目录

[生成式分类器 4](#_Toc167367675)

[贝叶斯最优分类器 4](#_Toc167367676)

[最小化错误率决策（最大后验概率） 4](#_Toc167367677)

[最小化风险决策 5](#_Toc167367678)

[引入拒识reject 6](#_Toc167367679)

[小结 6](#_Toc167367680)

[生成式分类器 7](#_Toc167367681)

[概率密度函数的参数估计 7](#_Toc167367682)

[极大似然估计 7](#_Toc167367683)

[贝叶斯估计 8](#_Toc167367684)

[常见分布的参数估计 8](#_Toc167367685)

[伯努利分布 8](#_Toc167367686)

[多项分布 9](#_Toc167367687)

[单变量高斯分布 10](#_Toc167367688)

[多元高斯分布 10](#_Toc167367689)

[朴素贝叶斯分类器 10](#_Toc167367690)

[高斯判别分析 10](#_Toc167367691)

[判别式分类器 11](#_Toc167367692)

[特征工程 12](#_Toc167367693)

[线性回归 13](#_Toc167367694)

[支持向量机 14](#_Toc167367695)

[统计学理论 15](#_Toc167367696)

[集成学习 16](#_Toc167367697)

[聚类 17](#_Toc167367698)

[降维 18](#_Toc167367699)

[半监督学习 19](#_Toc167367700)

[深度学习 20](#_Toc167367701)

[神经元结构 20](#_Toc167367702)

[典型神经元 20](#_Toc167367703)

[人工神经元（M-P模型） 20](#_Toc167367704)

[激活函数 21](#_Toc167367705)

[Relu变种 22](#_Toc167367706)

# 生成式分类器

* 模式分类：给定某个模式样本，确定其所属的类别（通过测量被识别对象的某些特征值，并将其作为一个判决规则的输入，按此规则来对样本进行分类）
* 确定性分类
  + 确定性现象：在获取模式的观测值时，有些事物具有确定的因果关系，即在一定的条件下，它必然会发生或必然不发生。
* 非确定性分类
  + 但在现实世界中，有许多客观现象的发生。就每一次观测来说，即使在基本条件保持不变的情况下也具有不确定性。只有在大量重复的观察下，其结果才能呈现出某种规律性，即对它们观察到的特征具有统计特性。
  + 特征的值不再是一个确定的向量，而是一个随机向量。此时，只能利用模式集合的统计特性来分类，以使分类器发生错误的概率最小

## 贝叶斯最优分类器

### 最小化错误率决策（最大后验概率）

* 希望决策的**平均错误率/平均误差概率**最小
* 如果对于每个样本，保证最小，则平均错误率就最小
* 给定观测值，判断其类别为的错误率是
* 要使错误率最小，则的概率最大。（就是找到样本最match的类别）
* 所以最小错误率决策等价于最大后验概率决策
* 根据贝叶斯规则
* 判别规则：
* 判别函数：
* 对数形式：
* 例题

地震，正常。表示出现生物异常。，。现假设观察到生物异常：

1. 直接求条件概率（后验）
2. 利用后验概率的等价形式

### 最小化风险决策

不同错误决策带来的损失可能不同（如癌症筛查） 引入**损失函数**或**代价函数**

* 损失函数：表示将本应属于类别的模式判别成属于类别的代价
* 平均损失
* 样本对应的条件风险：
* 期望风险：
* 对比平均错误率
  + 条件风险和错误率的作用相同，条件风险是错误率的推广
  + 选择对每个样本条件风险最小的分类规则，将使期望风险最小化
  + 当损失函数为0-1损失时，最小风险等价于最小错误率
* 最小化风险决策

**例题**

|  |  |  |
| --- | --- | --- |
| 课程质量 | 好（Y=1） | 差（Y=0） |
| 概率（先验） | 0.6 | 0.4 |
| P(X|Y) | 课程质量好 | 课程质量差 |
| 课堂有趣（X=1） | 0.8 | 0.1 |
| 课堂无聊（X=0） | 0.2 | 0.9 |

|  |  |  |
| --- | --- | --- |
|  | 课程质量好 | 课程质量差 |
| 选课（y=1） | 0 | 10 |
| 退课（y=0） | 20 | 0 |

* 听了一次课，觉得有趣，求最小化风险决策

退课风险远大于选课风险

### 引入拒识reject

在必要的情况下，分类器对某些样本可以拒绝给出一个输出结果，拒绝将样本判给C个类别中的任何一类

* 损失
* 拒识代价必须小于错分代价，否则永远不会对样本拒识
* 此时的条件风险为
* 引入拒识后的最小风险决策

### 小结

* 贝叶斯最优分类器：最小风险决策所决定出的贝叶斯分类器
* 相应地，风险被称为贝叶斯风险。
* 给定损失函数时，计算贝叶斯最优分类器的关键是计算后验概率

## 生成式分类器

* 计算后验概率，需要已知先验和类条件概率
* 此时亦被称为生成式分类器。因为已知 , 可得到联合分布，从而可以从联合分布通过采样生成数据
* 另一种方式是判别式分类器：直接计算后验概率或者判别函数
* 实际中，估计概率密度函数很困难。尤其是类条件概率， 因为通常是高维随机向量
* 类先验概率
* 类条件概率：由于为多维向量，条件分布建模困难，可根据数据的实际情况对其做适当假设或简化
* 朴素贝叶斯：在给定的情况下，的各维独立
* 高斯判别分析：在给定的情况下，为多元高斯分布

## 概率密度函数的参数估计

给定随机变量或随机向量的概率密度函数的形式，但其参数未知。如，但未知

估计模型参数的方法

* 矩方法
* 极大似然估计：频率学派
* 贝叶斯方法：贝叶斯学派

### 极大似然估计

令为来自分布的独立同分布的样本，定义似然函数：

似然函数在数值上是数据的联合密度，但它是参数的函数，不满足密度函数的性质（对积分不必为1）

极大似然估计是使得似然函数最大的，即

似然函数定义为似然函数的自然对数：

* 自然对数函数为单调增函数，所以和似然函数在相同的位置取极大值
* 数值计算更稳定：似然函数涉及多个小的概率值相乘，容易下溢出
* 计算更简单：很多概率密度函数是指数函数，取对数运算后更简单
* 在不引起混淆的情况下，有时记似然函数为似然函数

### 贝叶斯估计

* MLE认为参数只是一个值（点估计）
* 贝叶斯估计：：参数也是随机变量，亦可用概率分布描述其性质
  + 先验分布：在没有看到数据之前，参数的分布
    - 先验反映我们对参数取值的信念：通常偏好更简单或更光滑的模型
    - 为计算方便，我们一般采用共轭先验（先验分布与后验分布为同族分布）
  + 似然：同MLE相同，为
  + 后验分布： 在看到数据后，对参数分布的更新
    - 参数估计不再是一个点估计，而是一个分布（信息更多）
    - 也可以用后验分布的均值或众数得到参数的点估计

### 常见分布的参数估计

#### 伯努利分布

随机变量，

##### 极大似然估计

似然函数：

*令* ，求得

##### 贝叶斯估计

似然：

该似然对应的共轭先验为Beta分布：，假设

后验：

类先验的贝叶斯后验估计为：

伪先验的和称为先验的强度（先验的有效样本大小），与样本数的作用类似

参数的点估计可以取

* 最大后验估计（后验的众数，MAP）：（当时，退化成MLE）
* 后验的均值：（当时，，称为Laplace平滑）

#### 多项分布

*，*其中为指示函数，当成立时为1，否则为0

##### 极大似然估计

似然函数：

由于需要满足条件，采用拉格朗日乘子法：

分别求导：

得到：

##### 贝叶斯估计

似然：

该似然对应的共轭先验为Dirichlet分布：，假设

后验：

类先验的贝叶斯后验估计为：

参数的点估计可以取

* 最大后验估计（后验的众数，MAP）：（当所有时，MAP退化成MLE）
* 后验的均值：（当所有时，，称为Laplace平滑）

#### 单变量高斯分布

##### 极大似然估计

似然函数：

对参数求偏导：

得到：

##### 贝叶斯估计

（只讨论已知，的贝叶斯估计）

#### 多元高斯分布

##### 极大似然估计

##### 贝叶斯估计

## 朴素贝叶斯分类器（NBC）

假设有个类别，类别的先验分布

* 两类：
* 多类：

每个样本的特征为

朴素贝叶斯分类器：假设各维特征在给定类别标签的情况下条件独立

* 实际应用中即使特征条件独立的假设不严格满足，NBC性能也不错
* 因为NBC比较简单，不容易过拟合

### 模型训练

朴素贝叶斯模型的训练过程就是估计模型的参数

类先验分布：

类条件分布：

似然函数：

* 类先验分布的参数只与似然函数中第2项有关
* 类条件分布只与似然函数中第1项有关

#### 类先验分布

#### 类条件分布

## 高斯判别分析

# 判别式分类器

# 特征工程

# 线性回归

# 支持向量机

# 统计学理论

# 集成学习

# 聚类

# 降维

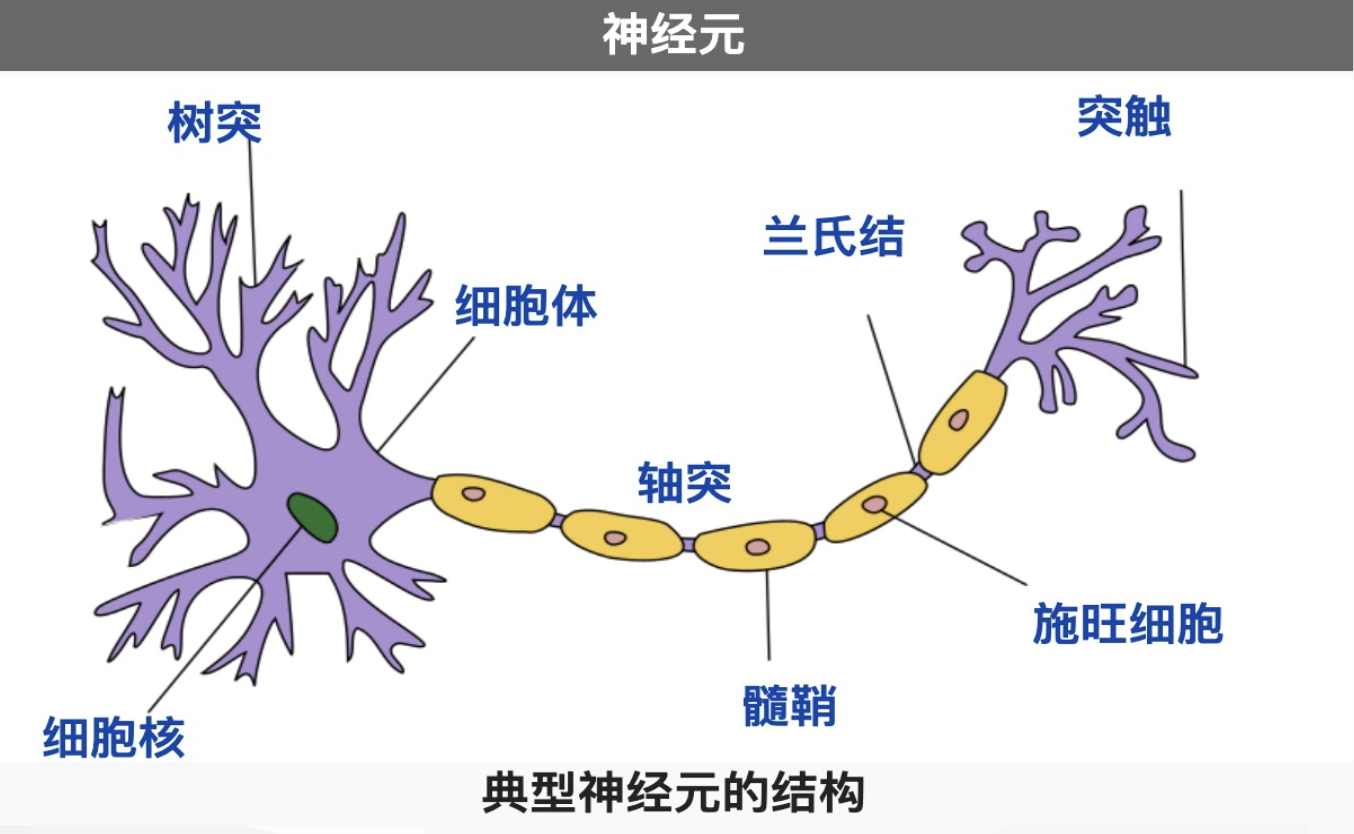
# 半监督学习

# 深度学习

## 神经元结构

### 典型神经元

* 神经细胞之间通过突触连接，突触通过复杂精巧的电化学过程传递信息
* 接收前面神经元的输入，汇总决策传递



### 人工神经元（M-P模型）

* 权重、偏置

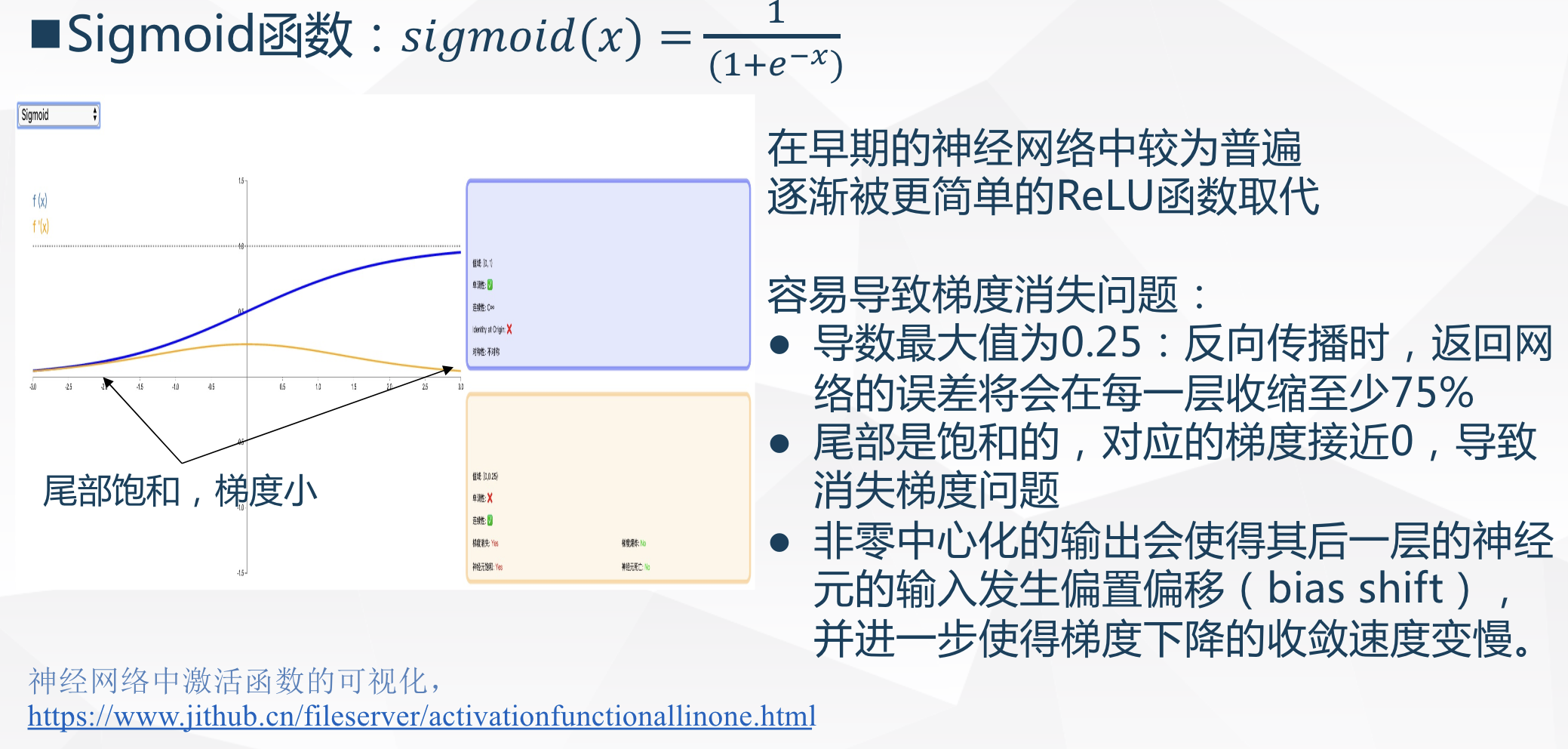
### 激活函数

激活函数性质

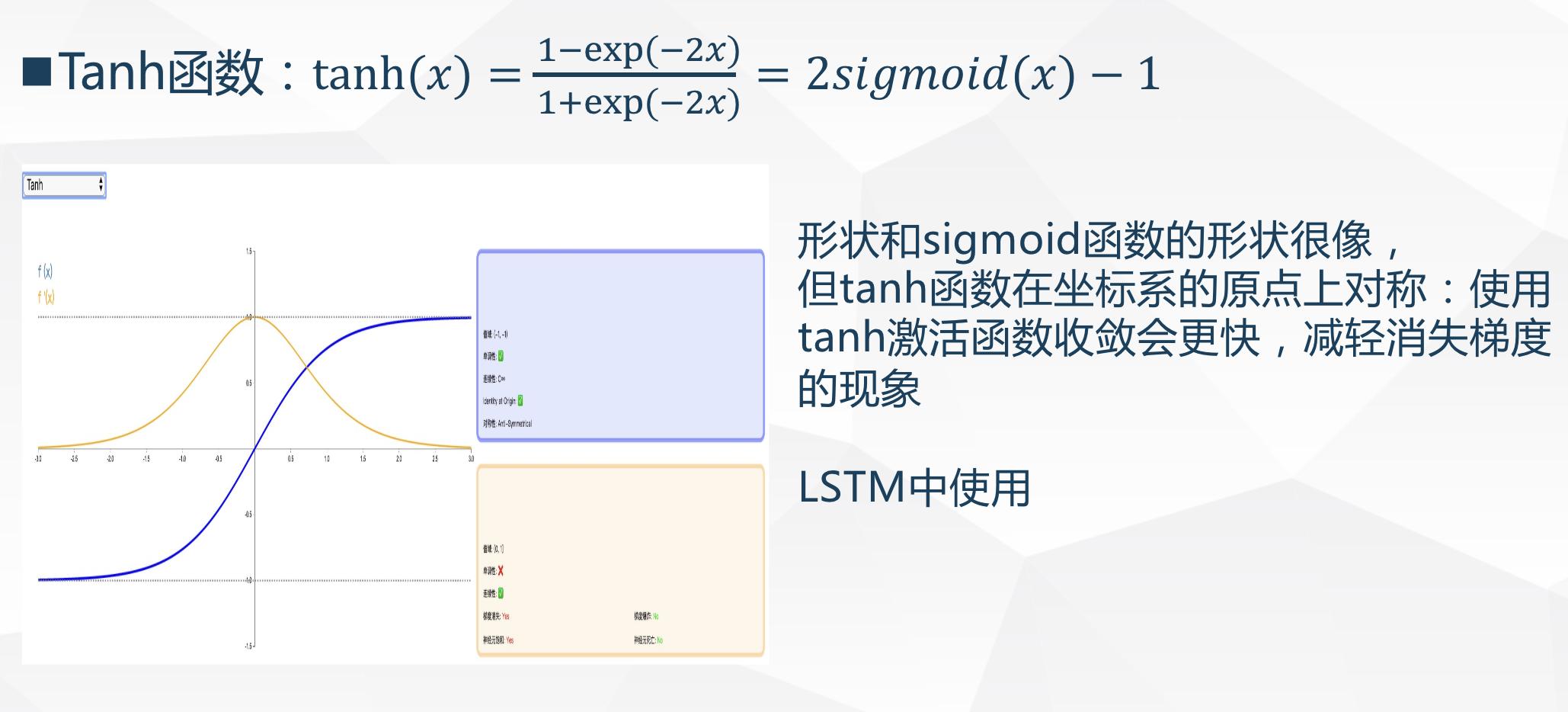
* 连续并可导（允许少数点上不可导）的非线性函数
  + 可导的激活函数可以直接利用数值优化的方法来学习网络参数
* 激活函数及其导函数要尽可能的简单
  + 有利于提高网络计算效率
* 导函数的值域要在一个合适的区间内
  + 不能太大也不能太小，否则会影响训练的效率和稳定性。

**例子**

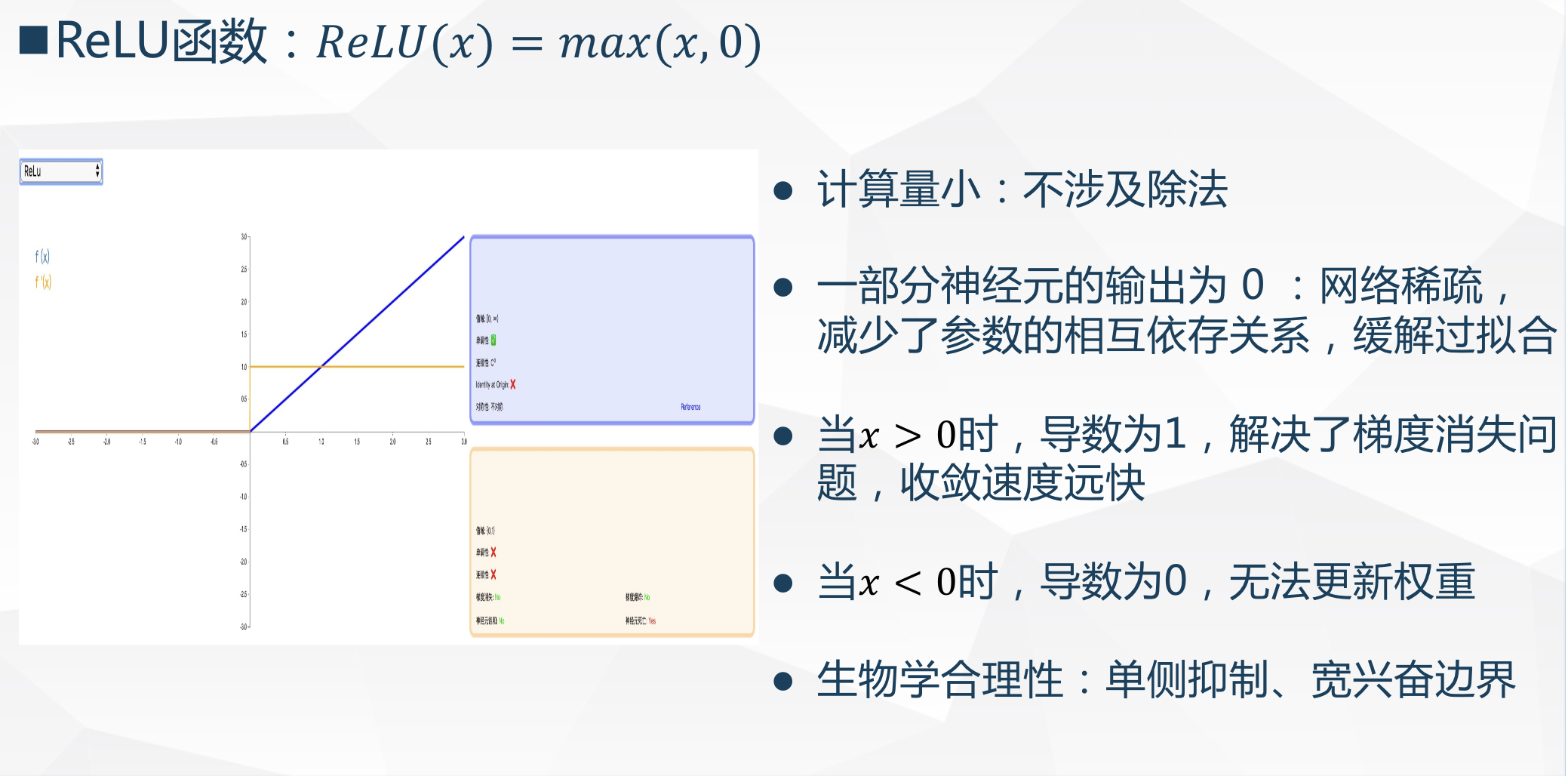
* Sigmoid



* Tanh



* Relu



### Relu变种