北京航空航天大学
COLLEGE OF SOFTWARE BEIHANG UNIVERSITY 软件学院

人工智能

第3讲：机器学习-有监督学习I

线性分类

张晶

2025春季

- 参考资料： 吴飞，《人工智能导论：模型与算法》，高等教育出版社
- 在线课程： https://www.icourse163.org/course/ZJU-1003377027?from=searchPage

提纲

一、机器学习基本概念

二、线性回归与线性分类

三、线性判别分析

四、支持向量机

五、决策树

六、集成学习

提纲

一、机器学习基本概念

二、线性回归与线性分类

三、线性判别分析

四、支持向量机

五、决策树

六、集成学习

机器学习 v.s. 传统编程

输入数据x

程序规则f

传统编程

输出结果y

```
def classification(x):
  if x==[0,0,1,1,0,0]:
    y=0
  elif x==[0,0,1,0,0,0]:
    y=1
  elif x ==[0,1,1,1,0,0]:
    y=2
  return y
```

输入数据x

输出结果y

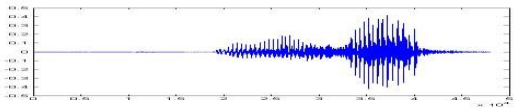
机器学习

程序规则f

```
x=[[0,0,1,1,0,0],
   [0,0,1,0,0,0],
   [0,1,1,1,0,0]]
y=[0,1,2]
def classification(x,y):
  model=linear()
  model.fit(x,y)
  return model
```

机器学习：从数据中学习知识 $f(x) = y$

- 语音识别

$f($  $)=$ “你好”

- 图像分类

$f($  $)=$ “猫”

- 围棋游戏

$f($  $)=$ “5–5”

（下一步落子位置）

监督学习的重要元素

标注数据	标识了类别信息的数据
学习模型	定义映射模型类型
损失函数	对映射结果进行度量
优化方法	对映射模型进行学习

提纲

一、机器学习基本概念

二、线性回归与线性分类

三、线性判别分析

四、支持向量机

五、决策树

六、集成学习

线性回归
Linear Regression

线性回归：一元线性回归

- 训练数据：$\mathcal{D}=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$
- 学习模型： $f(x_i)=wx_i+b\ (1\le i\le n)$
  - 记在当前参数下第i个训练样本x_i的预测值为$f(x_i)$
  - x_i的标注值（实际值）y_i与预测值$f(x_i)$之差记为$(y_i-f(x_i))^2$
- 损失函数： $L(w,b)=\sum_{i=1}^{n} loss(x_i,y_i)=\sum_{i=1}^{n}(y_i-(wx_i+b\ ))^2$
- 优化目标：寻找一组w和b，使得损失函数定义的误差总和$L(w,b)$值最小。
- 优化方法：最小二乘法
  - 对$L(w,b)$ 参数w和b分别求偏导，令其导数值为零，再求取参数w和b的取值。

线性回归：多元线性回归

- 训练数据：$\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_n, y_n)\}$
  - 其中，$\boldsymbol{x_i} = [x_{i,1}, x_{i,2}, \ldots, x_{i,D}] \in \mathbb{R}^D$为代表多元特征的第$i$个数据向量
  - D为数据向量的维度，也就是特征个数
- 学习模型：$f(\boldsymbol{x_i}) = \sum_{j=1}^{D} w_j\, x_{i,j} + w_0 = \boldsymbol{w^T x_i} + w_0, (1 \le i \le n)$
- 损失函数：$L(\boldsymbol{w}) = \sum_{i=1}^{n} loss(\boldsymbol{x}_i, y_i) = \sum_{i=1}^{n} (y_i - (\boldsymbol{w^T x_i} + w_0))^2$
- 优化方法：最小二乘法

线性分类
Linear Classification

线性分类：用线性回归做分类？

- 线性回归

$x \longrightarrow$ 模型 $\longrightarrow \hat{y} \longleftrightarrow y$

- 用线性回归做分类？

$x \longrightarrow$ 模型 $\longrightarrow \hat{y} \longleftrightarrow y$ 类别

0 = 类别 0

1 = 类别 1

线性分类模型

- 二分类问题定义：
  - 数据：假设训练集$\mathrm{D}=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\dots,(\boldsymbol{x}_N,y_N)\}$，样本$\boldsymbol{x}_i\in\mathbb{R}^d$的类别标签为$y_i$。其中，$y_i$的取值范围是$\{0,1\}$，即一共有2类样本。
  - 模型：线性分类器

$$\hat{y}(\boldsymbol{x})=g\big(f(\boldsymbol{x})\big)=g(\boldsymbol{w}^T\boldsymbol{x}+b)$$

  - 目标：正确分类问题

线性分类模型

二分类问题举例

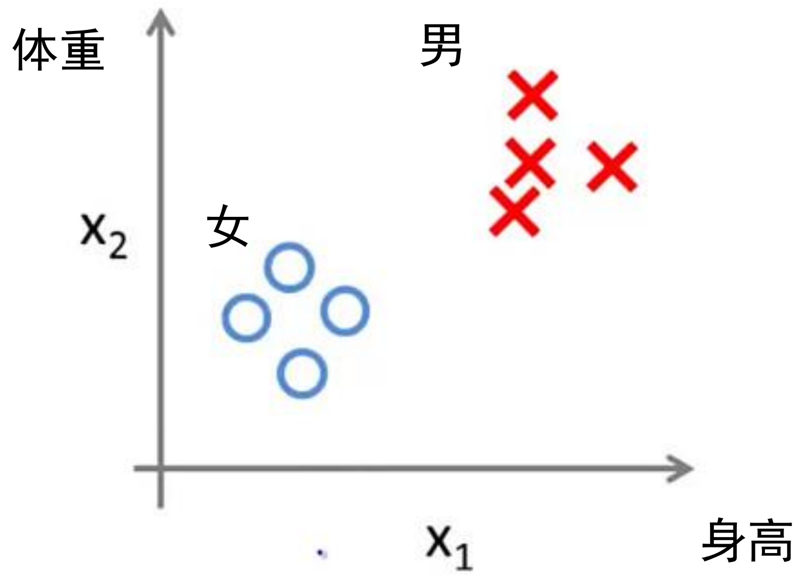
- 下表给出了一组同学的身高和体重，以及这些同学的性别。请根据身高和体重这两个特征，建立线性分类模型，预测性别。

身高 cm	160	162	165	170	177	180	185	190	…
体重 kg	55	52	57	50	70	75	80	75	…
性别	女	女	女	女	男	男	男	男	…

线性分类模型

二分类问题举例

- 下表给出了一组同学的身高和体重，以及这些同学的性别。请根据身高和体重这两个特征，建立线性分类模型，预测性别。



线性分类模型

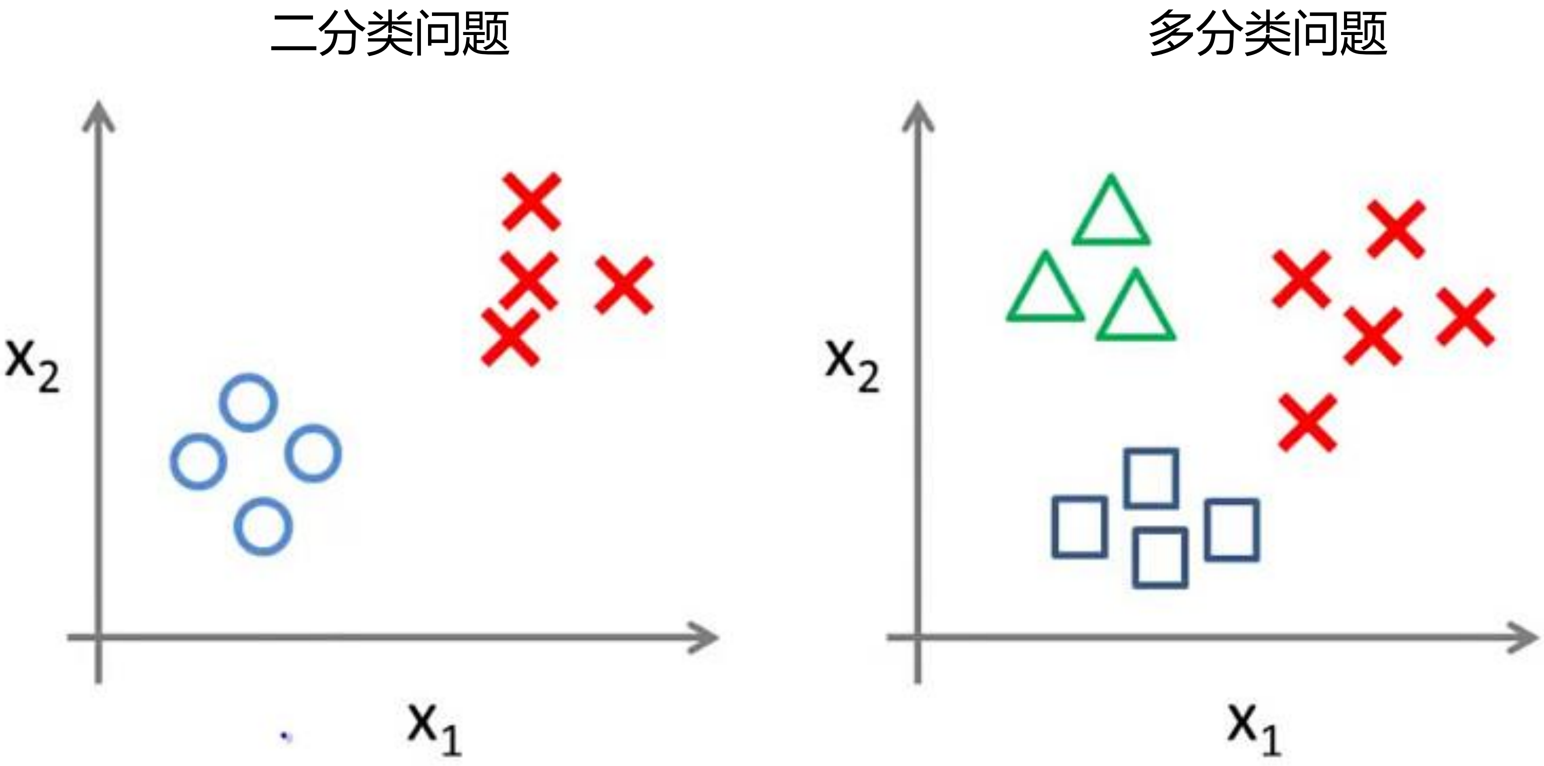
- 多分类问题定义：
  - 数据：假设训练集$\mathrm{D}=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_N,y_N)\}$，样本$\boldsymbol{x}_i\in\mathbb{R}^d$的类别标签为$y_i$。其中，$y_i$的取值范围是$\{0,1,2,\ldots\}$。
  - 模型：线性分类器

$$\hat{y}(\boldsymbol{x})=g\big(f(\boldsymbol{x})\big)=g(\boldsymbol{w}^T\boldsymbol{x}+b)$$

  - 目标：正确分类问题

线性分类模型

- 多分类问题定义：



主观题 10分

设置

直接用线性回归模型做分类任务会有什么问题?

作答

线性分类：用线性回归做分类？

- 线性回归

$$\boldsymbol{x} \longrightarrow \boxed{\text{模型}} \longrightarrow \hat{y} \longleftrightarrow y$$

- 用线性回归做分类?

$$\boldsymbol{x} \longrightarrow \boxed{\text{模型}} \longrightarrow \hat{y} \longleftrightarrow y \quad \text{类别}$$

0 = 类别 0

1 = 类别 1

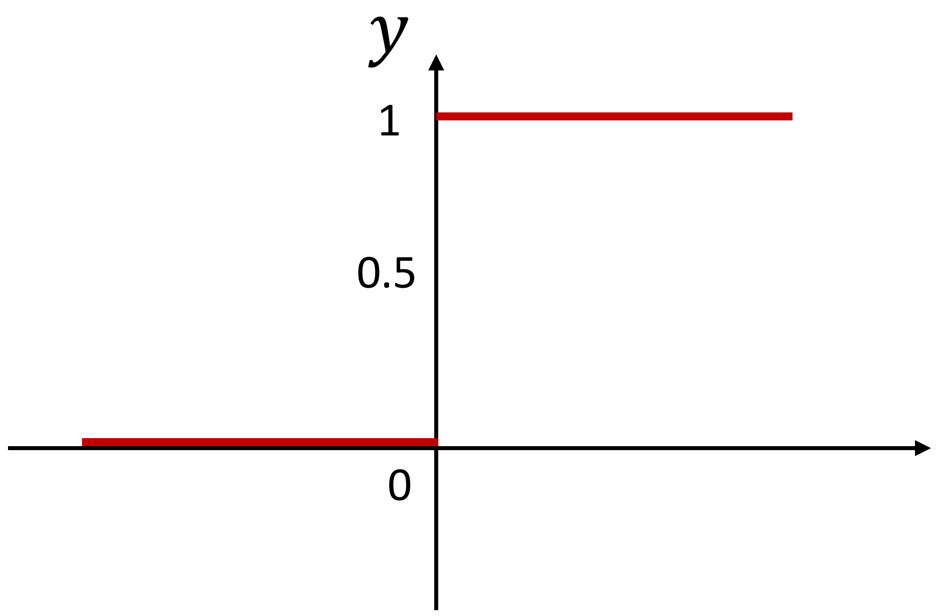
2 = 类别 2

更不同?

更相似?

线性分类：以二分类为例

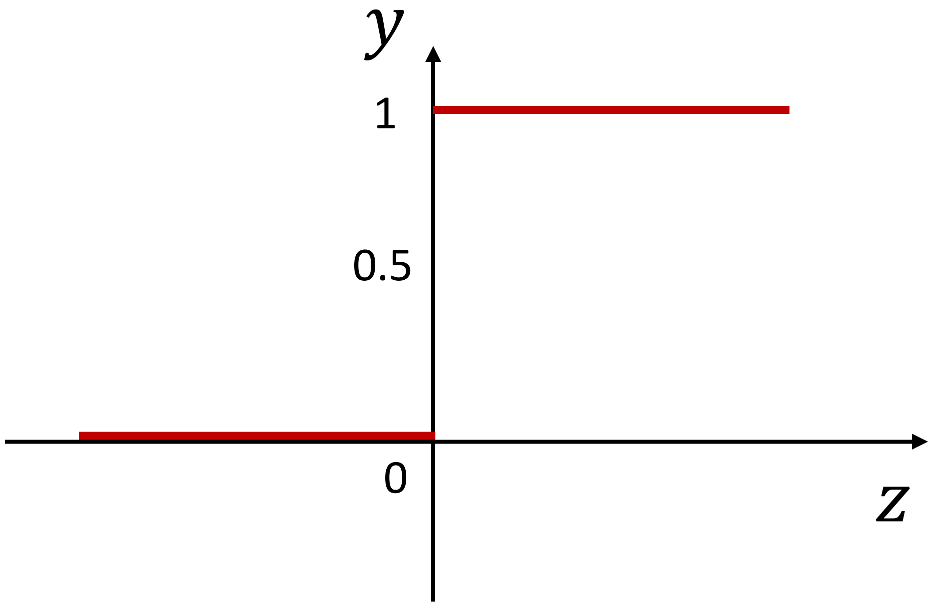
- 线性归回产生的实值输出：$z_i = wx_i + b$
- 线性分类期望的输出： $y_i \in \{0,1\}$
- $y_i = \begin{cases} 1 & z_i > 0 \\ 0.5 & z_i = 0 \\ 0 & z_i < 0 \end{cases}$

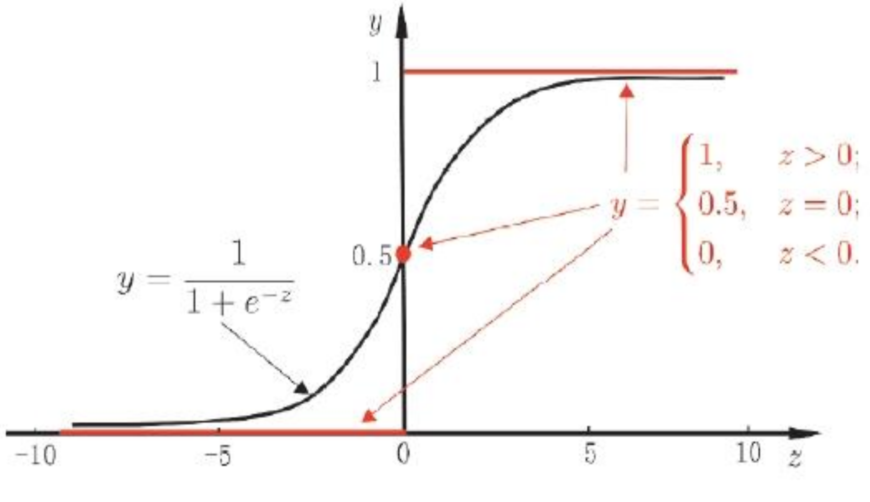


线性分类：以二分类为例

- 训练数据：$\mathcal{D}=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，其中$y_i\in\{0,1\}$
- 学习模型：？
- 损失函数：？
- 优化方法：？

线性分类：对数几率回归 (Logistic Regression)





Sigmoid函数

- 对数几率回归(logistic regression)就是在回归模型中引入 sigmoid函数的一种广义线性模型。

- Logistic回归模型可如下表示：

$$\hat{y} = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}$$ ，其中$\hat{y} \in (0,1), z = \boldsymbol{w}^T\boldsymbol{x} + \mathrm{b}$

- 这里 $\frac{1}{1+e^{-z}}$ 是sigmoid函数、$\boldsymbol{x} \in \mathbb{R}^d$ 是输入数据、$\boldsymbol{w} \in \mathbb{R}^d$ 和 $b \in \mathbb{R}$ 是回归函数的参数。

线性分类模型

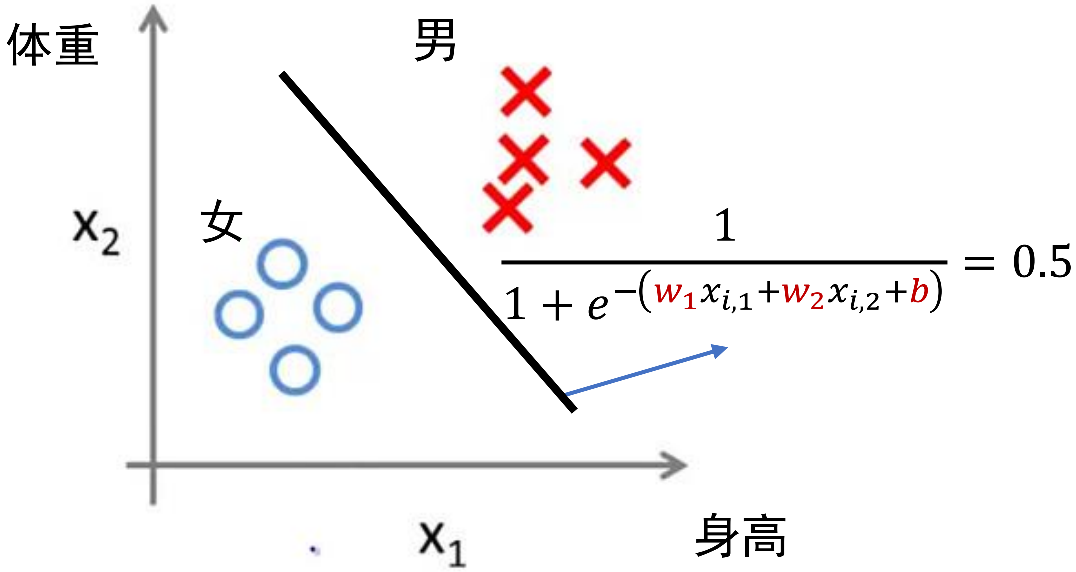
二分类问题举例

- 下表给出了一组同学的身高和体重，以及这些同学的性别。请根据身高和体重这两个特征，建立线性分类模型，预测性别。

身高 cm	160	162	165	170	177	180	185	190	…
体重 kg	55	52	57	50	70	75	80	75	…
性别	女	女	女	女	男	男	男	男	…

线性分类模型

二分类问题举例



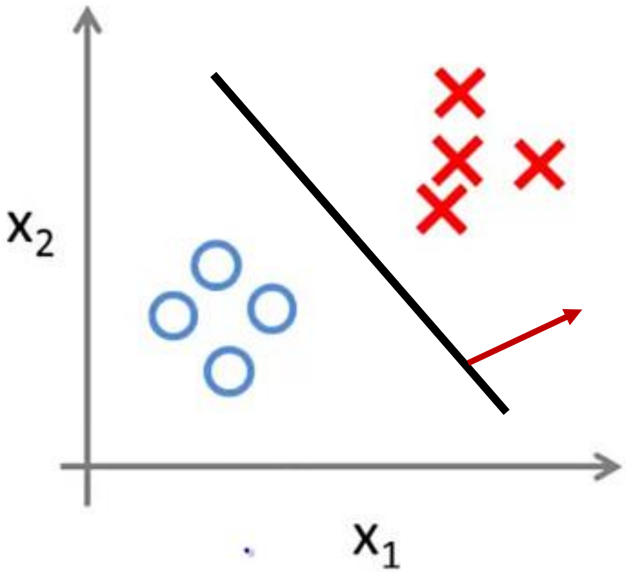
- 令身高为$x_{i,1}$，体重为$x_{i,2}$，性别为y
- 学习模型：$f(\boldsymbol{x}_i)=\frac{1}{1+e^{-(\boldsymbol{w}^T x_i+b)}}=\frac{1}{1+e^{-(w_1 x_{i,1}+w_2 x_{i,2}+b)}}\ (1\le i\le n)$

线性分类模型

- 线性分类模型的一般形式

$$\hat{y}(\boldsymbol{x}) = g(\boldsymbol{w}^T\boldsymbol{x} + b)\text{，其中}g(z) = \begin{cases} 1, & z > 0 \\ 0.5 & z = 0 \\ 0, & z < 0 \end{cases}$$

  - $g(\boldsymbol{w}^T\boldsymbol{x} + b) > 0.5 \Rightarrow \mathcal{C}_1$
  - $g(\boldsymbol{w}^T\boldsymbol{x} + b) < 0.5 \Rightarrow \mathcal{C}_2$
  - g为激活函数



决策边界：
$g(\boldsymbol{w}^T\boldsymbol{x} + b) = 0.5$

线性分类：对数几率回归（Logistic Regression）

- 线性模型：通过特征的线性组合来进行预测的函数

$$\hat{y} = f(x) = \boldsymbol{w}^T\boldsymbol{x} + b$$

- 广义线性模型：

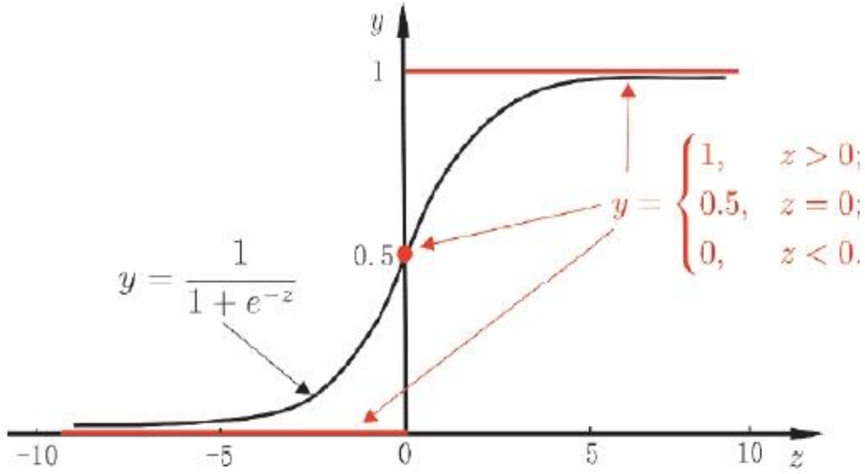
$$\hat{y} = g^{-1}(\boldsymbol{w}^T\boldsymbol{x} + b)$$

- 线性模型用于分类：
  - 找一个单调可微函数将分类任务的真实标记 y 与线性回归模型的预测值联系起来。
  - 比如对数几率回归中：

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \boldsymbol{w}^T\boldsymbol{x} + b$$

即

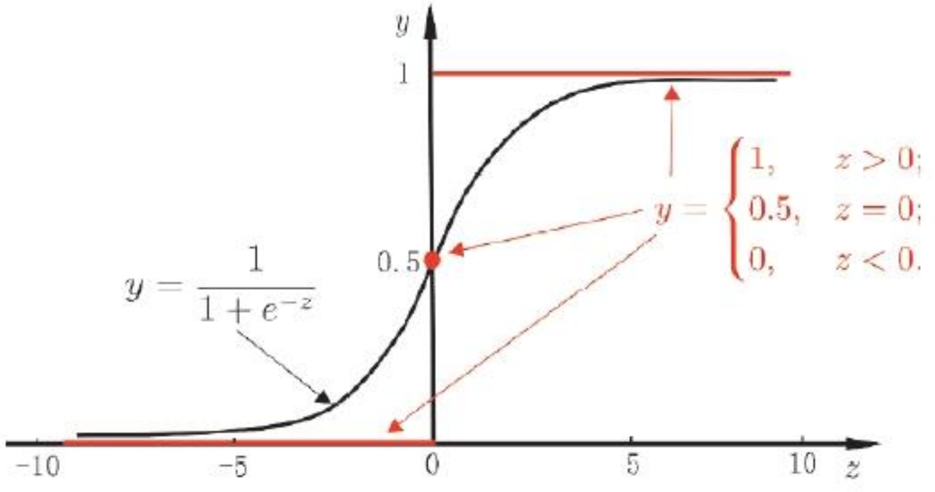
$$\hat{y} = \frac{1}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}$$



线性分类：对数几率回归 (Logistic Regression)

Sigmoid函数$\sigma(z) = \frac{1}{1+e^{-z}}$的特点

- 定义域：$(-\infty, +\infty)$，值域：$(0,1)$。即将取值范围$(-\infty, +\infty)$映射到$(0,1)$之间，更适宜表示预测的概率，即事件发生的“可能性”。
- 在(0,0.5) 处中心对称性：$\sigma(-z) = 1 - \sigma(z)$
- 函数在定义域内为连续光滑非线性函数
- 函数处处可导，导数为：$\sigma'(z) = \sigma(z)(1 - \sigma(z))$
- 反函数为：$z = \ln\left(\frac{\sigma}{1-\sigma}\right)$，可以表示对数几率 (log odds)



Sigmoid函数

线性分类：对数几率回归（Logistic Regression）

- 训练数据：$\mathcal{D}=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，其中$y_i\in\{0,1\}$
- 学习模型： $f(x_i)=\frac{1}{1+e^{-(w^Tx_i+b)}}(1\le i\le n)$
- 损失函数：？
- 优化方法：？

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 假设模型输出$f_\theta(x)$表示输入数据x属于类别1的后验概率$p(y=1|\boldsymbol{x})$

$$f_\theta(x)=p(y=1|\boldsymbol{x})=\frac{1}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}$$

- 那么x属于类别0的后验概率为

$$p(y=0|\boldsymbol{x})=1-p(y=1|\boldsymbol{x})=\frac{e^{-(\boldsymbol{w}^T\boldsymbol{x}+\mathrm{b})}}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+\mathrm{b})}}$$

线性分类：对数几率回归 (Logistic Regression)

几率（Odds）

- 几率定义为事件发生与不发生的概率比：

$$Odds = \frac{p(y=1|\boldsymbol{x})}{p(y=0|\boldsymbol{x})}$$

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 假设模型输出$f_\theta(x)$表示输入数据x属于类别1的后验概率$p(y=1|\boldsymbol{x})$

$$f_\theta(x)=p(y=1|\boldsymbol{x})=\frac{1}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}$$

$$p(y=0|\boldsymbol{x})=\frac{e^{-(\boldsymbol{w}^T\boldsymbol{x}+\mathrm{b})}}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+\mathrm{b})}}$$

- 于是有：

$$\mathrm{logit}\big(p(y=1|\boldsymbol{x})\big)=\ln\left(\underline{\frac{p(y=1|\boldsymbol{x})}{p(y=0|\boldsymbol{x})}}\right)=\underline{\ln\left(\frac{p(y=1|\boldsymbol{x})}{1-p(y=1|\boldsymbol{x})}\right)}=\underline{\boldsymbol{w}^T\boldsymbol{x}+\mathrm{b}}$$

几率(odds)：x属于正例的相对可能性

对数几率(log odds) 或logit

线性回归方程

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

$$\operatorname{logit}\big(p(y=1|\boldsymbol{x})\big)=\log\left(\frac{p(y=1|\boldsymbol{x})}{p(y=0|\boldsymbol{x})}\right)=\log\left(\frac{p}{1-p}\right)=\boldsymbol{w}^T\boldsymbol{x}+b$$

- 如果输入数据$\boldsymbol{x}$属于正例的概率大于其属于负例的概率，即$p(y=1|\boldsymbol{x})>0.5$，则输入数据$\boldsymbol{x}$可被判断属于正例。
- 等价于$\frac{p(y=1|\boldsymbol{x})}{p(y=0|\boldsymbol{x})}>1$，即$\log\left(\frac{p(y=1|\boldsymbol{x})}{p(y=0|\boldsymbol{x})}\right)>\log 1=0$，即$\boldsymbol{w}^T\boldsymbol{x}+\mathrm{b}>0$成立。

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

$$\text{logit}\big(p(y=1|\boldsymbol{x})\big) = \log\left(\frac{p(y=1|\boldsymbol{x})}{p(y=0|\boldsymbol{x})}\right) = \log\left(\frac{p}{1-p}\right) = \boldsymbol{w}^T\boldsymbol{x} + b$$

- $y=1$的对数几率是由输入$\boldsymbol{x}$的线性函数表示的模型$\boldsymbol{w}^T\boldsymbol{x}+b$，因此，logistic 回归是一个线性分类模型。
- 在预测时，可以计算线性函数$\boldsymbol{w}^T\boldsymbol{x}+b$取值是否大于0来判断输入数据$\boldsymbol{x}$的类别归属。

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 为了估计后验概率$f_\theta(x) = p(y = 1|\boldsymbol{x})$，可以用极大似然估计！
- 模型参数的似然函数被定义为$\mathcal{L}(\theta|\mathcal{D}) = \mathrm{p}(\mathcal{D}|\theta)$，其中$\mathcal{D} = \{(x_i, y_i\}(1 \leq \mathrm{i} \leq \mathrm{n})$表示所有观测数据（或训练数据），$\theta$表示模型参数（$\theta = \{\boldsymbol{w}, b\}$）。

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 对于二分类问题，标签y的取值为 0 或 1。可以将条件概率表示为：

$$p(y_i|x_i) = \left(f_\theta(x_i)\right)^{y_i}\left(1 - f_\theta(x_i)\right)^{1-y_i}$$

- 当 $y_i = 1$ 时， $p(y_i = 1|x_i) = f_\theta(x_i)$ 。
- 当 $y_i = 0$ 时， $p(y_i = 0|x_i) = 1 - f_\theta(x_i)$ 。

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 在最大化对数似然函数过程中，一般假设观测所得每一个样本数据是独立同分布 (independent and identically distributed, i.i.d），于是可得：

$$\mathcal{L}(\theta|\mathcal{D}) = \mathrm{p}(\mathcal{D}|\theta) = \prod_{i=1}^{n} \mathrm{p}(y_i|\mathrm{x},\theta) = \prod_{i=1}^{n} \left(f_\theta(x_i)\right)^{y_i} \left(1 - f_\theta(x_i)\right)^{1-y_i}$$

- 对上述公式取对数：

$$L(\theta) = \log\left(\mathcal{L}(\theta|\mathcal{D})\right) = \sum_{i=1}^{n} y_i \log\left(f_\theta(x_i)\right) + (1 - y_i)\log(1 - f_\theta(x_i))$$

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 最大似然估计目的是计算似然函数的最大值，而分类过程是需要损失函数最小化。因此，在上式前加一个负号得到损失函数(交叉熵)：

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

真实值　　预测值

- $L(\theta)$等价于：$L(\theta) = \begin{cases} -\log(f_\theta(x_i)) & if\ y=1 \\ -\log(1-f_\theta(x_i)) & if\ y=0 \end{cases}$

线性分类：对数几率回归（Logistic Regression）

- 训练数据：$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$，其中$y_i \in \{0,1\}$
- 学习模型：$f(x_i) = \frac{1}{1+e^{-(\boldsymbol{w}^T \boldsymbol{x}_i+b)}} (1 \le i \le n)$
- 损失函数：$L(\theta) = -(\sum_{i=1}^{n} y_i \log\left(f_\theta(x_i)\right) + (1 - y_i)\log(1 - f_\theta(x_i)))$
- 优化方法：？

主观题 10分

设置

根据$L(\theta)=\begin{cases}-\log\left(f_\theta(x_i)\right) & if\ y=1\\ -\log\left(1-f_\theta(x_i)\right) & if\ y=0\end{cases}$， $f_\theta(x_i)=\frac{1}{1+e^{-(w^Tx_i+b)}}$

请回答，该损失函数是否满足：

1）损失$L(\theta)$为非负数

2）当预测值$f_\theta(x_i)$与真值y非常接近时，损失接近0

作答

填空题 2分

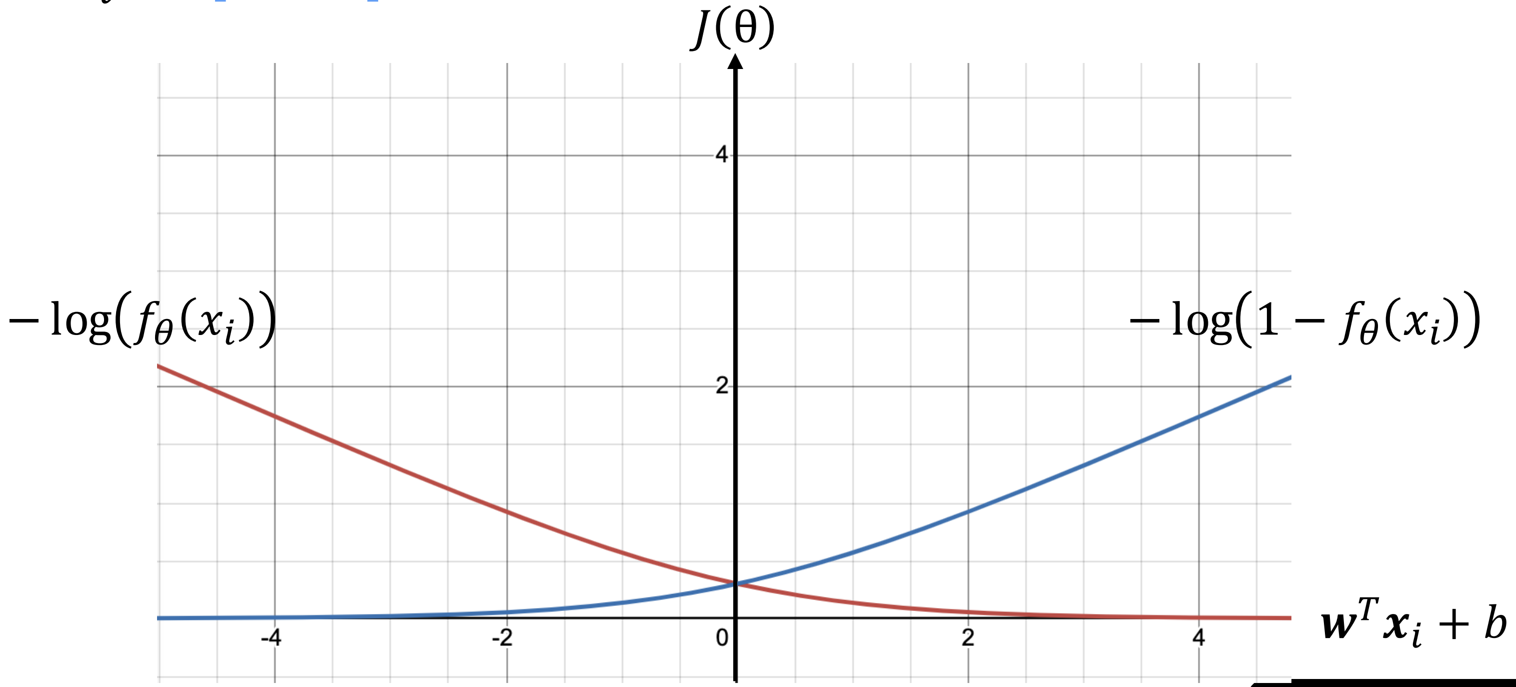
设置

根据 $L(\theta)=\begin{cases}-\log\big(f_\theta(x_i)\big) & if\ y=1\\ -\log\big(1-f_\theta(x_i)\big) & if\ y=0\end{cases}$，其中 $f_\theta(x_i)=\frac{1}{1+e^{-(w^T x_i+b)}}$

以及下图损失函数的示意，请分析，

$y=1$时，希望$\boldsymbol{w}^T\boldsymbol{x}_i+b$[填空1]，

$y=0$时，希望$\boldsymbol{w}^T\boldsymbol{x}_i+b$[填空2]。



作答

线性分类：对数几率回归（Logistic Regression）

- 训练数据：$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$，其中$y_i \in \{0,1\}$
- 学习模型： $f(x_i) = \frac{1}{1+e^{-(w^T x_i + b)}} (1 \le i \le n)$
- 损失函数： $L(\theta) = -(\sum_{i=1}^{n} y_i \log\left(f_\theta(x_i)\right) + (1 - y_i)\log(1 - f_\theta(x_i)))$
- 优化方法：梯度下降
  - 对$L(\theta)$ 参数θ分别求偏导， $\frac{\partial L(\theta)}{\partial \theta_j}$ 。
  - 更新 $\theta_j = \theta_j - \eta \frac{\partial L(\theta)}{\partial \theta_j}$，其中$\eta$为学习率

主观题 10分

设置

- 损失函数：

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

- 优化方法：梯度下降
  - 请推导对$L(\theta)$ 参数θ_j的偏导 $\frac{\partial L(\theta)}{\partial \theta_j}$

作答

线性分类：对数几率回归 (Logistic Regression)

概率输出：从回归到分类

- 需要最小化损失函数来求解参数。
- 损失函数对参数 θ 的偏导如下（其中， $f_\theta'(x) = f_\theta(x)\big(1 - f_\theta(x)\big), \log' x = \frac{1}{x}$ ）

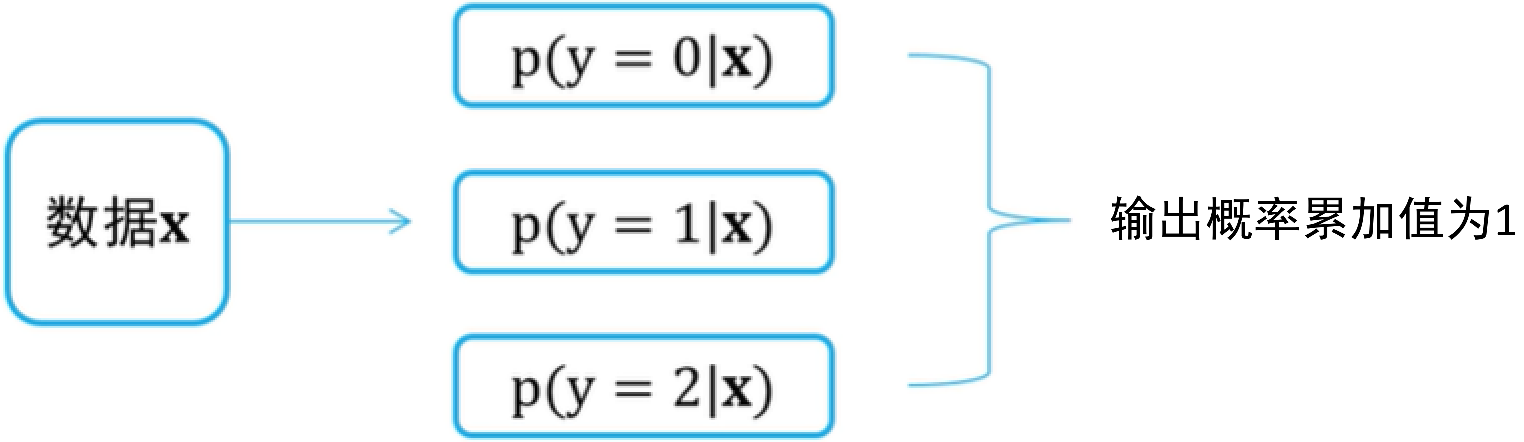
$$
\begin{aligned}
\frac{\partial L(\theta)}{\partial \theta_j} &= -\sum_{i=1}^{n}(y_i \frac{1}{f_\theta(x_i)} \frac{\partial f_\theta(x_i)}{\partial \theta_j} + (1-y_i)\frac{1}{1-f_\theta(x_i)} \frac{\partial(1-f_\theta(x_i))}{\partial \theta_j}) \\
&= -\sum_{i=1}^{n} \frac{1}{f_\theta(x_i)} (\frac{y_i}{f_\theta(x_i)} - \frac{1-y_i}{1-f_\theta(x_i)}) \\
&= -\sum_{i=1}^{n} x_i f_\theta(x_i)\big(1-f_\theta(x_i)\big)\left(\frac{y_i}{f_\theta(x_i)} - \frac{1-y_i}{1-f_\theta(x_i)}\right) \\
&= -\sum_{i=1}^{n} x_i(y_i\big(1-f_\theta(x_i)\big) - (1-y_i)f_\theta(x_i)) = \sum_{i=1}^{n}(y_i - f_\theta(x_i))x_i
\end{aligned}
$$

- 将求导结果代入梯度下降迭代公式得： $\theta_j = \theta_j - \eta \sum_{i=1}^{n}\big(y_i - f_\theta(x_i)\big)x_i$

线性分类：对数几率回归 (Logistic Regression)

从回归到分类：从两类分类到多类分类

- Logistic回归只能用于解决二分类问题
- 将它进行推广为多项逻辑斯蒂回归模型（multi-nominal logistic model，也即softmax函数），用于处理多类分类问题，可以得到处理多类分类问题的softmax回归



三个类别互斥

线性分类：用线性回归做分类？

- 线性回归

$$x \longrightarrow \boxed{\text{模型}} \longrightarrow \hat{y} \longleftrightarrow y$$

- 用线性回归做分类?

$$x \longrightarrow \boxed{\text{模型}} \longrightarrow \hat{y} \longleftrightarrow y$$ 类别

更不同?

0 = 类别 0

1 = 类别 1

2 = 类别 2

更相似?

线性分类：对数几率回归 (Logistic Regression)

从回归到分类（ softmax 分类）：从两类分类到多类分类

- 多类分类问题需要先把真值y表示为独热向量：

类别 1

$$y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

类别2

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

类别3

$$y = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

线性分类：对数几率回归 (Logistic Regression)

从回归到分类（ softmax 分类）：从两类分类到多类分类

类别 1 $\begin{bmatrix}1\\0\\0\end{bmatrix}$　类别2 $\begin{bmatrix}0\\1\\0\end{bmatrix}$　类别3 $\begin{bmatrix}0\\0\\1\end{bmatrix}$

$$z_1 \xrightarrow{4} e^{z_1}=54.6 \longrightarrow y_1 = \frac{e^{z_1}}{\sum_{j=1}^{3} e^{z_j}} = \underline{0.9525}$$

$$z_2 \xrightarrow{1} e^{z_2}=2.7 \longrightarrow y_2 = \frac{e^{z_2}}{\sum_{j=1}^{3} e^{z_j}} = \underline{0.0471}$$

$$z_3 \xrightarrow{-4} e^{z_3}=0.02 \longrightarrow y_3 = \frac{e^{z_3}}{\sum_{j=1}^{3} e^{z_j}} = \underline{0.0004}$$

$$0 < y_i < 1, \sum_i y_i = 1$$

线性分类：对数几率回归（Logistic Regression）

- 训练数据：$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$，其中$y_i$为$K$类标签
- 学习模型： $p(y_{i,k}|x_i) = \frac{exp\left(\boldsymbol{w}^T\boldsymbol{x}_i+b\right)_k}{\sum_j exp\left(\boldsymbol{w}^T\boldsymbol{x}_i+b\right)_j}\ (1 \le i \le n, 1 \le j \le K)$
- 损失函数： $L(\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{K} y_{i,j}\log\left(p(y_{i,j}|x_i)\right)$
- 优化方法：梯度下降
  - 对$L(\theta)$ 参数θ分别求偏导，$\frac{\partial L(\theta)}{\partial \theta_j}$。
  - 更新 $\theta_j = \theta_j - \eta\frac{\partial L(\theta)}{\partial \theta_j}$，其中$\eta$为学习率