}{2\sigma^2}\right)$$

$$\Rightarrow P(\omega_1|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_1)+p(\boldsymbol{x}|\omega_2)}$$ （将$p(\boldsymbol{x}|\omega_i)$代入）

$$= \frac{\frac{1}{\left(\sqrt{2\pi\sigma}\right)^d}\exp\left(-\frac{(x-\boldsymbol{\mu}_1)^T(x-\boldsymbol{\mu}_1)}{2\sigma^2}\right)}{\frac{1}{\left(\sqrt{2\pi\sigma}\right)^d}\exp\left(-\frac{(x-\boldsymbol{\mu}_1)^T(x-\boldsymbol{\mu}_1)}{2\sigma^2}\right)+\frac{1}{\left(\sqrt{2\pi\sigma}\right)^d}\exp\left(-\frac{(x-\boldsymbol{\mu}_2)^T(x-\boldsymbol{\mu}_2)}{2\sigma^2}\right)}$$

Logistic 回归的模型

$$= \frac{\exp\left(-\frac{x^T x - 2\mu_1{}^T x + \mu_1{}^T \mu_1}{2\sigma^2}\right)}{\exp\left(-\frac{x^T x - 2\mu_1{}^T x + \mu_1{}^T \mu_1}{2\sigma^2}\right) + \exp\left(-\frac{x^T x - 2\mu_2{}^T x + \mu_2{}^T \mu_2}{2\sigma^2}\right)}$$

$$= \frac{1}{1 + \exp\left[\left(-\frac{x^T x - 2\mu_2{}^T x + \mu_2{}^T \mu_2}{2\sigma^2}\right) - \left(-\frac{x^T x - 2\mu_1{}^T x + \mu_1{}^T \mu_1}{2\sigma^2}\right)\right]}$$

$$= \frac{1}{1 + \exp\left(-\frac{(\mu_1{}^T - \mu_2{}^T)x}{\sigma^2} + \frac{\mu_1{}^T \mu_1 - \mu_2{}^T \mu_2}{2\sigma^2}\right)}$$

$$= \frac{1}{1 + \exp(-(\boldsymbol{w}^T x + b))}$$

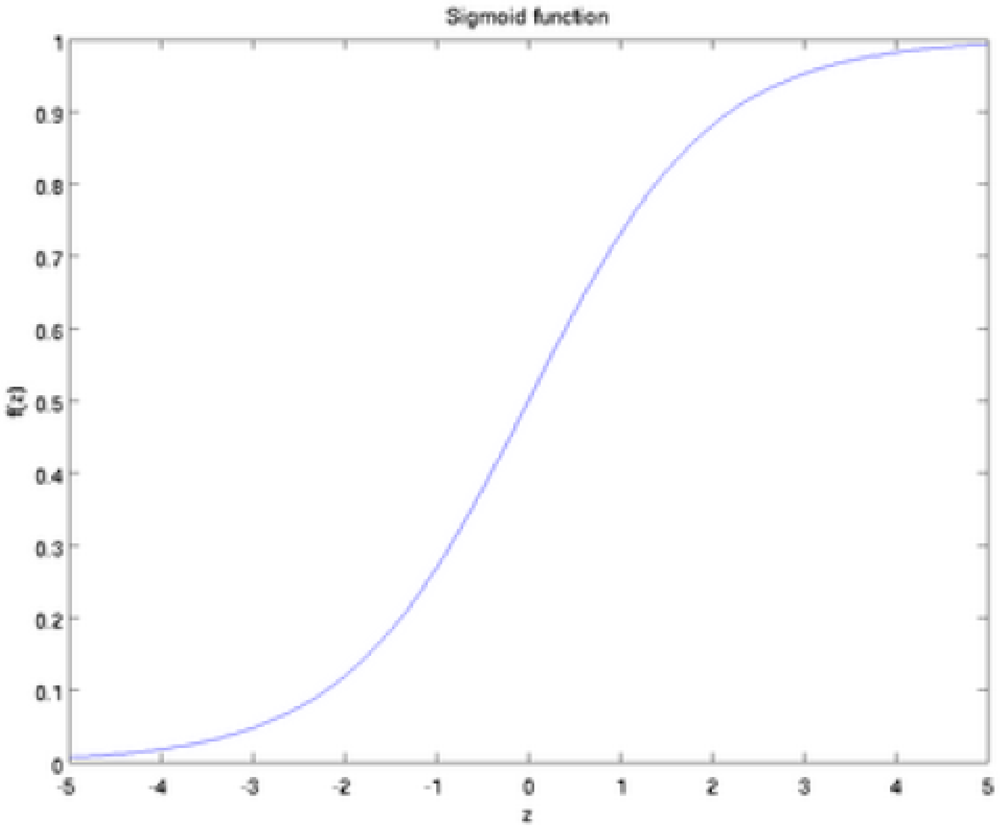
$$= \frac{1}{1 + \exp(-z)} = \sigma(z)$$

Logistic 回归的模型

Sigmoid:

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

$$\sigma'(z) = \sigma(z)\big(1 - \sigma(z)\big)$$



Logistic 回归的模型

$$P(\omega_1|\boldsymbol{x}) = P(y=1|\boldsymbol{x}) = \frac{1}{1+\exp(-\boldsymbol{W}^T\boldsymbol{x})}$$

$$P(\omega_2|\boldsymbol{x}) = P(y=-1|\boldsymbol{x}) = 1 - P(\omega_1|\boldsymbol{x})$$

$$= 1 - \frac{1}{1+\exp(-\boldsymbol{W}^T\boldsymbol{x})}$$

$$= \frac{\exp(-\boldsymbol{W}^T\boldsymbol{x})}{1+\exp(-\boldsymbol{W}^T\boldsymbol{x})}$$

$$= \frac{1}{1+\exp(\boldsymbol{W}^T\boldsymbol{x})}$$

其中$\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x}\\1\end{bmatrix}$，$\boldsymbol{W} = \begin{bmatrix}\boldsymbol{w}\\b\end{bmatrix}$。根据结果可得$P(\omega_i|\boldsymbol{x}) = \dfrac{1}{1+\exp(-y_i\boldsymbol{W}^T\boldsymbol{x})}$

Logistic 回归的策略

- logistic分类器是由一组权值系数组成的，最关键的问题就是如何获取这组权值，通过极大似然函数估计获得
- 似然函数是统计模型中参数的函数。给定输出$\boldsymbol{x}$时，关于参数$\boldsymbol{\theta}$的似然函数$L(\boldsymbol{\theta}|\boldsymbol{x})$（在数值上）等于给定参数$\boldsymbol{\theta}$后变量$X$的概率：

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = P(X = \boldsymbol{x}; \boldsymbol{\theta})$$

- 似然函数的重要性不是它的取值，而是当参数变化时概率密度函数到底是变大还是变小。
- 最大似然估计：似然函数取得最大值表示相应的参数能够使得统计模型最为合理

最大似然估计

- N noisy measurements are made to observe the constant μ:

$$z_i = \mu + v_i \quad i = 1, 2, \cdots, N \quad \textit{with} \quad \textit{i.i.d.}\ v_i \sim N(0, \sigma^2)$$

- For $\theta = (\mu, \sigma^2)$,
$p(z_1, z_2, \cdots, z_N|\theta) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i - \mu)^2)$
- MLE:

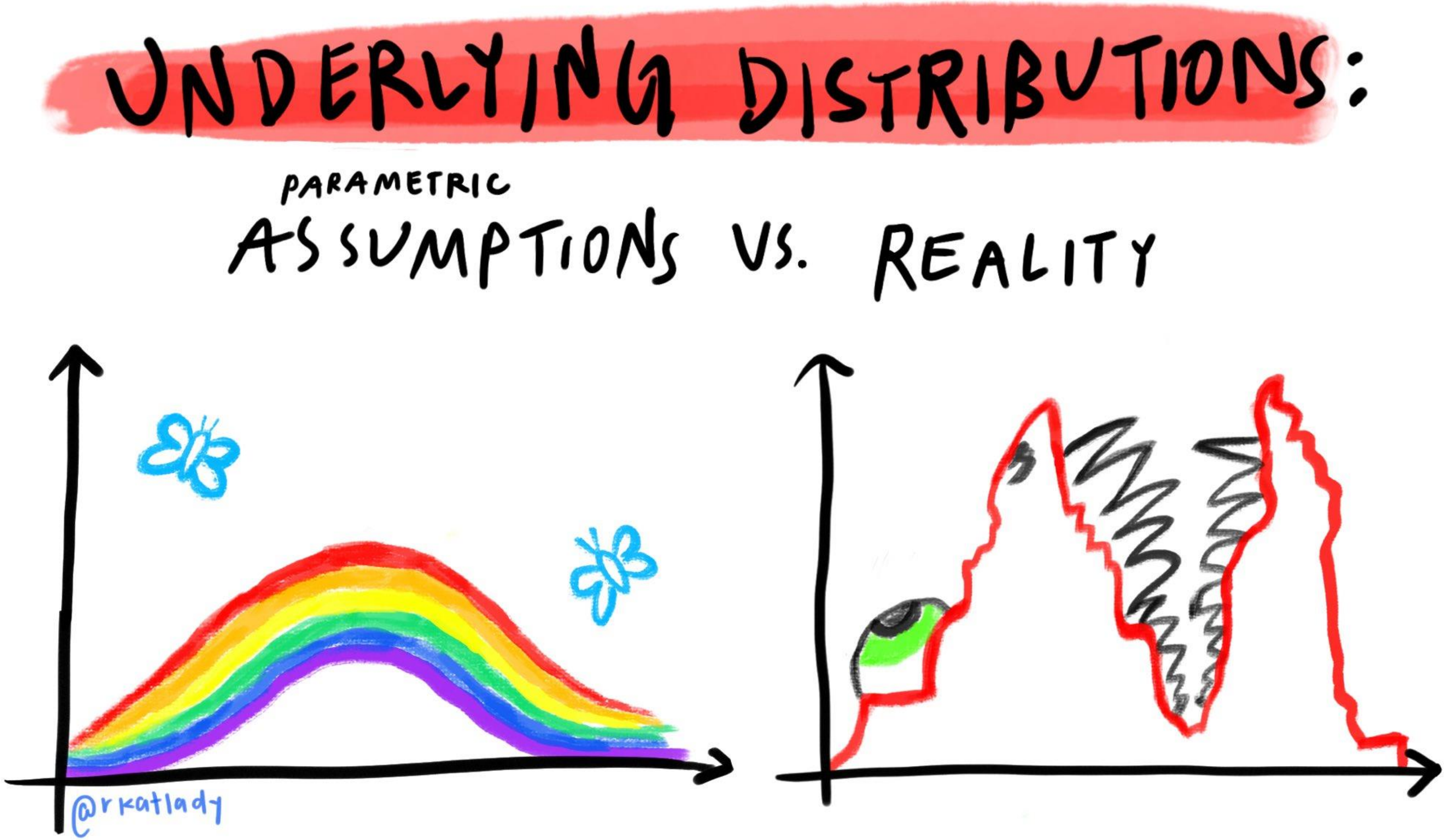
$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} z_i = \textit{sample mean}$$

$$\hat{\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{\mu})^2 = \textit{sample variance}$$

最大似然估计

- Result of tossing a coin is $\in$ {Heads, Tails}
- Random var $X \in \{1,0\}$
    - Bernoulli: $P(X = x) = p_0^x (1 - p_0)^{(1-x)}$
- Sample: $X = \{x^t\}_{t=1}^N$
- $\ln p(D|\theta) = \ln \prod_{t=1}^N p_0^{x^t} (1 - p_0)^{(1-x^t)}$
- MLE $\hat{p}_0 = \frac{\sum_{i=1}^N x^i}{N}$

最大似然估计


UNDERLYING DISTRIBUTIONS:
PARAMETRIC
ASSUMPTIONS VS. REALITY
@rkatlady

Logistic 回归的目标函数

- 由于$\boldsymbol{x}_i$是独立同分布（i.i.d.），似然函数为：

$$L(W)=\prod_{i=1}^{N}p(\omega_i|\boldsymbol{x}_i)=\prod_{i=1}^{N}\frac{1}{1+\exp(-y_iW^T\boldsymbol{x}_i)}$$

- 一般使用似然函数的负对数函数（Negative Log Likelihood，NLL）来作为Logistic回归的损失函数：

$$NLL(W)=-\ln L(W)=-\sum_{i=1}^{N}\ln\frac{1}{1+\exp(-y_iW^T\boldsymbol{x}_i)}=-\sum_{i=1}^{N}\ln\sigma(z_i)$$

$$\Downarrow$$

$$(z_i=y_iW^T\boldsymbol{x}_i)$$

Logistic 回归的算法

- Logistic回归就是要求W,使得$NLL(W)$最小。
- 用梯度下降法来求解：

$$\frac{\partial NLL}{\partial W} = -\sum_{i=0}^{\mathrm{N}} \frac{1}{\sigma(z_i)} \sigma(z_i)[1-\sigma(z_i)] y_i \boldsymbol{x}_i$$

$$= -\sum_{i=1}^{N} \frac{\exp(-y_i W^T \boldsymbol{x}_i)}{1+\exp(-y_i W^T \boldsymbol{x}_i)} y_i \boldsymbol{x}_i$$

$$= -\sum_{i=1}^{N} \frac{1}{1+\exp(y_i W^T \boldsymbol{x}_i)} y_i \boldsymbol{x}_i$$

Logistic 回归的算法

- 用牛顿法来求解：

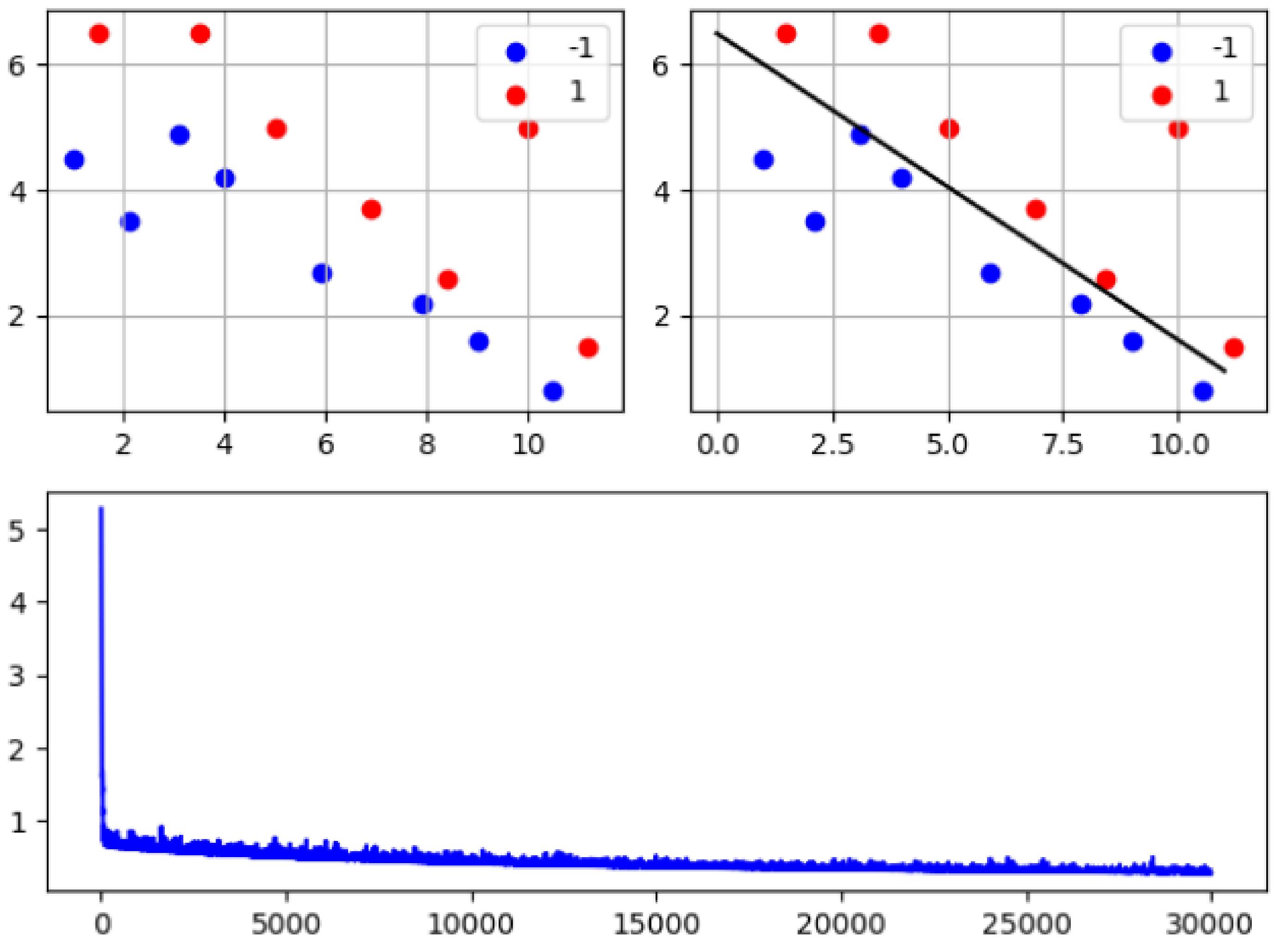
$$\frac{\partial^2 NLL}{\partial W^T \partial W} = \sum_{i=1}^{N} \frac{\exp(y_i W^T \boldsymbol{x}_i)\boldsymbol{x}_i \boldsymbol{x}_i{}^T}{(1+\exp(y_i W^T \boldsymbol{x}_i))^2}$$

$$= X diag[\sigma(-y_i W^T \boldsymbol{x}_i)(1-\sigma(-y_i W^T \boldsymbol{x}_i))]X^T$$

$$= \ XSX^T \qquad （X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots \boldsymbol{x}_N], S\text{是一个对角阵}）$$

可以发现 XSX^T 半正定。

Logistic 回归的例子



谢谢各位同学！