

Artificial Intelligence

人工智能实验

机器学习基础

中山大学计算机学院
2024年春季

目录

1. 理论课内容回顾

1.1 人工神经网络介绍

1.2 分类算法

1.3 梯度下降

1.4 数据标准化和归一化

1.5 过拟合和欠拟合

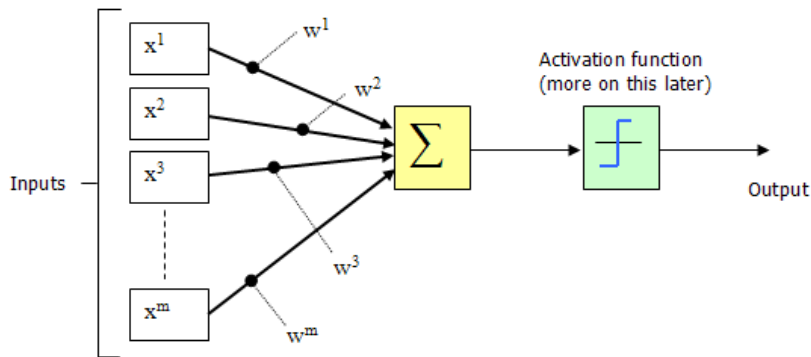
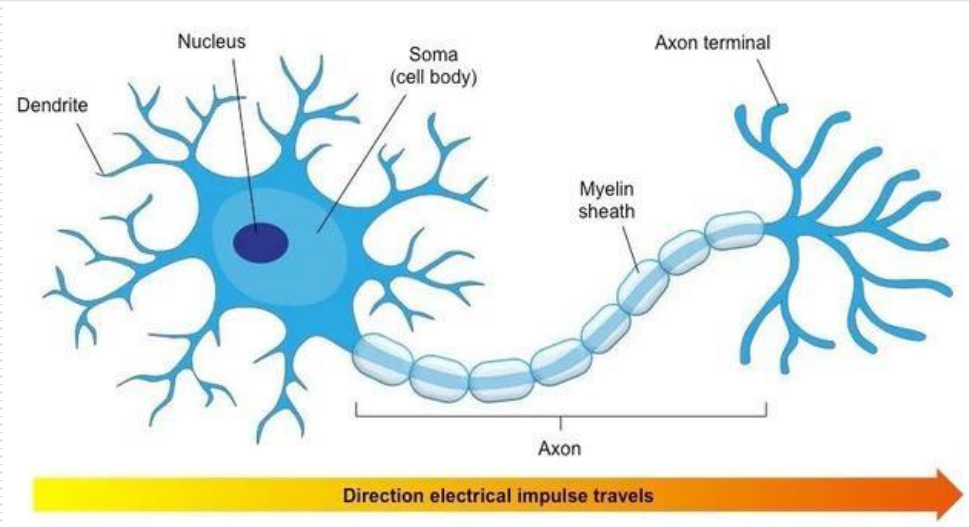
2. 实验任务

2.1 购房预测分类任务

3. 作业提交说明

1.1 人工神经网络介绍

- 神经网络采用了仿生学的思想，通过模拟生物神经网络的结构和功能来实现建模

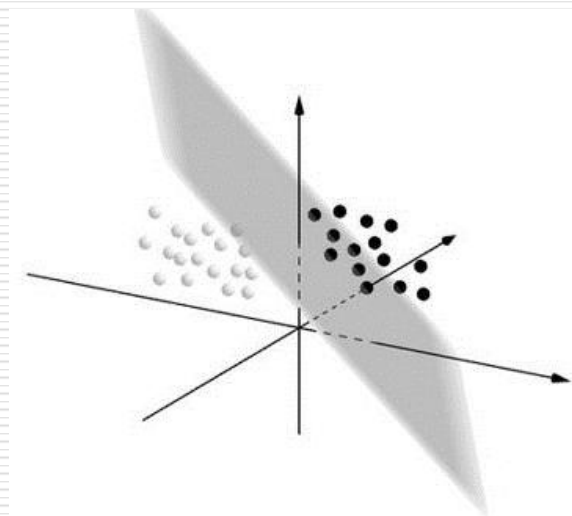
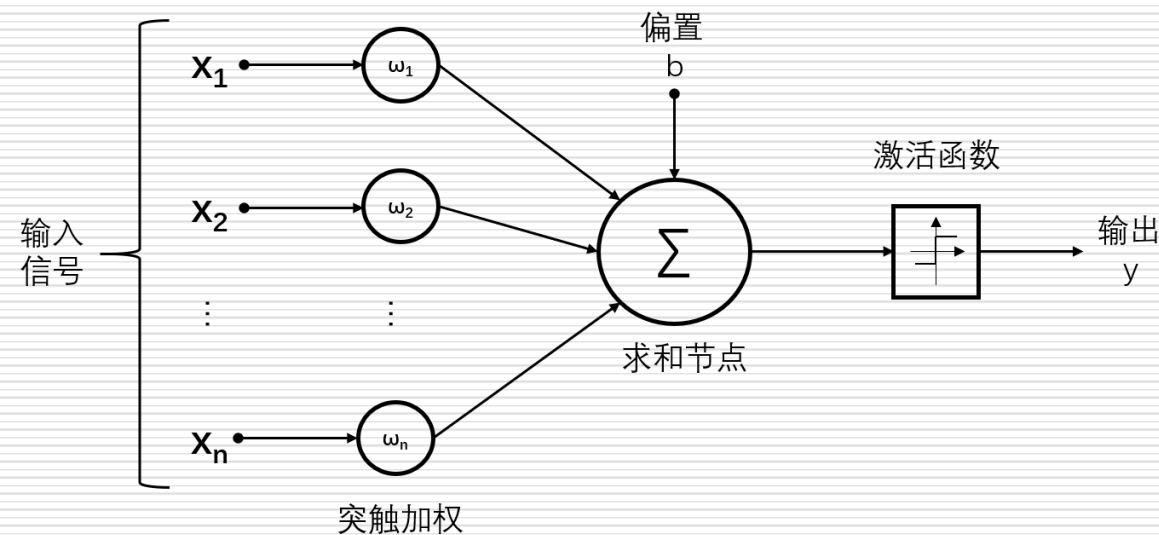


$$y = f(\sum_{i=1}^n w_i x_i - \theta)$$

1.1 人工神经网络介绍

□ 单层感知机

- 由于M-P神经元模型参数需要事先设定好，为了能够自适应学习出所需要的参数，有研究人员就提出了单层感知机(Perceptron)
- 感知机的基本公式为： $y(x)=sign(wx+b)$
- $sign$ 为符号函数，当自变量为正数时取值为1，否则取值为0



1.1 人工神经网络介绍

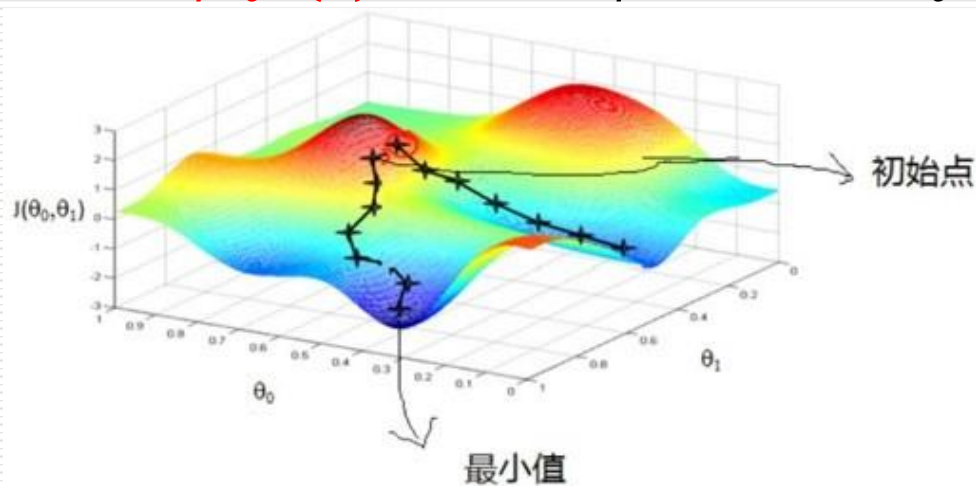
□ 损失函数

- 作用：为了衡量网络表现是否良好，并为之后的网络参数优化提供指导。
- 常见的用在分类任务上的损失函数：
 - 均方误差(MSE): $L_{MSE} = \frac{1}{2n} \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2$
 - 交叉熵: $L_{CE} = - \sum_i^n y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})$

1.1 人工神经网络介绍

□ 梯度下降

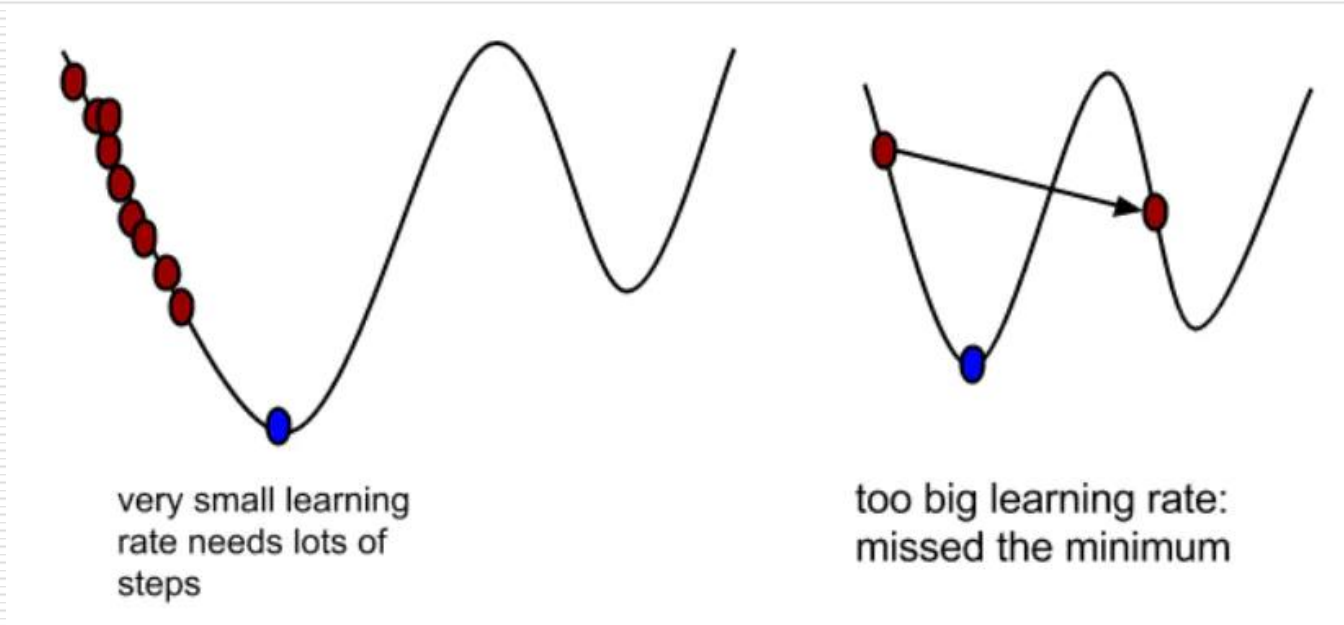
- 梯度定义：梯度是一个向量，表示某一函数在该点出的方向导数沿着该方向取得最大值。
- 也就是说该点处沿着梯度的方向变化最快，变化率最大
 - 沿着梯度方向容易找到函数最大值
 - 沿着梯度方向的反方向，容易找到函数最小值
- 梯度下降的一般公式为： $\theta = \theta - \eta \nabla_{\theta} L(\theta)$ ，其中， η 是学习率， ∇_{θ} 是对 θ 的梯度， θ 是参数



1.1 人工神经网络介绍

□ 学习率

- 学习率限制了下一步能到达的地方
- 如果学习率太小，可能很难达到最小值
- 如果学习率太大，则会错过最小值，无法收敛



1.1 人工神经网络介绍

□ 单变量的梯度下降

- 假设有一个单变量的函数 $J(\theta) = \theta^2$ ，我们找该函数的最小值
- 该函数的微分计算公式为： $J'(\theta) = 2\theta$
- 我们可以初始化起点 $\theta^0 = 1$
- 迭代过程为：

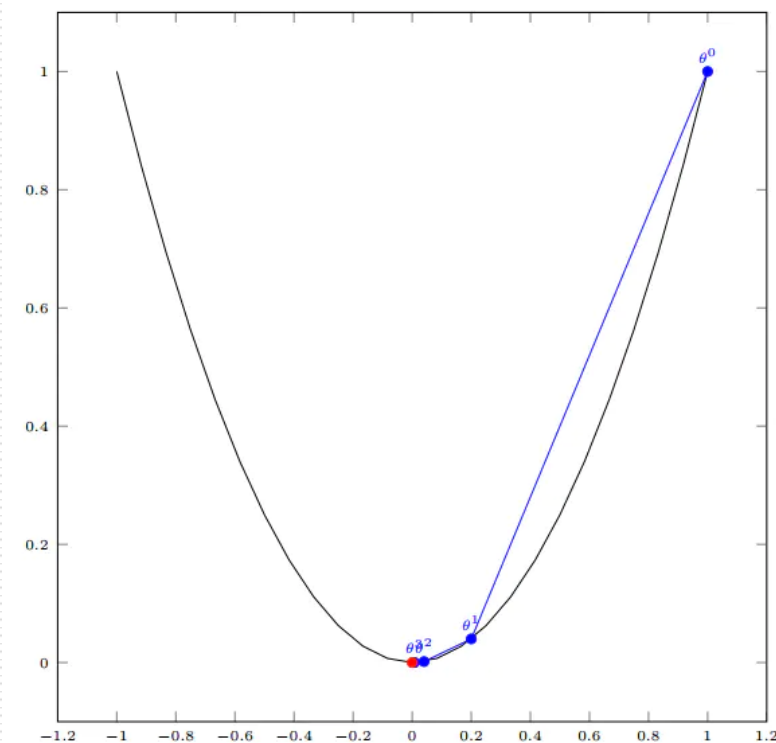
$$\theta^0 = 1$$

$$\begin{aligned}\theta^1 &= \theta^0 - \alpha * J'(\theta^0) \\ &= 1 - 0.4 * 2 \\ &= 0.2\end{aligned}$$

$$\begin{aligned}\theta^2 &= \theta^1 - \alpha * J'(\theta^1) \\ &= 0.04\end{aligned}$$

$$\theta^3 = 0.008$$

$$\theta^4 = 0.0016$$



1.2 分类

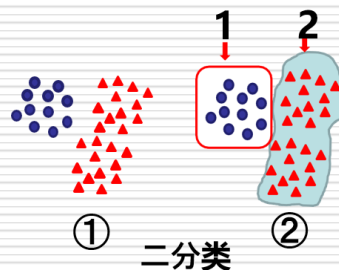
□ 分类 (Classification)

标签离散

- 身高1.85m, 体重100kg的男人穿什么尺码的T恤?
- 根据肿瘤的体积、患者的年龄来判断良性或恶性?
- 根据用户的年龄、职业、存款数量来判断信用卡是否会违约?
- 输入变量可以是离散的, 也可以是连续的

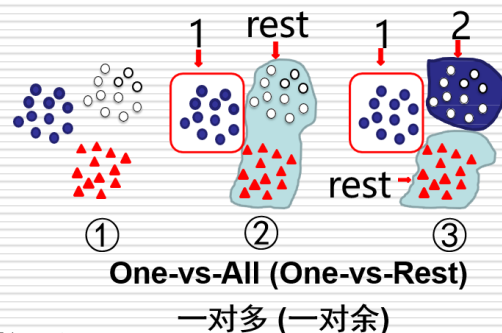
■ 二分类

- 只需要分类一次, 步骤: ①->②



■ 多分类

- 每次先定义其中一类为类型1 (正类), 其余数据为负类 (rest); 去掉类型1数据, 剩余部分再次二分类, 分成类型2和负类; 如果有n类, 那就需要分类n-1次。步骤: ①->②->③->.....
- 只分类一次, 直接输出类别1-n的其中之一作为分类结果



1.2 分类

□ 分类 (Classification)

■ Sigmoid函数

- $\sigma(z)$ 代表一个常用的逻辑函数 (logistic function) 为S形函数 (Sigmoid function)

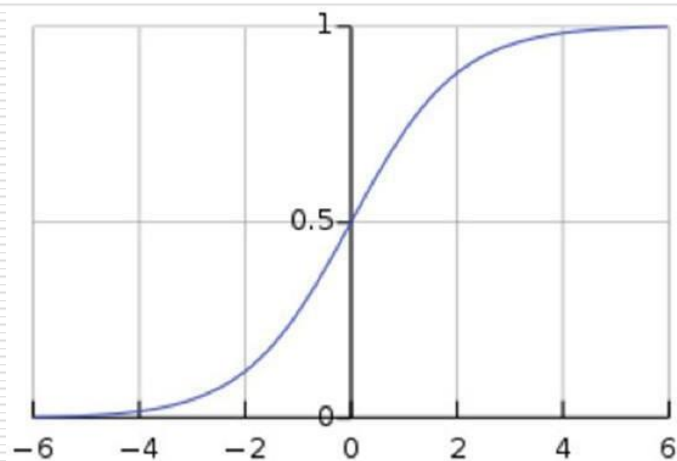
$$\sigma(z) = g(z) = \frac{1}{1 + e^{-z}}, z = w^T x + b$$
$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$

- 回归模型的假设函数

$$L(\hat{y}, y) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$$

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

其中, y 是真实的标签, \hat{y} 是预测值



✓ $\sigma(z)$ 大于等于0.5时, 预测 $y=1$

✓ $\sigma(z)$ 小于0.5时, 预测 $y=0$

1.3 梯度下降

□ 梯度下降

■ 批量梯度下降 (Batch Gradient Descent, BGD)

- 梯度下降的每一步中，都用到了所有的训练样本

参数更新

$$w_j := w_j - \alpha \frac{1}{m} \sum_{i=1}^m \left((h(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)} \right)$$

(同步更新 w_j , ($j=0,1,\dots,n$))

学习率

梯度

■ 随机梯度下降 (Stochastic Gradient Descent, SGD)

- 梯度下降的每一步中，用到一个样本，在每一次计算之后便更新参数，而不需要首先将所有的训练集求和

参数更新

$$w_j := w_j - \alpha (h(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

(同步更新 w_j , ($j=0,1,\dots,n$))

■ 小批量梯度下降 (Mini-Batch Gradient Descent, MBGD)

- 梯度下降的每一步中，用到了一定批量的训练样本

$$w_j := w_j - \alpha \frac{1}{b} \sum_{k=i}^{i+b-1} (h(x^{(k)}) - y^{(k)}) x_j^{(k)}$$

(同步更新 w_j , ($j=0,1,\dots,n$))

$b=1$ (随机梯度下降,SGD)
 $b=m$ (批量梯度下降,BGD)
 $b=\text{batch_size}$, 通常是2的指数倍, 常见有32,64,128等。
(小批量梯度下降,MBGD)

1.4 数据标准化和归一化

□ 数据标准化和归一化

- **归一化**：数据归一化的目的是使得各特征对目标变量的影响一致，会将特征数据进行伸缩变化，所以数据归一化是会改变特征数据分布的。

- 将数据映射到[0,1]区间

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

- **Z-Score**：数据标准化为了不同特征之间具备可比性，经过标准化变换之后的特征数据分布没有发生改变

- 当数据特征取值范围或单位差异较大时，最好是做一下标准化处理
- 处理后的数据均值为0，方差为1

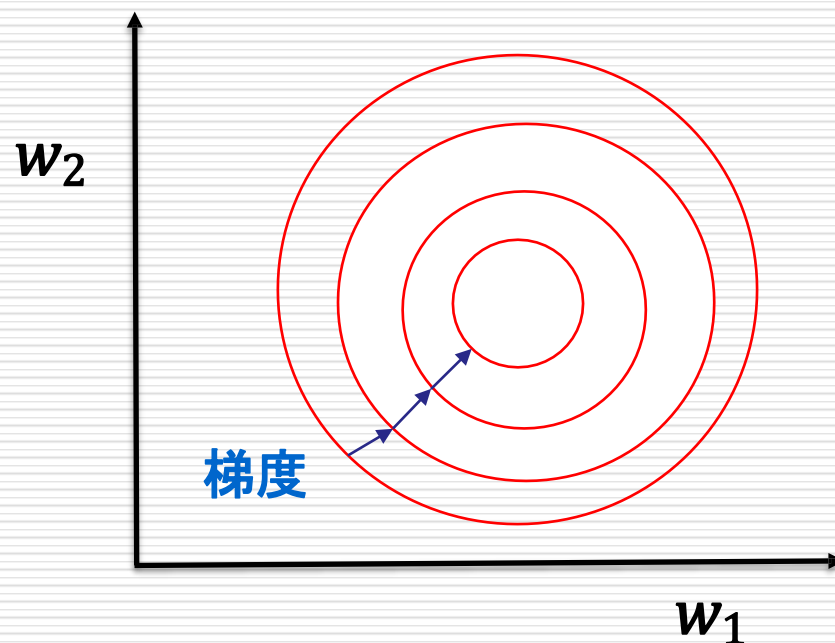
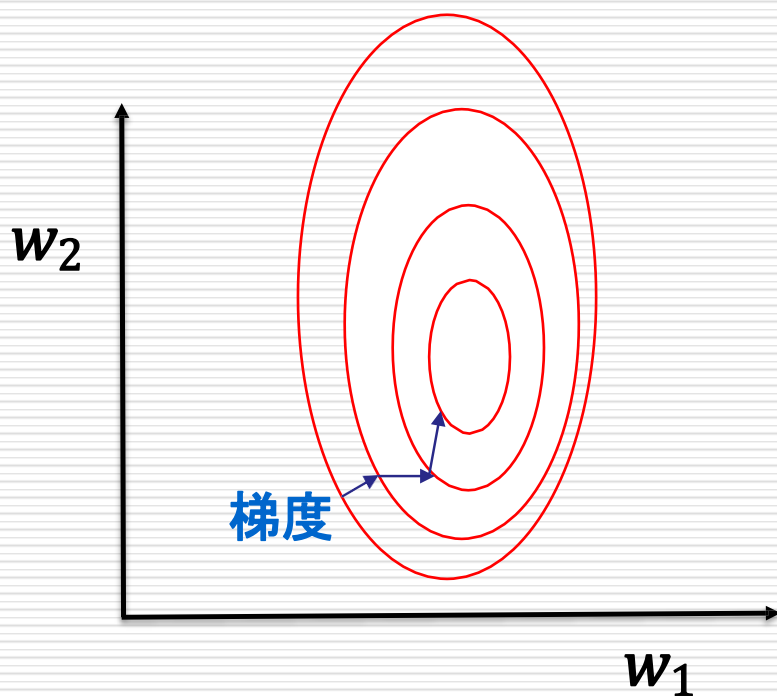
$$x^* = \frac{x - \mu}{\sigma}$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)^2$$
$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$

1.4 数据标准化和归一化

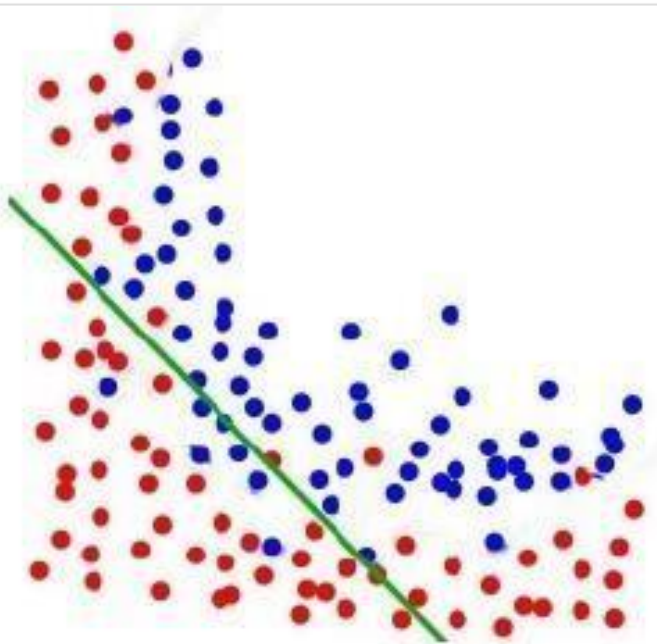
□ 数据标准化和归一化

- **提升模型精度**：不同维度之间的特征在数值上有一定比较性，可以大大提高分类器的准确性
- **加速模型收敛**：最优解的寻优过程明显会变得平缓，更容易正确的收敛到最优解

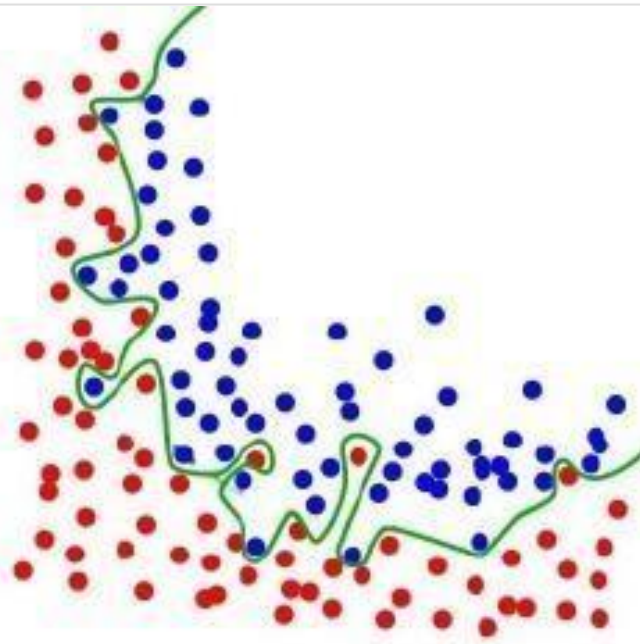


1.5 过拟合和欠拟合

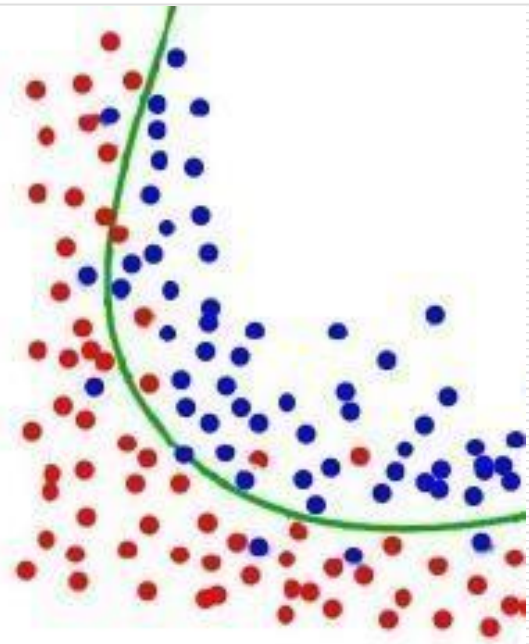
□ 过拟合和欠拟合



欠拟合



过拟合



正合适

1.5 过拟合和欠拟合

□ 过拟合和欠拟合

■ 过拟合的处理:

1. **获得更多的训练数据**。使用更多的训练数据是解决过拟合问题最有效的手段，因为更多的样本能够让模型学习到更多更有效的特征，减小噪声的影响
2. **降维**。即丢弃一些不能帮助我们正确预测的特征。可以是手工选择保留哪些特征，或者使用一些模型选择的算法来帮忙（例如PCA）。
3. **正则化**。正则化(regularization)的技术，保留所有的特征，但是减少参数的大小（magnitude），它可以改善或者减少过拟合问题。
4. **集成学习方法**。集成学习是把多个模型集成在一起，来降低单一模型的过拟合风险。
5.

1.5 过拟合和欠拟合

□ 过拟合和欠拟合

■ 欠拟合的处理

1. **添加新特征**。当特征不足或者现有特征与样本标签的相关性不强时，模型容易出现欠拟合。通过挖掘组合特征等新的特征，往往能够取得更好的效果。
2. **增加模型复杂度**。简单模型的学习能力较差，通过增加模型的复杂度可以使模型拥有更强的拟合能力。例如，在线性模型中添加高次项，在神经网络模型中增加网络层数或神经元个数等。
3. **减小正则化系数**。正则化是用来防止过拟合的，但当模型出现欠拟合现象时，则需要有针对性地减小正则化系数。
4.

2. 实验任务

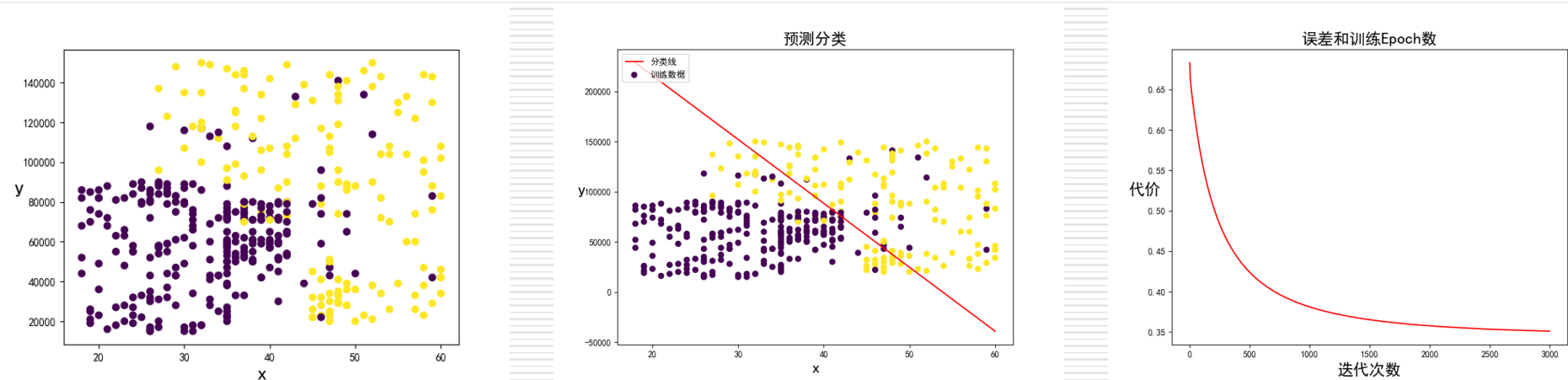
□ 购房预测分类任务

- 利用逻辑回归算法和感知机算法在给定数据集完成购房预测分类训练。

- 要求：

- 选择合适的损失函数，利用训练集完成网络训练，画出数据可视化图、loss曲线图，计算模型收敛后的分类准确率。

- 示例：



3. 作业提交说明

- 压缩包命名为：“学号_姓名_作业编号”，例：
20240423_张三_实验7。
- 每次作业文件下包含两部分：code文件夹和实验报告PDF文件。
 - code文件夹：存放实验代码；
 - PDF文件格式参考发的模板。
- 如果需要更新提交的版本，则在后面加_v2，_v3。如第一版是“学号_姓名_作业编号.zip”，第二版是“学号_姓名_作业编号_v2.zip”，依此类推。
- 截至日期：**2024年5月14日晚24点**。
- 提交邮箱：zhangyc8@mail2.sysu.edu.cn。