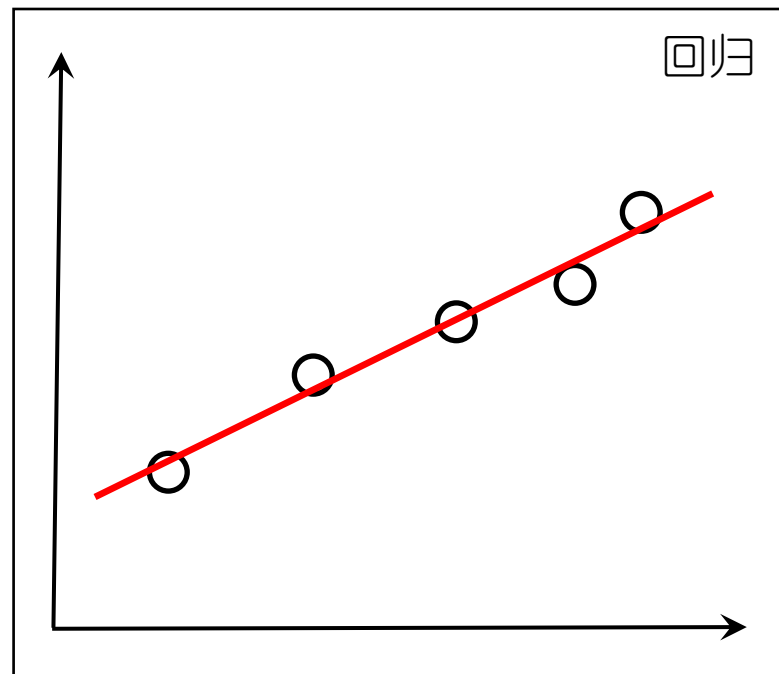
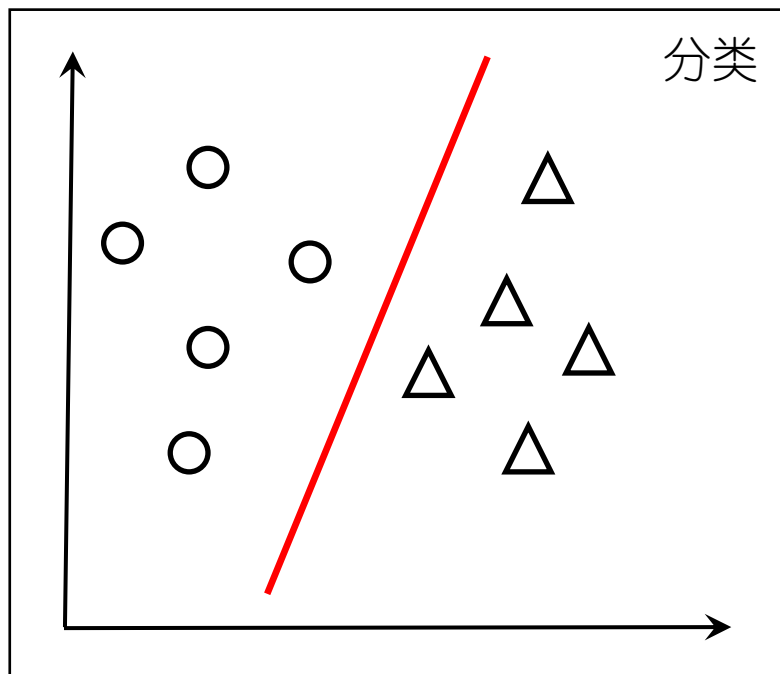


## 三、线性模型

主讲教师：周志华

---

# 线性模型



线性模型(linear model)试图学得一个通过属性的线性组合来进行预测的函数

$$f(x) = w_1x_1 + w_2x_2 + \dots + w_dx_d + b$$

向量形式:  $f(x) = w^T x + b$

简单、基本、可理解性好

# 线性回归 (linear regression)

---

$$f(x_i) = wx_i + b \text{ 使得 } f(x_i) \simeq y_i$$

离散属性的处理：若有“序”(order)，则连续化；  
否则，转化为 k 维向量

$$\begin{aligned} \text{令均方误差最小化, 有 } (w^*, b^*) &= \arg \min_{(w, b)} \sum_{i=1}^m (f(x_i) - y_i)^2 \\ &= \arg \min_{(w, b)} \sum_{i=1}^m (y_i - wx_i - b)^2 \end{aligned}$$

$$\text{对 } E_{(w, b)} = \sum_{i=1}^m (y_i - wx_i - b)^2 \text{ 进行最小二乘参数估计}$$

# 线性回归

---

分别对  $w$  和  $b$  求导：

$$\frac{\partial E_{(w,b)}}{\partial w} = 2 \left( w \sum_{i=1}^m x_i^2 - \sum_{i=1}^m (y_i - b) x_i \right)$$
$$\frac{\partial E_{(w,b)}}{\partial b} = 2 \left( mb - \sum_{i=1}^m (y_i - wx_i) \right)$$

令导数为 **0**，得到闭式(closed-form)解：

$$w = \frac{\sum_{i=1}^m y_i (x_i - \bar{x})}{\sum_{i=1}^m x_i^2 - \frac{1}{m} \left( \sum_{i=1}^m x_i \right)^2} \quad b = \frac{1}{m} \sum_{i=1}^m (y_i - wx_i)$$

# 多元(multi-variate)线性回归

$$f(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b \quad \text{使得} \quad f(\mathbf{x}_i) \simeq y_i$$

$$\mathbf{x}_i = (x_{i1}; x_{i2}; \dots; x_{id}) \quad y_i \in \mathbb{R}$$

把  $\mathbf{w}$  和  $b$  吸收入向量形式  $\hat{\mathbf{w}} = (\mathbf{w}; b)$ , 数据集表示为

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T & 1 \\ \mathbf{x}_2^T & 1 \\ \vdots & \vdots \\ \mathbf{x}_m^T & 1 \end{pmatrix} \quad \mathbf{y} = (y_1; y_2; \dots; y_m)$$

# 多元线性回归

同样采用最小二乘法求解，有

$$\hat{\mathbf{w}}^* = \arg \min_{\hat{\mathbf{w}}} (\mathbf{y} - \mathbf{X}\hat{\mathbf{w}})^T (\mathbf{y} - \mathbf{X}\hat{\mathbf{w}})$$

令  $E_{\hat{\mathbf{w}}} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{w}})^T (\mathbf{y} - \mathbf{X}\hat{\mathbf{w}})$ ，对  $\hat{\mathbf{w}}$  求导：

$$\frac{\partial E_{\hat{\mathbf{w}}}}{\partial \hat{\mathbf{w}}} = 2\mathbf{X}^T (\mathbf{X}\hat{\mathbf{w}} - \mathbf{y}) \quad \text{令其为零可得 } \hat{\mathbf{w}}$$

然而，麻烦来了：涉及矩阵求逆！

□ 若  $\mathbf{X}^T\mathbf{X}$  满秩或正定，则  $\hat{\mathbf{w}}^* = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T\mathbf{y}$

□ 若  $\mathbf{X}^T\mathbf{X}$  不满秩，则可解出多个  $\hat{\mathbf{w}}$

此时需求助于归纳偏好，或引入 正则化 (regularization) → 第6、11章

# 线性模型的变化

对于样例  $(x, y)$ ,  $y \in \mathbb{R}$ , 若希望线性模型的预测值逼近真实标记, 则得到线性回归模型  $y = w^T x + b$

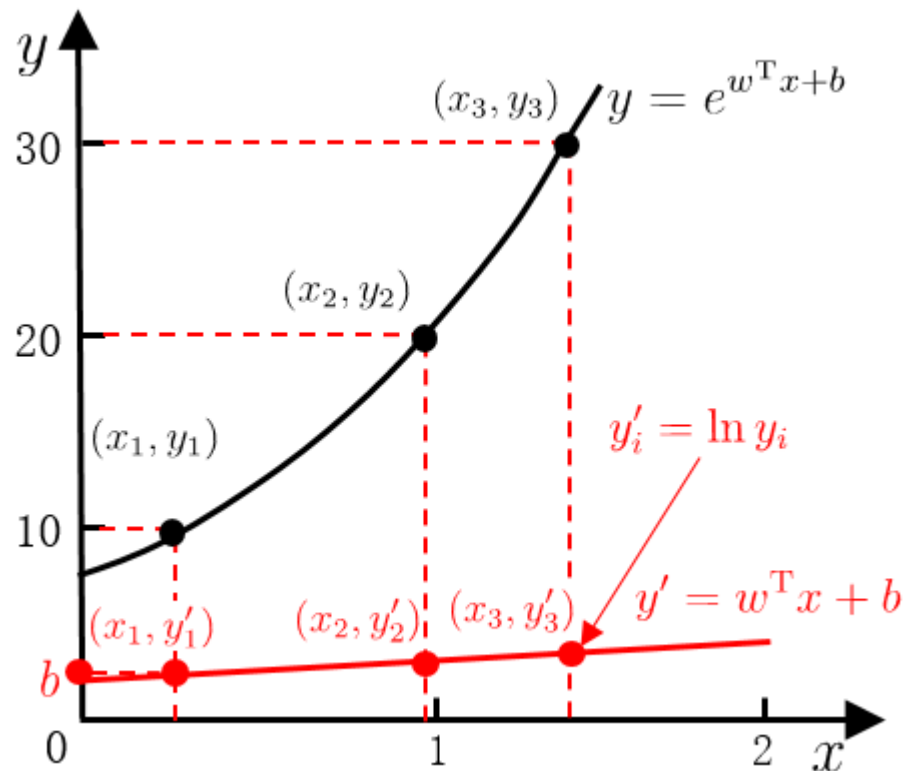
令预测值逼近  $y$  的衍生物 ?

若令  $\ln y = w^T x + b$

则得到对数线性回归

(log-linear regression)

实际是在用  $e^{w^T x + b}$  逼近  $y$



# 广义(generalized)线性模型

---

一般形式:  $y = g^{-1}(\boldsymbol{w}^T \boldsymbol{x} + b)$



单调可微的 联系函数 (link function)

令  $g(\cdot) = \ln(\cdot)$  则得到 对数线性回归

$$\ln y = \boldsymbol{w}^T \boldsymbol{x} + b$$

...



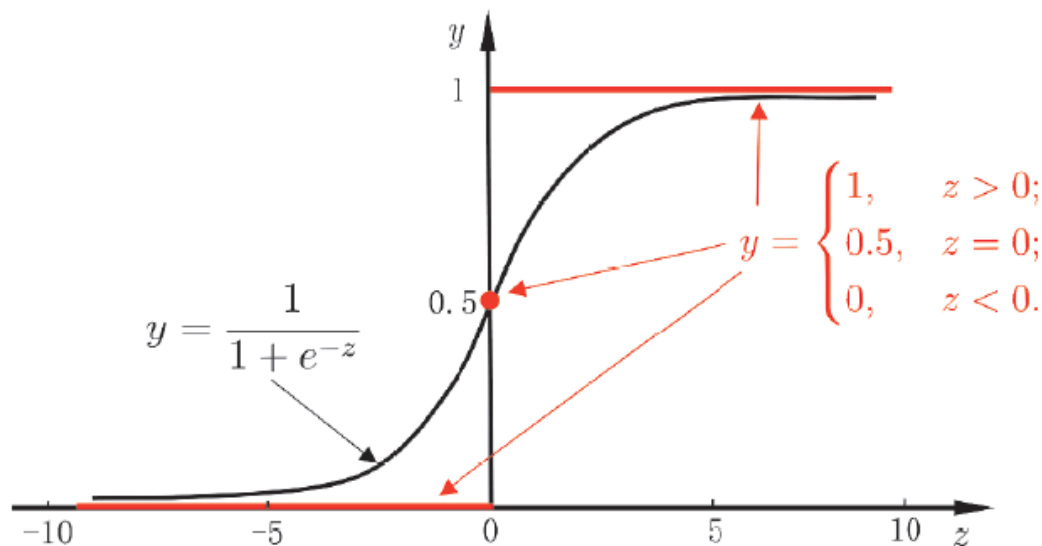
# 二分类任务

线性回归模型产生的实值输出  $z = \mathbf{w}^T \mathbf{x} + b$   
期望输出  $y \in \{0, 1\}$

找  $z$  和  $y$  的  
联系函数

理想的“单位阶跃函数”  
(unit-step function)

$$y = \begin{cases} 0, & z < 0; \\ 0.5, & z = 0; \\ 1, & z > 0, \end{cases}$$



性质不好,  
需找“替代函数”  
(surrogate function)

常用  
单调可微、任意阶可导

$$y = \frac{1}{1 + e^{-z}}$$

对数几率函数  
(logistic function)  
简称“对率函数”

注意: Logistic与“逻辑”没有半毛钱关系!

1. Logistic 源自 Logit, 不是Logic; 2. 实数值, 并非“非0即1”的逻辑值

# 对率回归

以对率函数为联系函数：

$$y = \frac{1}{1 + e^{-z}} \quad \text{变为} \quad y = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}}$$

即：

$$\ln \frac{y}{1 - y} = \mathbf{w}^T \mathbf{x} + b$$

“对数几率”

(log odds, 亦称 logit)

几率(odds), 反映了  $\mathbf{x}$  作为正例的相对可能性

“对数几率回归” (logistic regression)  
简称 “对率回归”

- 无需事先假设数据分布
- 可得到 “类别” 的近似概率预测
- 可直接应用现有数值优化算法求取最优解

注意：它是  
分类学习算法！

# 求解思路

---

若将  $y$  看作类后验概率估计  $p(y = 1 \mid \mathbf{x})$ , 则

$$\ln \frac{y}{1-y} = \mathbf{w}^T \mathbf{x} + b \quad \text{可写为} \quad \ln \frac{p(y = 1 \mid \mathbf{x})}{p(y = 0 \mid \mathbf{x})} = \mathbf{w}^T \mathbf{x} + b$$

于是, 可使用 “极大似然法”  $\longrightarrow$  第7章  
(maximum likelihood method)

给定数据集  $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$

最大化 “对数似然” (log-likelihood) 函数

$$\ell(\mathbf{w}, b) = \sum_{i=1}^m \ln p(y_i \mid \mathbf{x}_i; \mathbf{w}, b)$$

# 求解思路

令  $\boldsymbol{\beta} = (\mathbf{w}; b)$ ,  $\hat{\mathbf{x}} = (\mathbf{x}; 1)$ , 则  $\mathbf{w}^T \mathbf{x} + b$  可简写为  $\boldsymbol{\beta}^T \hat{\mathbf{x}}$

$$\text{再令 } p_1(\hat{\mathbf{x}}_i; \boldsymbol{\beta}) = p(y = 1 \mid \hat{\mathbf{x}}; \boldsymbol{\beta}) = \frac{e^{\mathbf{w}^T \mathbf{x} + b}}{1 + e^{\mathbf{w}^T \mathbf{x} + b}}$$

$$p_0(\hat{\mathbf{x}}_i; \boldsymbol{\beta}) = p(y = 0 \mid \hat{\mathbf{x}}; \boldsymbol{\beta}) = 1 - p_1(\hat{\mathbf{x}}_i; \boldsymbol{\beta}) = \frac{1}{1 + e^{\mathbf{w}^T \mathbf{x} + b}}$$

则似然项可重写为  $p(y_i \mid \mathbf{x}_i; \mathbf{w}_i, b) = y_i p_1(\hat{\mathbf{x}}_i; \boldsymbol{\beta}) + (1 - y_i) p_0(\hat{\mathbf{x}}_i; \boldsymbol{\beta})$

$$\text{于是, 最大化似然函数 } \ell(\mathbf{w}, b) = \sum_{i=1}^m \ln p(y_i \mid \mathbf{x}_i; \mathbf{w}, b)$$

$$\text{等价于最小化 } \ell(\boldsymbol{\beta}) = \sum_{i=1}^m \left( -y_i \boldsymbol{\beta}^T \hat{\mathbf{x}}_i + \ln \left( 1 + e^{\boldsymbol{\beta}^T \hat{\mathbf{x}}_i} \right) \right)$$

高阶可导连续凸函数, 可用经典的数值优化方法  
如梯度下降法/牛顿法 [Boyd and Vandenberghe, 2004]

---

To be continued

---

---

---