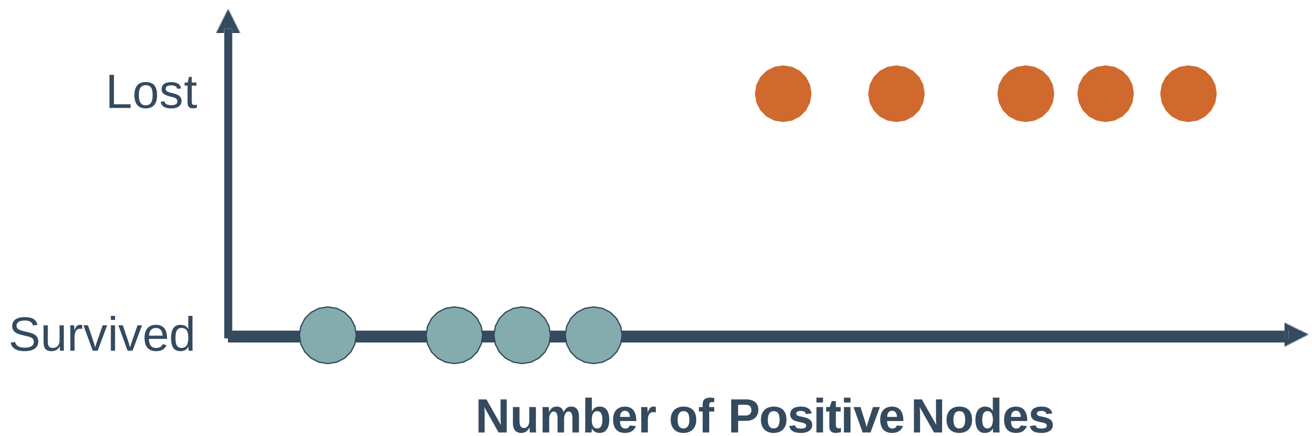
第7章 逻辑回归

线性回归做分类？

癌症病人治疗5年之后的状况



逻辑回归

分类（Classification）：

y=1或y= 0

线性回归（Linear Regression）：

$h_\beta(x)$可以＞1或者＜0

逻辑回归（Logistic Regression）：

$0 \le h_\beta(x) \le 1$

逻辑回归

找一组函数：

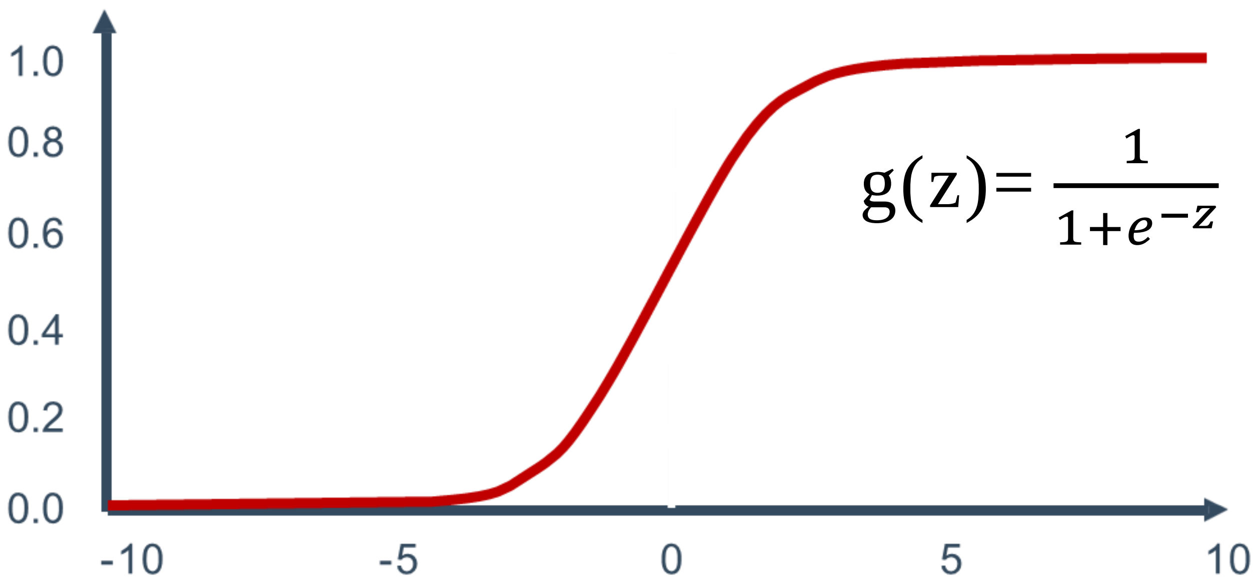
$P(y=1|x) \geq 0.5$，输出y=1

$P(y=1|x) < 0.5$，输出y=0

$$P(y=1|x) = g(z)$$

$$z = \beta_0 + \beta_1 x$$

Logistic函数



一种Sigmod函数

逻辑回归

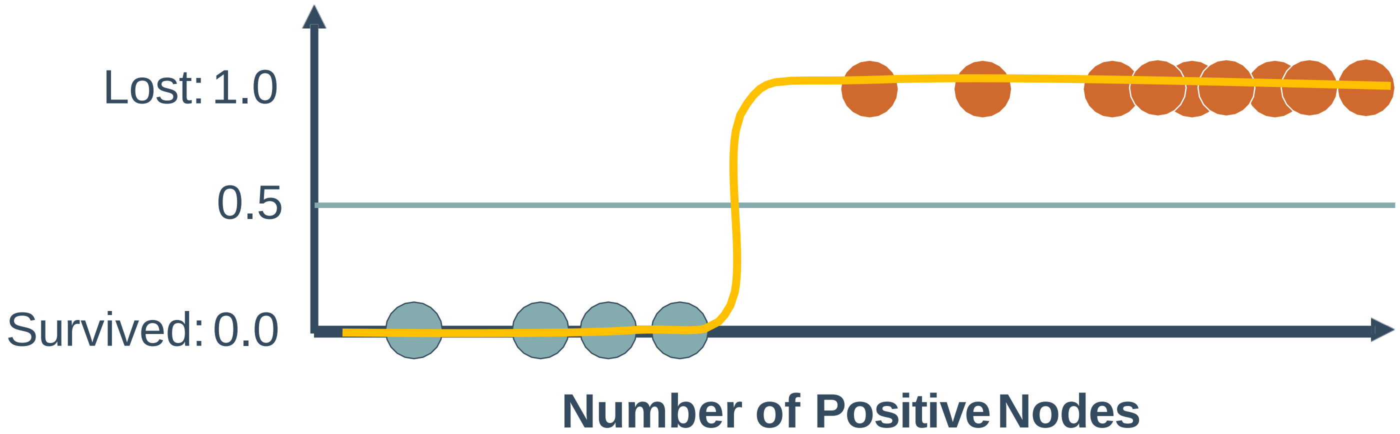
逻辑回归模型：

$$h_\beta(\mathrm{X}) = \mathrm{g}(\beta^{\mathrm{T}} X)$$

$$= \frac{1}{1+e^{-\beta^{\mathrm{T}} X}}$$

逻辑回归

癌症病人治疗5年之后的状况



$$h_\beta(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

逻辑回归和线性回归的关系

Logistic 函数

$$h_\beta(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

$$= \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

逻辑回归和线性回归的关系

Logistic 函数

$$h_\beta(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

Odds Ratio

$$\frac{h_\beta(x)}{1-h_\beta(x)} = e^{\beta_0+\beta_1 x}$$

用以衡量一个特定群体中，属性**A**的出现与否和属性**B**的出现与否的关联性大小

逻辑回归和线性回归的关系

Logistic 函数

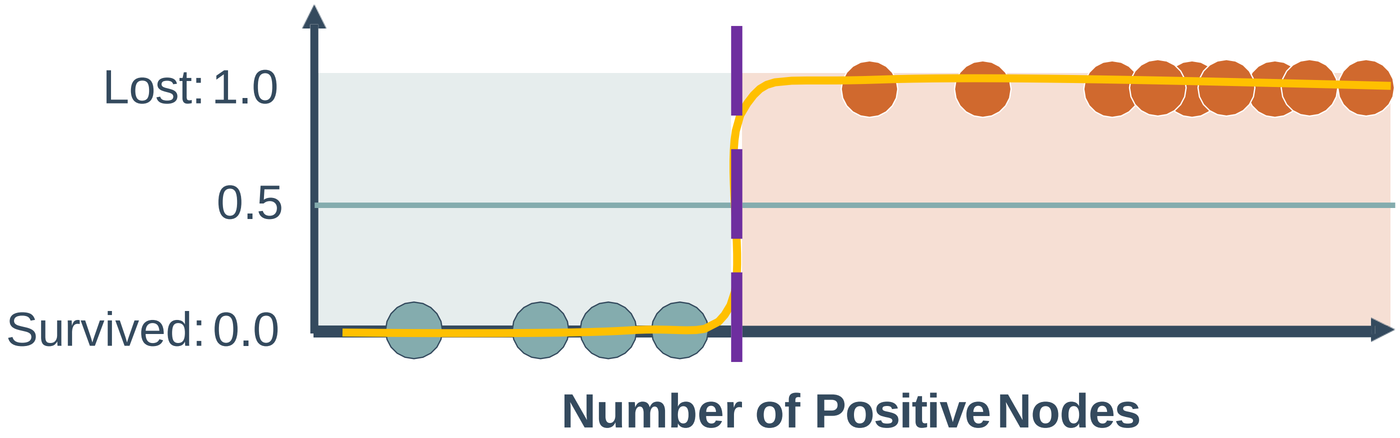
$$h_\beta(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

Log Odds

$$\log\left(\frac{h_\beta(x)}{1-h_\beta(x)}\right) = \boxed{\beta_0 + \beta_1 x}$$

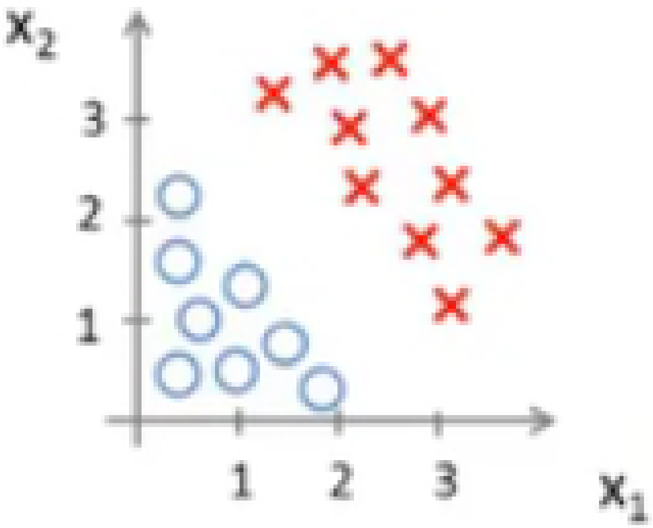
决策边界

癌症病人治疗5年之后的状况

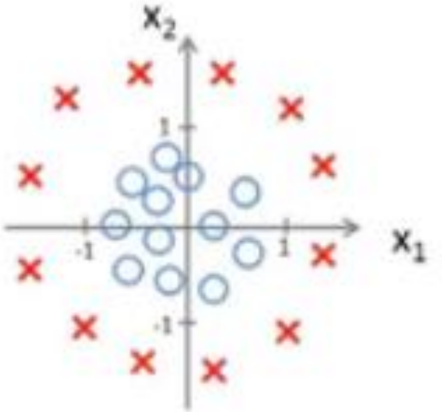


$$h_\beta(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

决策边界



决策边界



代价函数

线性回归：

$$\mathrm{J}(\beta_0,\beta_1)=\frac{1}{2m}\sum_{i=1}^{m}\left(h_\beta\left(x^{(i)}\right)-y^{(i)}\right)^2$$

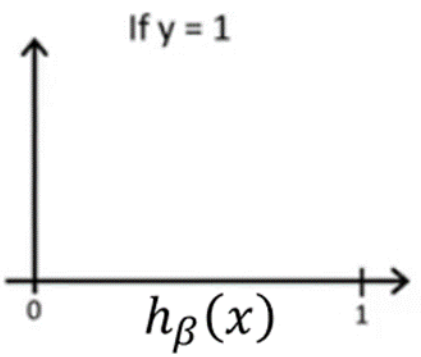
$$=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left(h_\beta\left(x^{(i)}\right)-y^{(i)}\right)^2$$

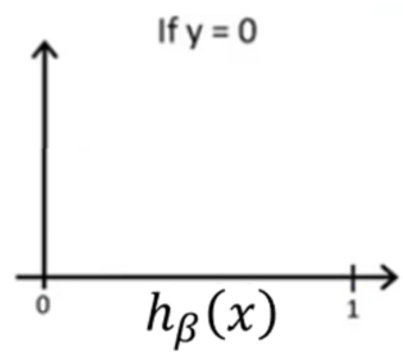
逻辑回归：

$$\mathrm{J}(\beta_0,\beta_1)=\frac{1}{m}\sum_{i=1}^{m}\mathrm{Cost}(h_\beta\left(x^{(i)}\right),y^{(i)})$$

代价函数

$$\mathrm{Cost}\big(h_\beta(x), y\big) = \begin{cases} -\log(h_\beta(x)), & y = 1 \\ -\log(1 - h_\beta(x)), & y = 0 \end{cases}$$





代价函数

$$\mathrm{Cost}\left(h_\beta(x), y\right) = -\mathrm{y} \times \log(h_\beta(x)) - (1 - \mathrm{y}) \times \log(1 - h_\beta(x))$$

$$\mathrm{J}(\beta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \log(h_\beta\left(x^{(i)}\right)) + (1 - y^{(i)}) \times \log(1 - h_\beta\left(x^{(i)}\right))\right]$$

$$\min_{\beta_0,\beta_1} \mathrm{J}(\beta)$$

代价函数

梯度下降法：

$$\mathrm{J}(\beta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \log(h_\beta\left(x^{(i)}\right)) + (1 - y^{(i)}) \times \log(1 - h_\beta\left(x^{(i)}\right))\right]$$

$$\beta_j := \beta_j - \alpha \frac{\partial}{\partial \beta_j}\mathrm{J}(\beta)$$

$$\frac{\partial}{\partial \beta_j}\mathrm{J}(\beta) = ?$$

代价函数

梯度下降法：

$$\mathrm{J}(\beta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \log(h_\beta\left(x^{(i)}\right)) + (1 - y^{(i)}) \times \log(1 - h_\beta\left(x^{(i)}\right))\right]$$

$$\beta_j := \beta_j - \alpha \frac{\partial}{\partial \beta_j}\mathrm{J}(\beta)$$

$$\frac{\partial}{\partial \beta_j}\mathrm{J}(\beta) = \frac{1}{m}\sum_{i=1}^{m}(h_\beta\left(x^{(i)}\right) - y^{(i)})x^{(i)}$$

代价函数

为什么不能用均方误差?

1. 非凸函数

$$\mathrm{J}(\beta_0, \beta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\beta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$h_\beta(x) = \mathrm{g}(z) = \frac{1}{1+e^{-z}}$$

代价函数

为什么不能用均方误差?

2. 梯度下降法求解过程问题

$$J(\beta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\beta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$\beta_j := \beta_j - \alpha \frac{\partial}{\partial \beta_j} J(\beta)$$

$$\frac{\partial}{\partial \beta_j} J(\beta) = ?$$

逻辑回归和线性回归

	逻辑回归	线性回归
Step 1	$h_\beta(x) = \mathrm{g}(\beta_0 + \beta_1 x)$ $0 \le h_\beta(x) \le 1$	$h_\beta(x) = \beta_0 + \beta_1$ $h_\beta(x)$:任意值
Step 2	训练数据：$(x^n, \hat{y}^n)$ $\hat{y}^n$：1或0 $\mathrm{J}(\beta) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{Cost}(h_\beta(x^{(i)}), y^{(i)})$	训练数据：$(x^n, \hat{y}^n)$ $\hat{y}^n$：任意实数 $\mathrm{J}(\beta) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}(h_\beta(x^{(i)}) - y^{(i)})^2$
Step 3	$\beta_i := \beta_i - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\beta(x^{(i)}) - y^{(i)})x^{(i)}$	

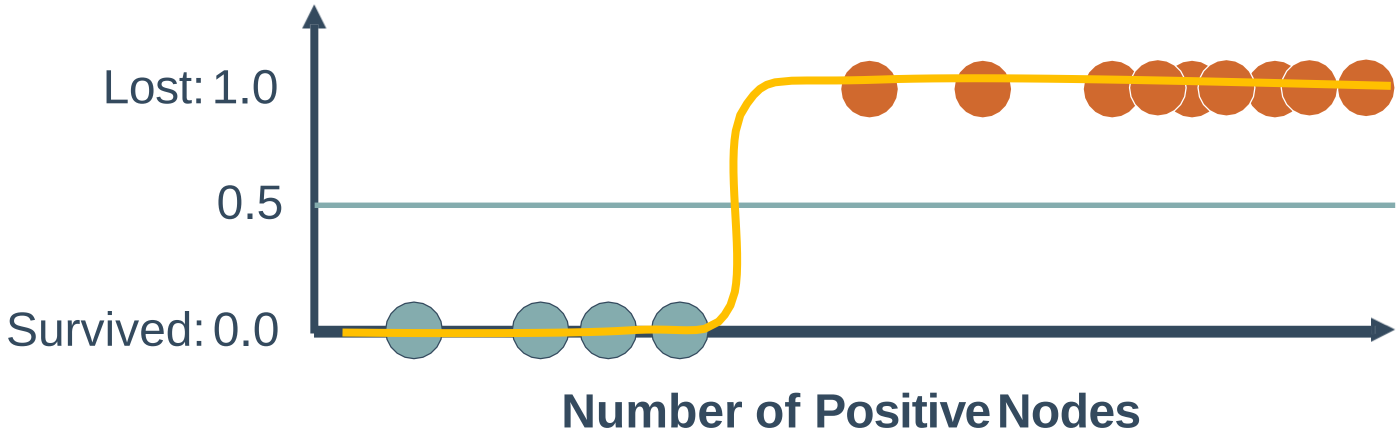
逻辑回归的优点

- 直接对分类可能性建模，无需事先假设数据分布
- 不是仅预测出“类别”，还可得到对近似概率的预测
- 对数几率（logistic）函数是任意阶可导的凸函数，有很好的数学性质

逻辑回归做分类

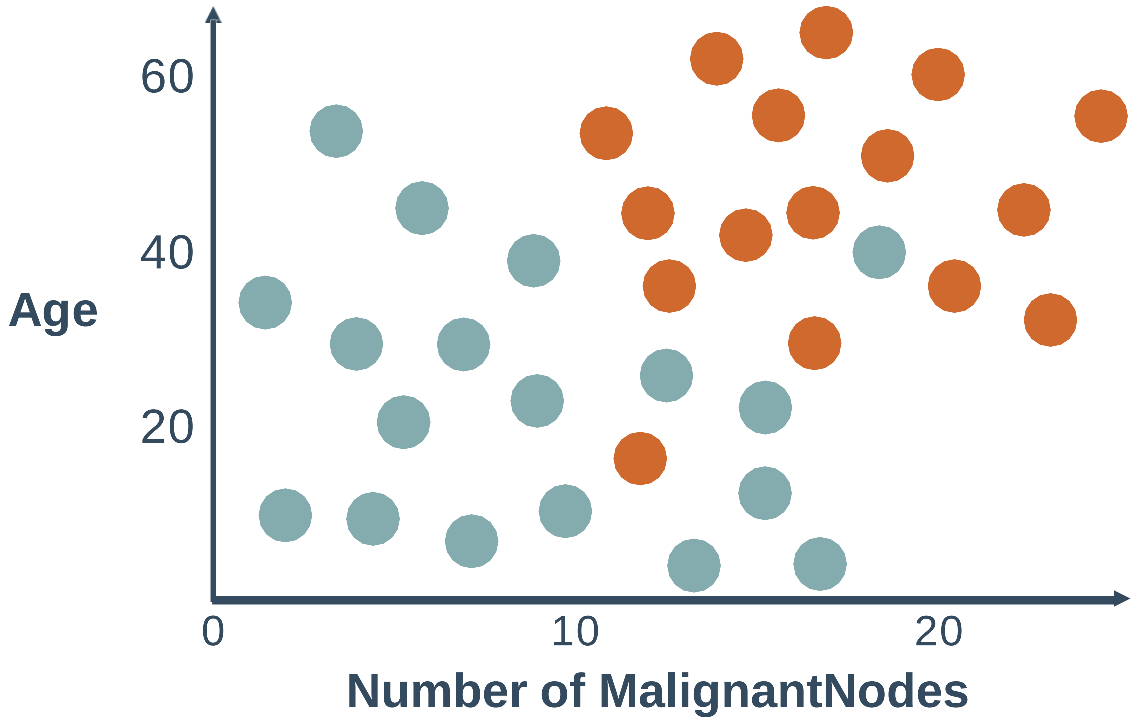
一个特征 (nodes)
两个类标签 (survived, lost)

癌症病人治疗5年之后的状况



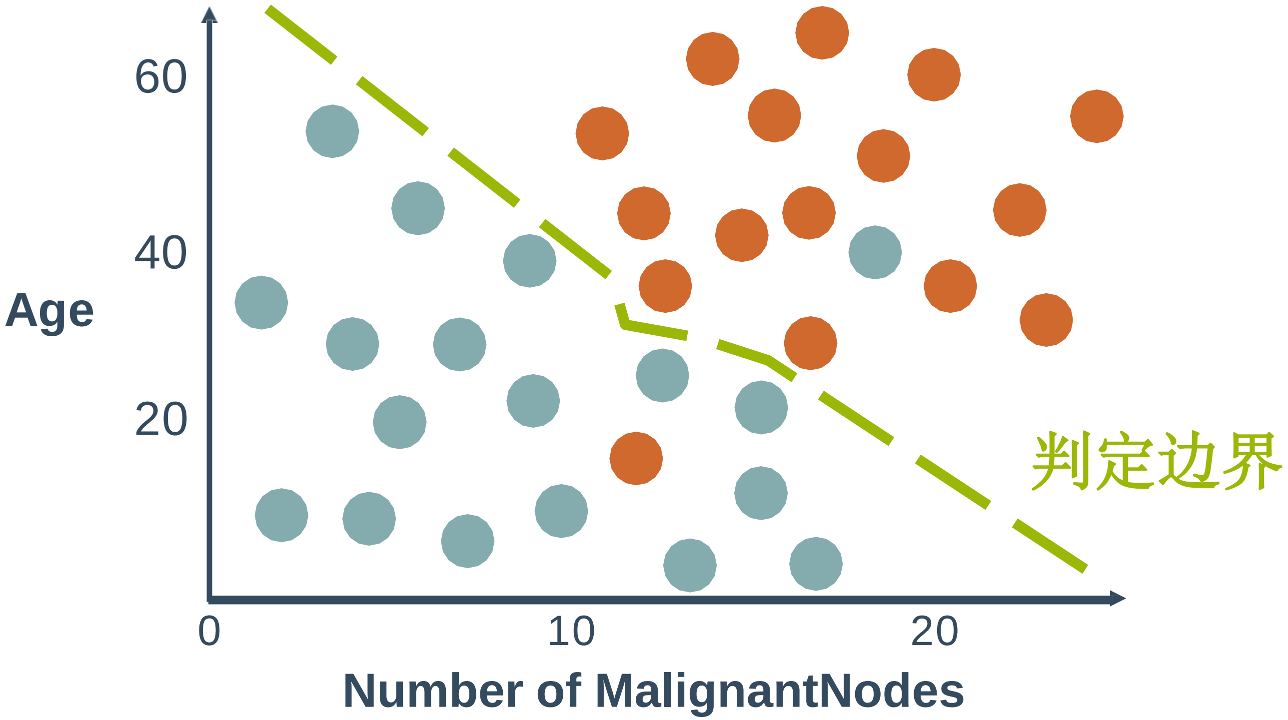
逻辑回归做分类

两个特征 (nodes, age)
两个类标签 (survived, lost)



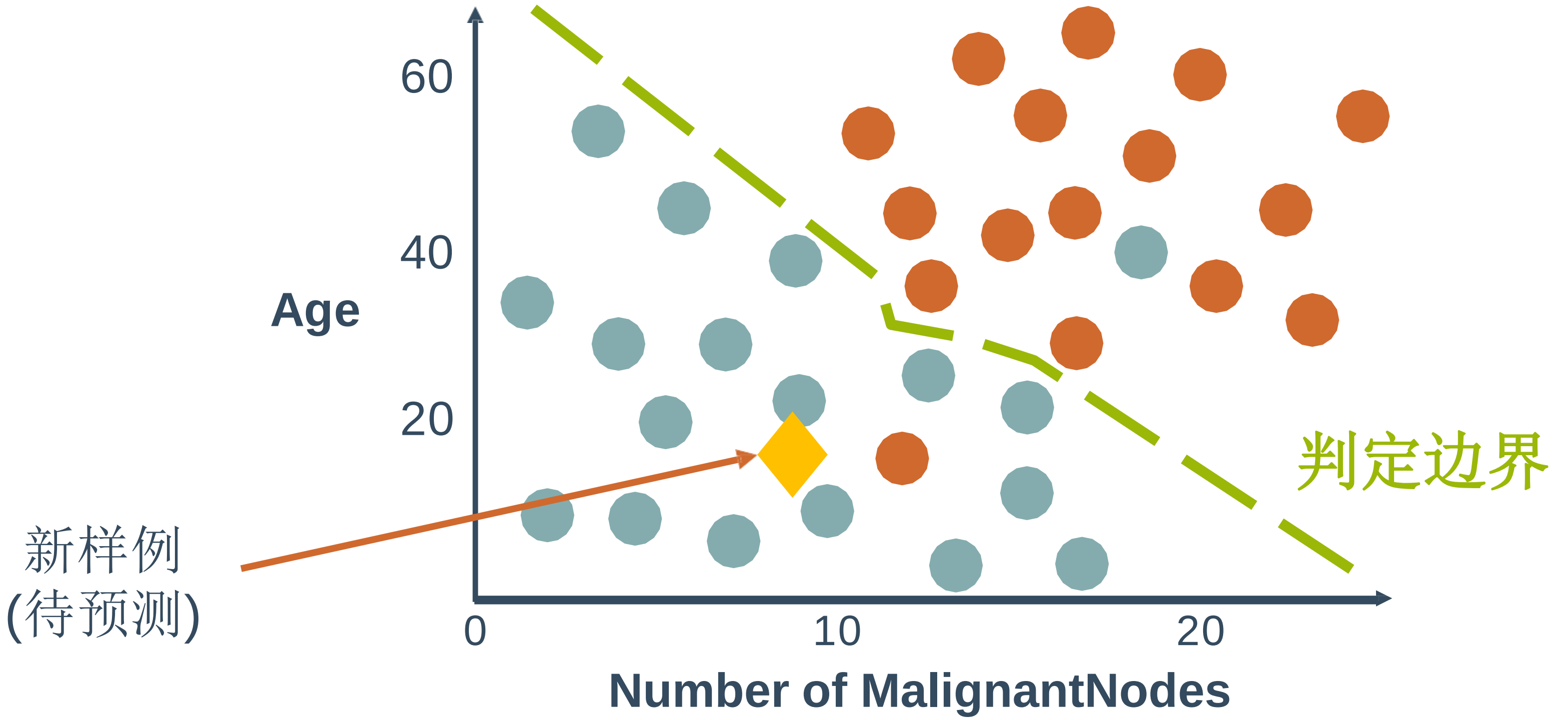
逻辑回归做分类

两个特征 (nodes, age)
两个类标签 (survived, lost)



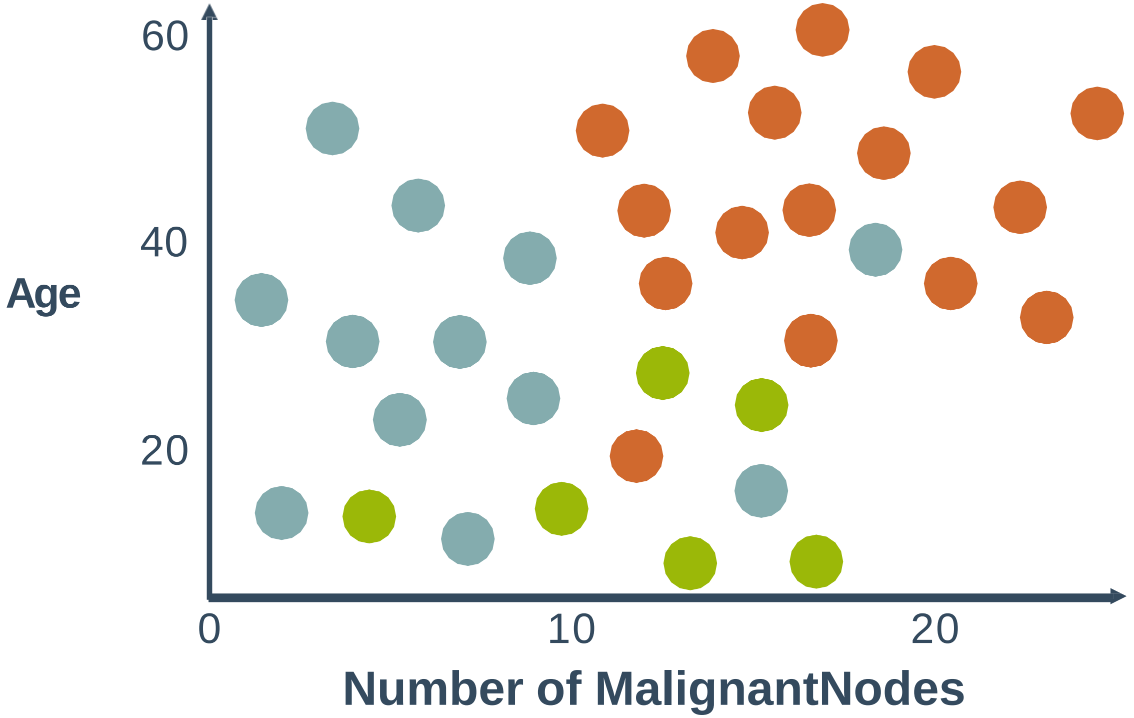
逻辑回归做分类

两个特征 (nodes, age)
两个类标签 (survived, lost)

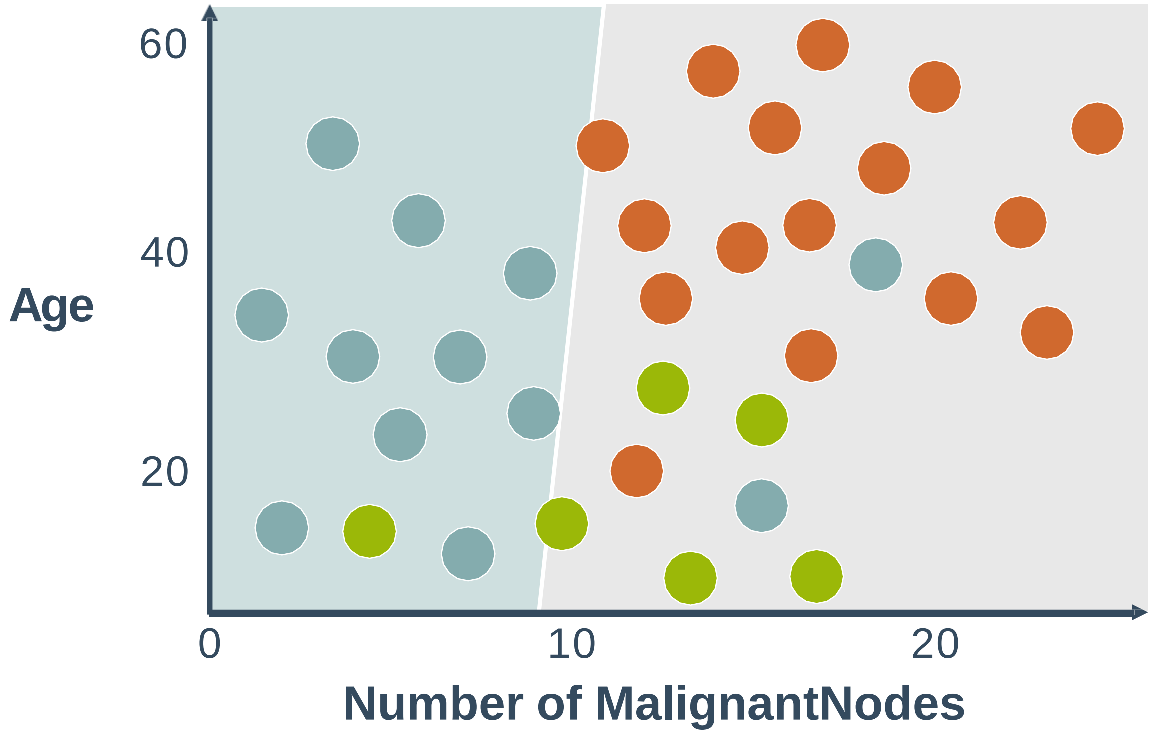


逻辑回归做多分类

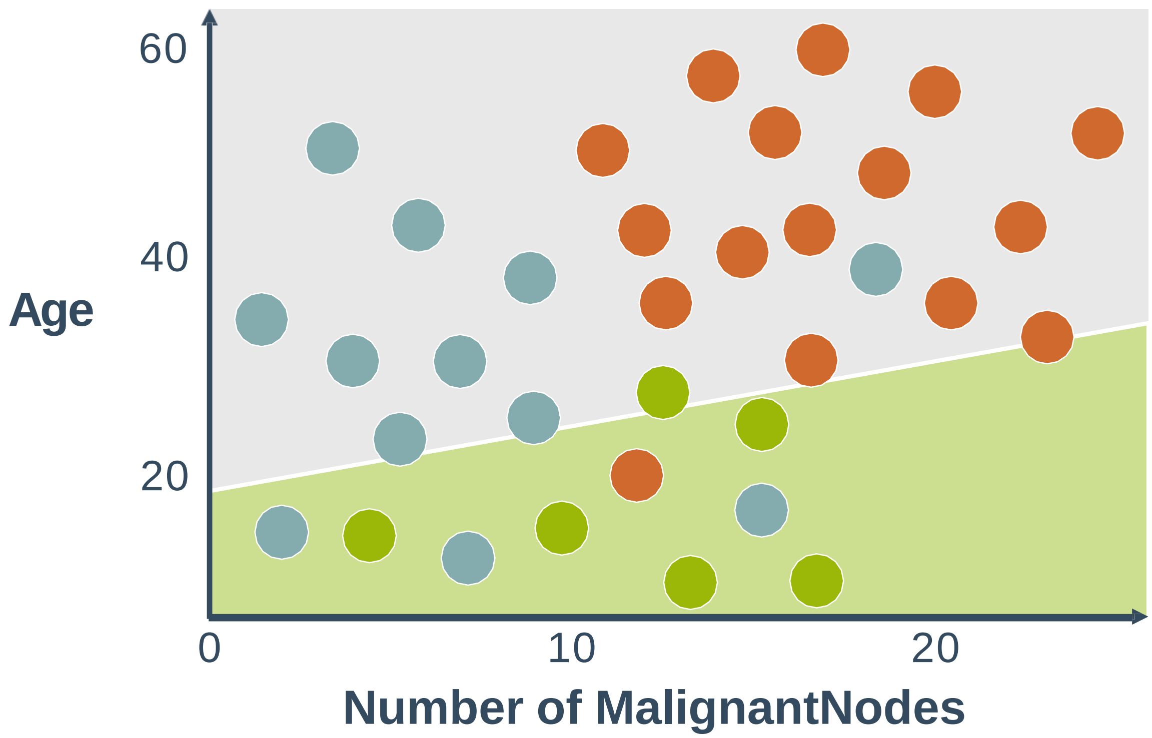
两个特征 (nodes, age)
三个类标签 (survived, complications, lost)



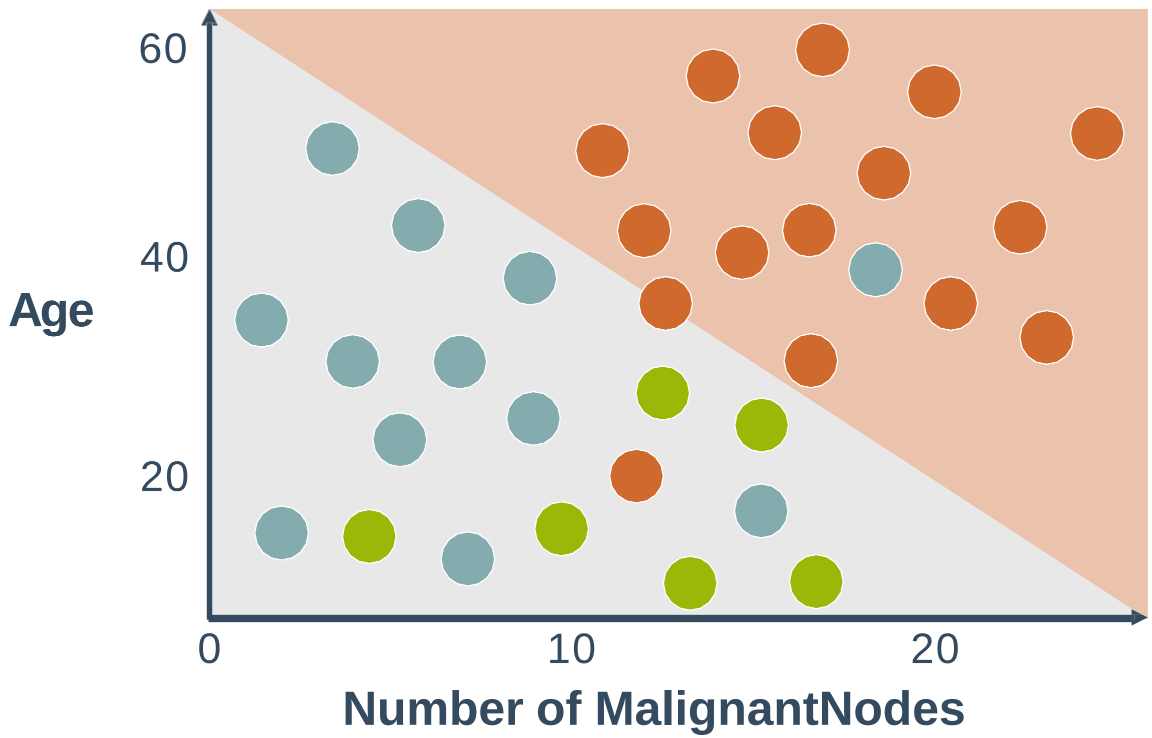
One vs All: Survived vs All



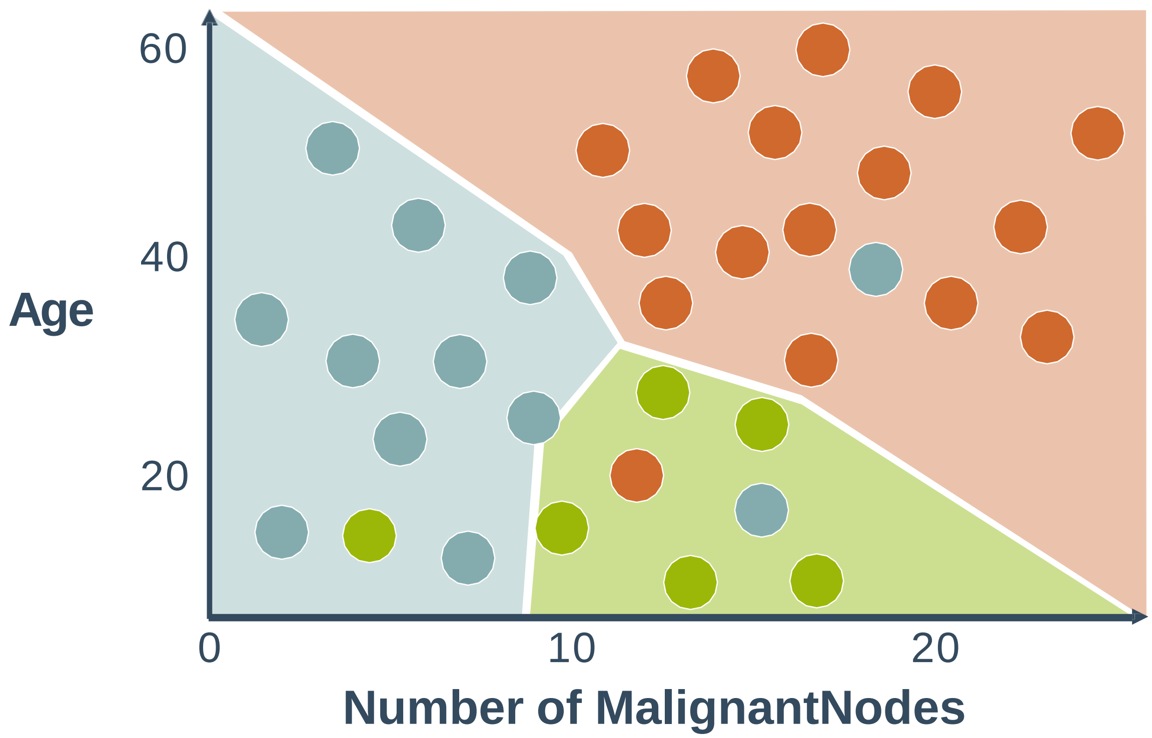
One vs All: Complications vs All



One vs All: Lost vs All



多分类判定边界



每个区域属于其概率最大的类

正则化

$$\mathrm{J}(\beta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \times \log(h_\beta\left(x^{(i)}\right)) + (1 - y^{(i)}) \times \log(1 - h_\beta\left(x^{(i)}\right))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\beta_j^2$$

$$\beta_j \coloneqq \beta_j - \alpha\frac{\partial}{\partial\beta_j}\mathrm{J}(\beta)$$

$$\frac{\partial}{\partial\beta_j}\mathrm{J}(\beta) = \frac{1}{m}\sum_{i=1}^{m}(h_\beta\left(x^{(i)}\right) - y^{(i)})x^{(i)} + \frac{\lambda}{m}\beta_j$$