

从零开始学数学建模

神经网络

主讲人：北海

b站/公众号：数学建模BOOM

神经网络

主讲人：北海

b站/公众号：数学建模BOOM

- 模型简介
- 适用赛题
- 原理讲解
- MATLAB工具箱

□模型简介

□ 什么是“神经网络”

- (以下视频素材取自youtube: Explained In A Minute: Neural Networks, by Samuel Arzt)



□模型简介

□机器学习与神经网络

- 机器学习是一类实现人工智能的方法总称，让计算机模拟或实现人类的学习行为
- 神经网络是实现机器学习的一种模型
- 实现机器学习的模型还有支持向量机，决策树，朴素贝叶斯分类器等

□神经网络能用来做什么

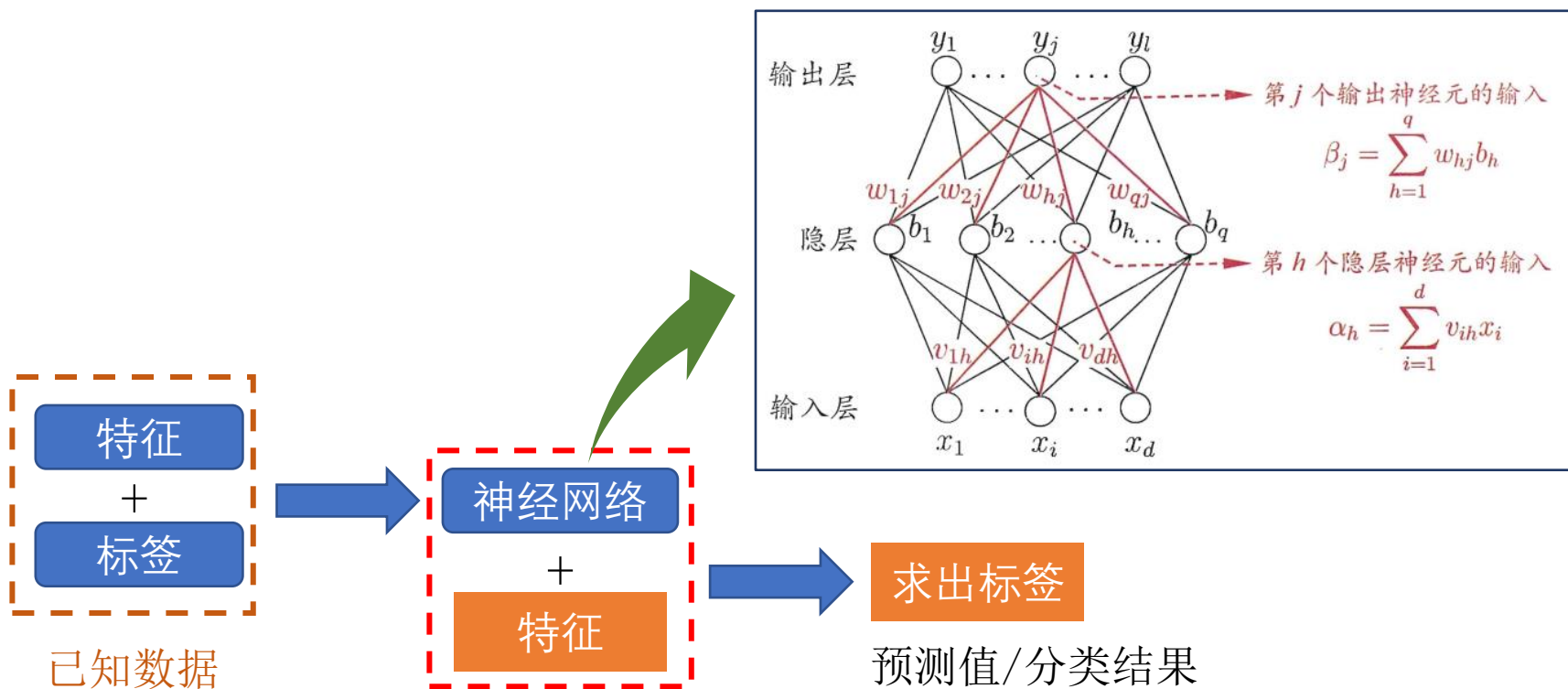
- 已知7人的身高和体重，以及是否达到8岁的数据
- 现在又来了个人，身高128 cm，体重24 kg
- 如何判断此人是否达到8岁？

序号	特征，输入值 x_1 和 x_2		标签，输出值 y
	身高(cm)	体重(kg)	到8岁
1	131	26	是
2	98	19	否
3	155	45	是
4	133	19	否
5	121	21	是
6	160	50	是
7	102	23	否

□模型简介

□ 翻译翻译什么叫做“神经网络”

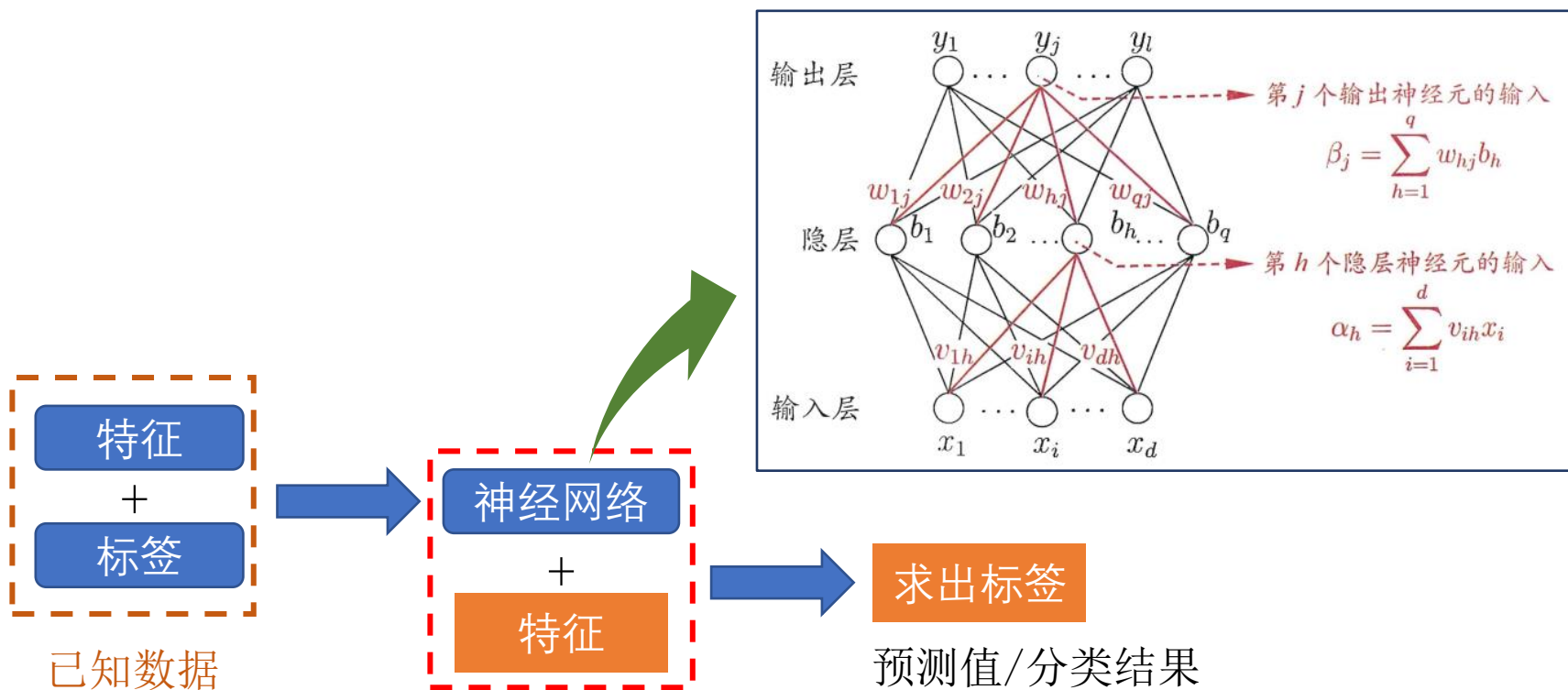
- 有现成的一组数据，既有其特征(自变量 x ，身高体重)也有其标签(因变量 y ，是否8岁了)
- 找出特征和标签之间的“关系”(建立神经网络模型)
- 从而利用该“关系”，把另一组特征对应的标签给求出来



□模型简介

□ 翻译翻译什么叫做“神经网络”

- 神经网络由**多层**的神经元(节点)构成，不同层之间的神经元通过**权重**连接
- 隐藏层和输出层的神经元中有**激活函数**，收到的数据作为自变量，计算出结果
- 计算出的**结果与权重相乘后累加**传递到下一层，或输出最终结果



求出标签
预测值/分类结果

神经网络

主讲人：北海
b站/公众号：数学建模BOOM

- 模型简介
- 适用赛题
- 原理讲解
- MATLAB工具箱

□适用赛题

□ 预测类问题

- 预测未来的结果，一般为**回归问题**，即输出值 y 是连续的
- 美赛2022年C题比特币和黄金**投资策略**，LSTM(长短期记忆网络)进行预测
- 短时交通流量预测(小波神经网络的时间序列)
- 电力负荷预测(Elman神经网络，具有延时记忆，适合动态建模)

□ 分类问题

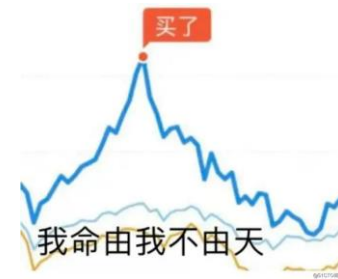
- 对现有的对象进行分类，一般为**离散问题**，即输出值 y 是有限个离散值
- 美赛2021年C题黄蜂传播规律和目击准确性，卷积神经网络进行图像识别
- 人体肿瘤诊断(LVQ神经网络，无需数据预处理，适合模式识别和优化)
- 柴油机故障判断(SOM神经网络，适合需要聚类的问题)

□ 评价类问题

- 嘉陵江水质评价(模糊神经网络)，高校科研能力评价(离散Hopfield神经网络)
- 本质上与预测和分类相同，只不过题目问的是“评价结果”而已

□ 先学会最基础的神经网络原理，遇到题目后再去翻书现学现用

参考《MATLAB神经网络43个案例分析》



□ 注意事项

□ 神经网络基本模型+其他模型=沃兹基胡硕德模型

- 需要针对具体问题量身定做，而不是把神经网络当做万能模板
- 不懂别乱用，否则将喜获成功参赛奖
- 近些年数模竞赛中，神经网络等启发式算法出现了严重的滥用



没有什么问题是
沃兹基胡硕德模型
解决不了的



low死了

很前沿×
烂大街✓

□ 举个例子：计算 $1+1=?$

- 正常做法：由数学公理可知，结果为2
- 舍本逐末：基于蚁群算法优化BP神经网络求得近似解为1.9987



狗听了都上吊

不是不能用，而是不要滥用！

神经网络

主讲人：北海
b站/公众号：数学建模BOOM

- 模型简介
- 适用赛题
- 原理讲解
- MATLAB工具箱

□从感知机到多层神经网络

□ 景区规定8周岁以下儿童可半价票，如何根据一个人的身高和体重，判断其是否可以买半价票？

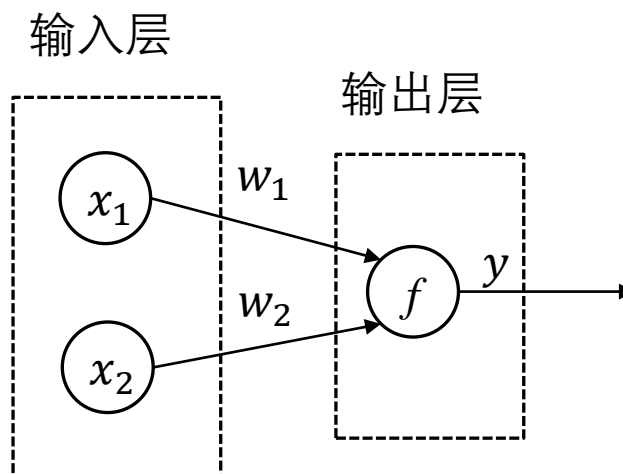
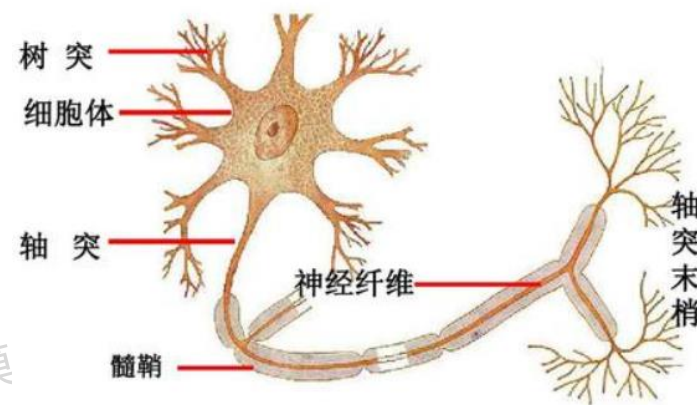
□ 生物学的神经网络

- 神经元树突收到输入信号
- 信号刺激细胞核，达到阈值后会产生输出
- 输出就是进一步的信号，传到下一个神经元

□ 感知机模型：最简单的神经网络

- 输入层：输入身高 x_1 和体重 x_2
- 激活函数：综合判断输入信号是否达到阈值
- 输出层：激活函数的函数值就是输出值 y

□ 只有输入层和输出层的神经网络，称为感知机



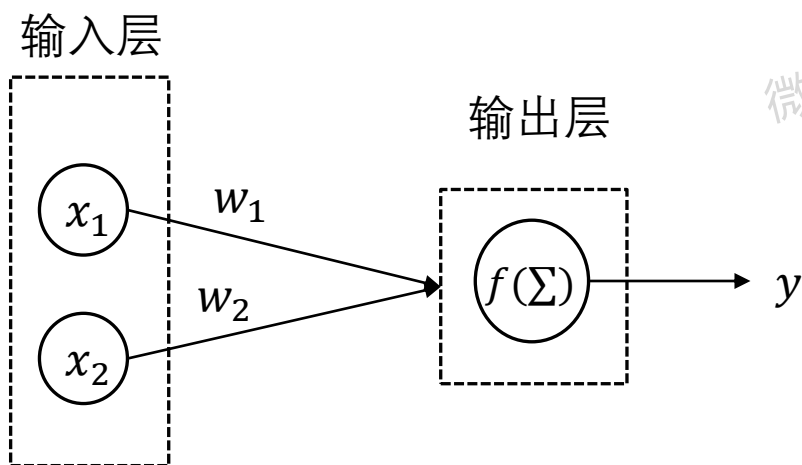
□ 从感知机到多层神经网络

□ 问题：如何综合考虑身高(x_1)和体重(x_2)，判断是否达到了8周岁(阈值 b)

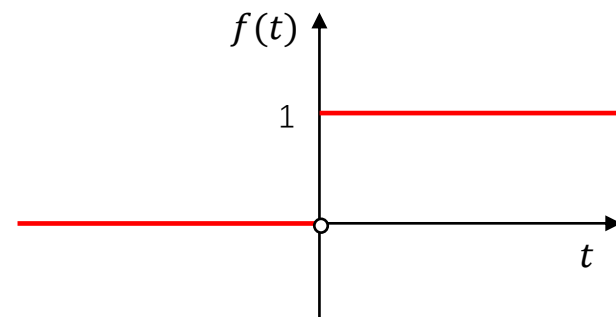
- 结果只能有“未达到”和“达到”
- 需要的函数值只有0和1，可用阶跃函数作为激活函数

□ 1、怎样实现“综合考虑”？

- 线性求和 $\sum w_i x_i$ (权重 w_i 代表变量 x_i 的重要性)
- 求和意味着身高和体重(即所有变量)都考虑到了



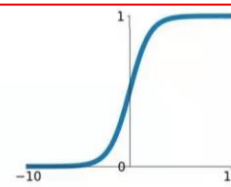
$$\text{阶跃函数: } f(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$$



其他激活函数：

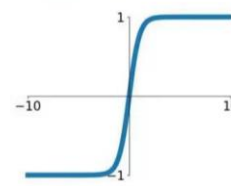
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



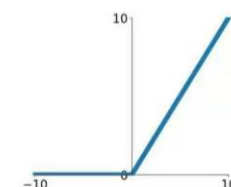
tanh

$$\tanh(x)$$



ReLU

$$\max(0, x)$$



□从感知机到多层神经网络

□ 2、如何根据“综合考虑”的结果进行判断？

- 当设线性求和的结果足够大，与阈值的差值作为激活函数的自变量 $t = \sum w_i x_i - b$

$$\text{阶跃函数: } f(t) = \begin{cases} 1, t \geq 0 \\ 0, t < 0 \end{cases}$$

极端例子：

- 假设某人身高10光年、体重100吨
- 代入 $t = \sum w_i x_i - b$ ，会是个非常大的数 ($t \geq 0$)
- 很容易判断已超过了8周岁 ($f(t) = 1$)

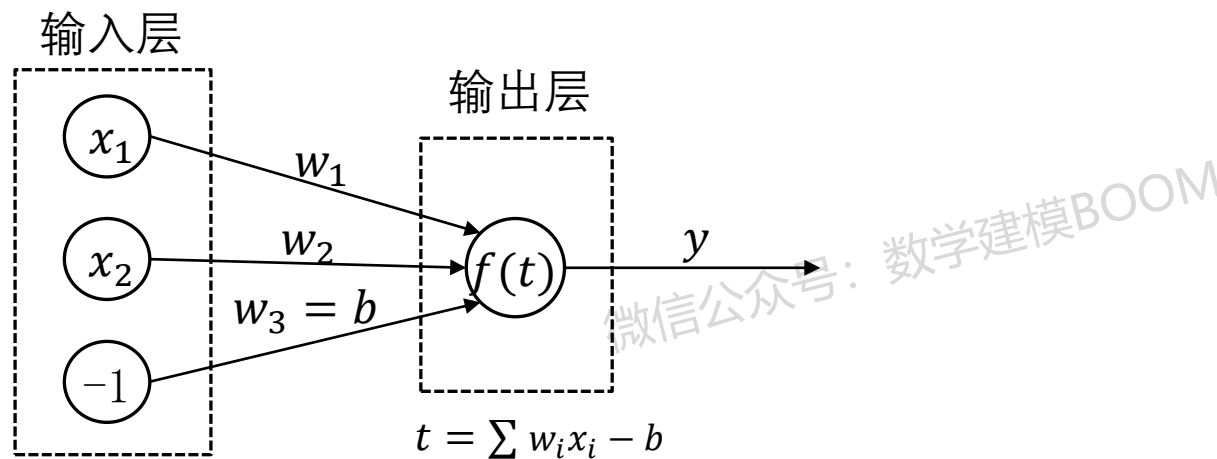


- w_i 是第 i 个变量的权重，区分不同变量对结果的影响程度
- 输入数据后，权重和阈值决定了输出结果

□从感知机到多层神经网络

□ 3、输出结果/题目答案

- 可能结果： $t = \sum w_i x_i - b \geq 0 \Leftrightarrow f(t) = 1 \Leftrightarrow$ “综合考虑” 身高和体重超过了 “阈值”
- 感知机输出值为1，则判定此人达到了8周岁
- 阈值 b 看作数值固定为-1的节点(即 $x_3=-1$)的权重 w_3



- 权重 w 如何设定？假如一个人身高1米8、体重90kg，感知机输出未达到8岁怎么办？
- 神经网络的核心，就在于如何确定权重，以确保输出正确答案

□从感知机到多层神经网络

□ 基本模型： $y = f(\sum_1^2 w_i x_i - b) = f(\sum_1^3 w_i x_i) = f(t) = \begin{cases} 1, t \geq 0, & \text{达到了8周岁} \\ 0, t < 0, & \text{未达到8周岁} \end{cases}$

- x_i 为第 i 个已知的参数
- y 为待求的输出值
- 该模型**还需确定**权值 w_i

□ 问题转化为：权值 w_i 该怎么设定

- 模型初始化时， w_i 随便设置个初始值
- **学习**：从已知数据中学得模型参数(确定权重 w_i)



微信公众号：数学建模BOOM

□从感知机到多层神经网络

□ 学习：从已知数据中学得模型(确定权重 w_i)

★□ 本质：不断地更改权重 w_i ，使得模型求出的预测值尽可能地接近真实值

- Step 1: 模型初始化，人为设置权重 w_i 初始值
- Step 2: 搜集多个人的身高 x_1 、体重 x_2 和是否达到8周岁 y 的数据

查文献和搜集数据的技巧方法在b站“快速入门”系列课程已讲过
<https://www.bilibili.com/video/BV1Rq4y1S7S8?p=6>

III P6 1-4 查文献技巧

P7 1-5 去哪找数据???

P8 1-6 数据预处理

- Step 3: 将搜集到的 x_1 和 x_2 代入模型，求得估计值 \hat{y}
- Step 4: 将模型求的估计值 \hat{y} 与搜集到的实际值 y 比较，差别越小越好
- Step 5: 直到满足终止条件，否则继续对权重 w_i 进行优化，重复Step 3到5
- 终止条件：估计值和实际值的差值小到一定程度为止

□ 从感知机到多层神经网络

□ 学习：从已知数据中学得模型(确定权重 w_i)

□ 问题转化为：Step 5中，如何对权重 w_i 进行优化？

□ 优化方法： w_i 改为 $w_i + \Delta w_i = w_i + \eta(y - \hat{y}) x_i$

- 当 $(y - \hat{y})$ 大于某定值时进行优化
- $\eta \in (0,1)$ 为学习率， η 过大容易震荡，过小则收敛速度太慢
- 达到终止条件后当前的 w_i 就作为最终的权重，模型就建好了

$$\text{基本模型: } y = f(\sum_1^2 w_i x_i - b) = f(\sum_1^3 w_i x_i) = f(t) = \begin{cases} 1, t \geq 0, & \text{达到了8周岁} \\ 0, t < 0, & \text{未达到8周岁} \end{cases}$$

- 本示例以 $(y - \hat{y})$ 作为对模型好坏的评估指标，更常见的有均方误差、平方绝对误差等

□从感知机到多层神经网络

□ 总结：景区规定8周岁的儿童以下可买半价票。但儿童往往没有身份证，难以准确判断年龄，如何根据一个人的身高和体重，判断其是否可以买半价票？

- 本题感知机：

$$\text{基本模型: } y = f(\sum_1^2 w_i x_i - b) = f(\sum_1^3 w_i x_i) = f(t) = \begin{cases} 1, t \geq 0, & \text{达到了8周岁} \\ 0, t < 0, & \text{未达到8周岁} \end{cases}$$

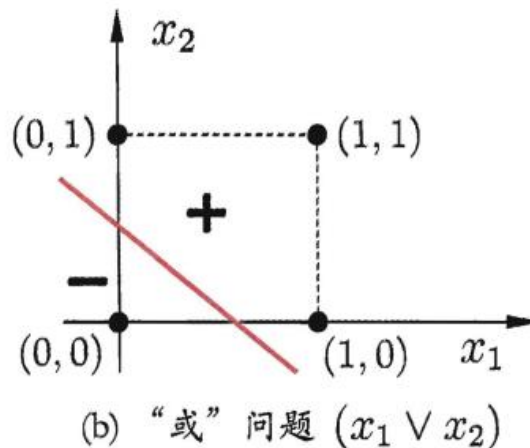
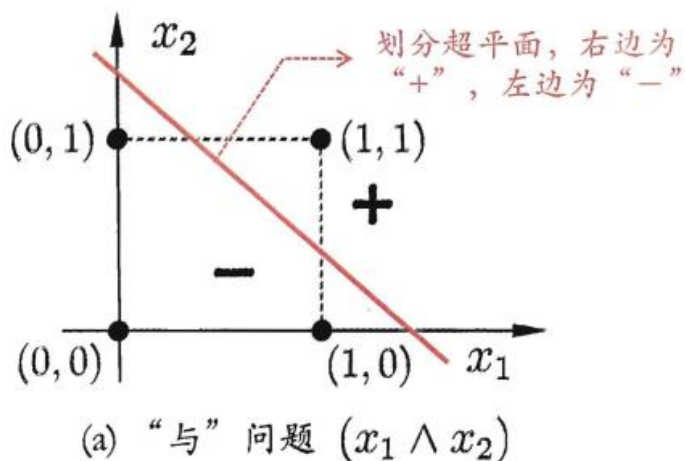
- 激活函数 f 的表达式已知，身高 x_1 和体重 x_2 已知，需要确定权重 w_i
- 根据搜集到的训练数据集，不断优化权重 w_i
- 直到达到终止条件后模型就建好了
- **Question**：既然单层的感知机能解决问题，为什么常见的神经网络都是多层的？



□从感知机到多层神经网络

□ 更复杂的情况

- 感知机只拥有输出层进行激活函数处理，若问题是线性可分的，则感知机一定会收敛
- 若数据集里超过八岁的人大多数都是身高较高、体重较大的
- 训练出的模型判定标准可能是“身高较高且体重较大的判定为达到8岁”，如图一所示
- (下图源自《机器学习》)



- 思考：什么样的数据集可能训练出图二“逻辑或”的结果？

□ 从感知机到多层神经网络

□ 更复杂的情况

- 非线性可分问题：异或问题，感知机无法收敛
- 例如：随机选取一组两个人，判断是否有能力生育出一个孩子
- 男性为1，女性为0作为输入，能生育为1、不能生育为0作为输出

□ 使用数据集进行训练时

- 假如数据集里某一段全是(0, 0)、(1, 0)和(0, 1)
- w_1 和 w_2 会收敛于某组定值(对应图中下方红线，划分两种模式)
- 假如数据集里某一段全是(1, 1)、(1, 0)和(0, 1)
- w_1 和 w_2 会收敛于另一组定值(对应图中上方红线)

但全面的数据集里肯定四种组合都有，则权重会不断震荡、无法收敛
本质就是权重无法同时在两条边界(图中红线)收敛

如何理解「异或」的含义？

知乎 · 36 个回答 · 375 关注 >



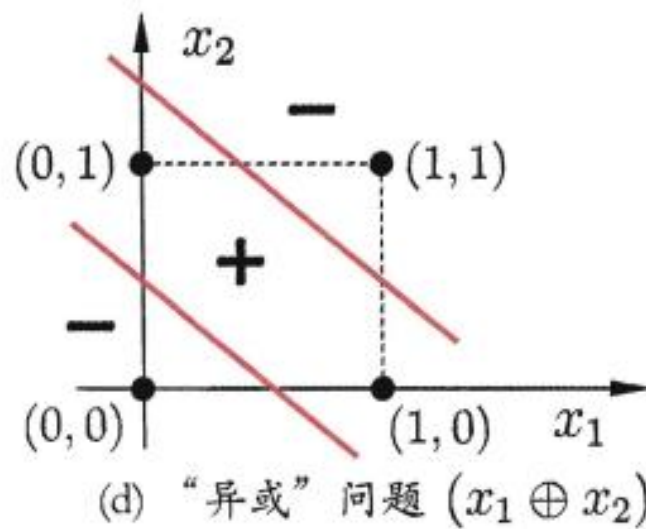
SEEN

唯有自律，才是解决人生痛苦的根本...

384 人赞同了该回答

男性和女性能生出孩子，否则就不行。

发布于 2016-11-20 22:07 · 著作权归作者所有



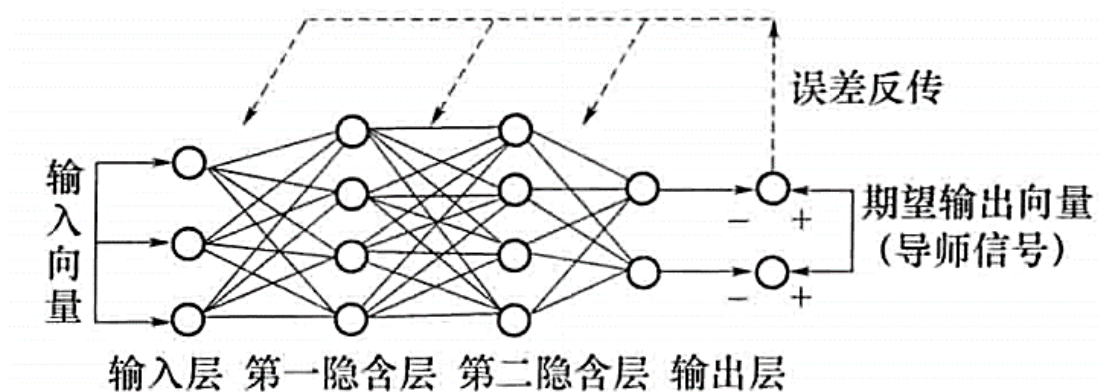
□从感知机到多层神经网络

□ BP神经网络 (back propagation, 误差逆传播)

- BP神经网络等于 **叠加多个感知机来实现非线性可分**
- 梯度下降策略：以目标的负梯度方向对参数进行调整（此外还有牛顿法、最小二乘等策略）
- 第 k 个训练例的均方误差：

$$E_k = \frac{1}{n} \sum_{j=1}^l (y_j^k - \hat{y}_j^k)^2$$

- 更新权重 $w_{hj} \leftarrow w_{hj} + \Delta w_{hj} = -\eta \frac{\partial E_k}{\partial w_{hj}}$ ，偏导展开最终表达式还是用 $(y_j^k - \hat{y}_j^k)$ 来计算
- 因此基本原理与前面感知机相同



□从感知机到多层神经网络

□ 总结

- MATLAB和GitHub上都有现成的工具包，无需原创代码
- 再复杂的神经网络，思路都可总结为以下三点：

- 搜集数据集 (x_i, y) ，将 x_i 代入初始的神经网络模型，计算估计值 \hat{y}
- 根据 \hat{y} 与 y 的差异不断 **更改权重** w_i ，使其误差尽可能的小(梯度下降)
- 大量数据训练后，再输入新的数据 x ，模型就可以求出 y 用于分类/预测/评价了

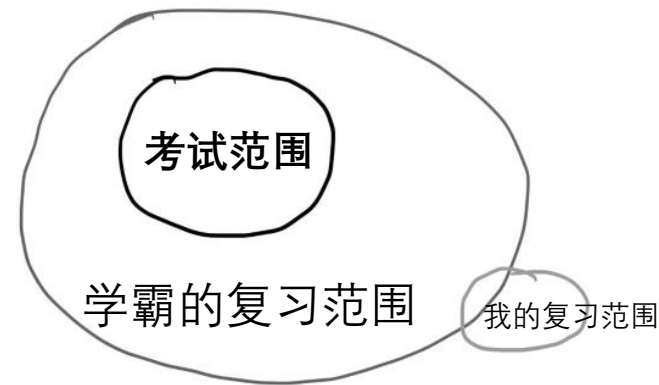
- **数据集**是题目给的，或者自己搜集的，决定了整个模型的生死
- 记得做好数据预处理(处理缺失值、异常值)

□从感知机到多层神经网络

□ 模型的泛化能力：确保模型真的能用

- 如何证明你建立的神经网络模型是靠谱的？

任务	方法	判断标准	是否优秀
神经网络建模	学习(优化权重)	达到终止条件	测试新的数据
北海准备高考	做习题、做错了改	全都会做了	参加模拟考试



- 翻译翻译：课后题都会做了，并不能保证考试一定能考好

根本原因：虽然课后题全都会做了，但考试不一定考你会的啊

- 留出法：将搜集到的数据划分为训练集、验证集和测试集
- 训练集进行学习、训练参数（权重）
- 验证集来检验性能（学习率等），以调整超参数或及时停止训练
- 测试集给出客观的评价

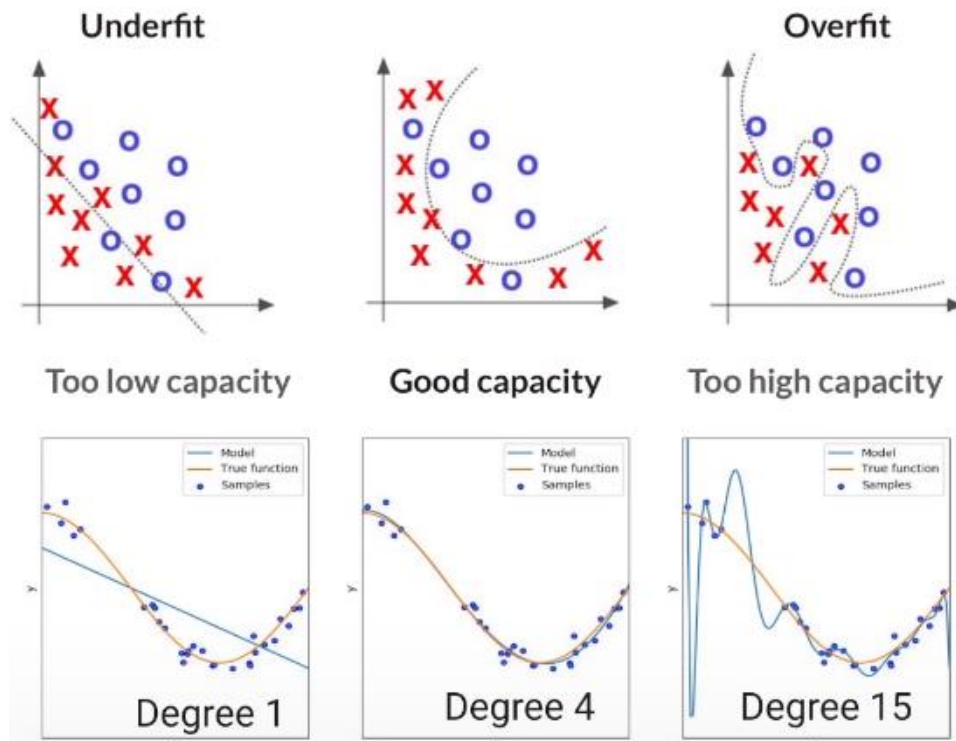
- (此外还有交叉验证法、自助法)

□从感知机到多层神经网络

□可能遇到的问题

- 过拟合：把特殊当一般(见过的天鹅都是白的，就以为天鹅只有白色的)
- 欠拟合：一般性质都没学好(“后浪”里讲述的都是年轻人，就以为年轻人都是后浪)

过拟合是因为“死记硬背”，欠拟合是因为“啥都不会”



□从感知机到多层神经网络

□ 过拟合的原因

1. 数据集有噪声(缺失值、异常值, 无用特征等等)
2. 训练集的数据量不足(训练集的分布与总体样本的不一致)
3. 训练模型过度, 导致模型非常复杂(完全拟合了训练集、无法接受新数据)

□ 解决方法

- 尽量寻找更多的数据集, 并做好预处理
- 去除无用的特征(判断一个人是否达到8岁, 与他鞋子的颜色无关)
- 早停策略: 在即将过拟合的时候立刻停止训练
- 正则化策略: 损失函数中加入惩罚项来降低模型复杂度, 特征越多惩罚越大

□ 欠拟合

- 与过拟合相反, 是由于训练不足造成的, 可在模型中增加特征来增大网络复杂度

(以上概念了解即可, 具体数学证明不用掌握。想深入学可看周志华《机器学习》第2和第5章)

□从感知机到多层神经网络

□ 学习建议

- 本期课程帮大家快速了解神经网络的基本思想，避开了大量数学推理证明
- 想深入学习的话，推荐：

(1)吴恩达机器学习课程，b站可以搜到，较为适合入门

(2)周志华《机器学习》，b站也有讲解课程

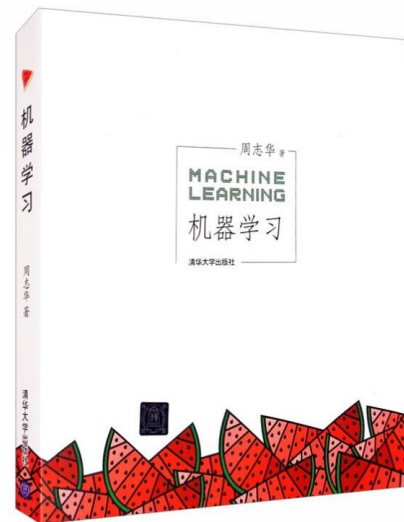
(3)学Python，要会用Numpy, Pandas, seaborn和matplot四个包

(4)MATLAB的工具箱，MATLAB帮助中心有讲解：

<https://ww2.mathworks.cn/help/deeplearning/getting-started-with-deep-learning-toolbox.html>



[中英字幕]吴恩达机器学习系列课程



- 对于做数模竞赛，先理解本节课讲的概念，然后学会使用MATLAB实现BP神经网络(不需要过于深究原理)，等比赛时遇到赛题再去翻书查资料
- 机器学习/神经网络/深度学习背后的水很深，立志深耕该领域的话做好脱发的准备

神经网络

主讲人：北海

b站/公众号：数学建模BOOM

- 模型简介
- 适用赛题
- 原理讲解
- **MATLAB工具箱**

□代码求解

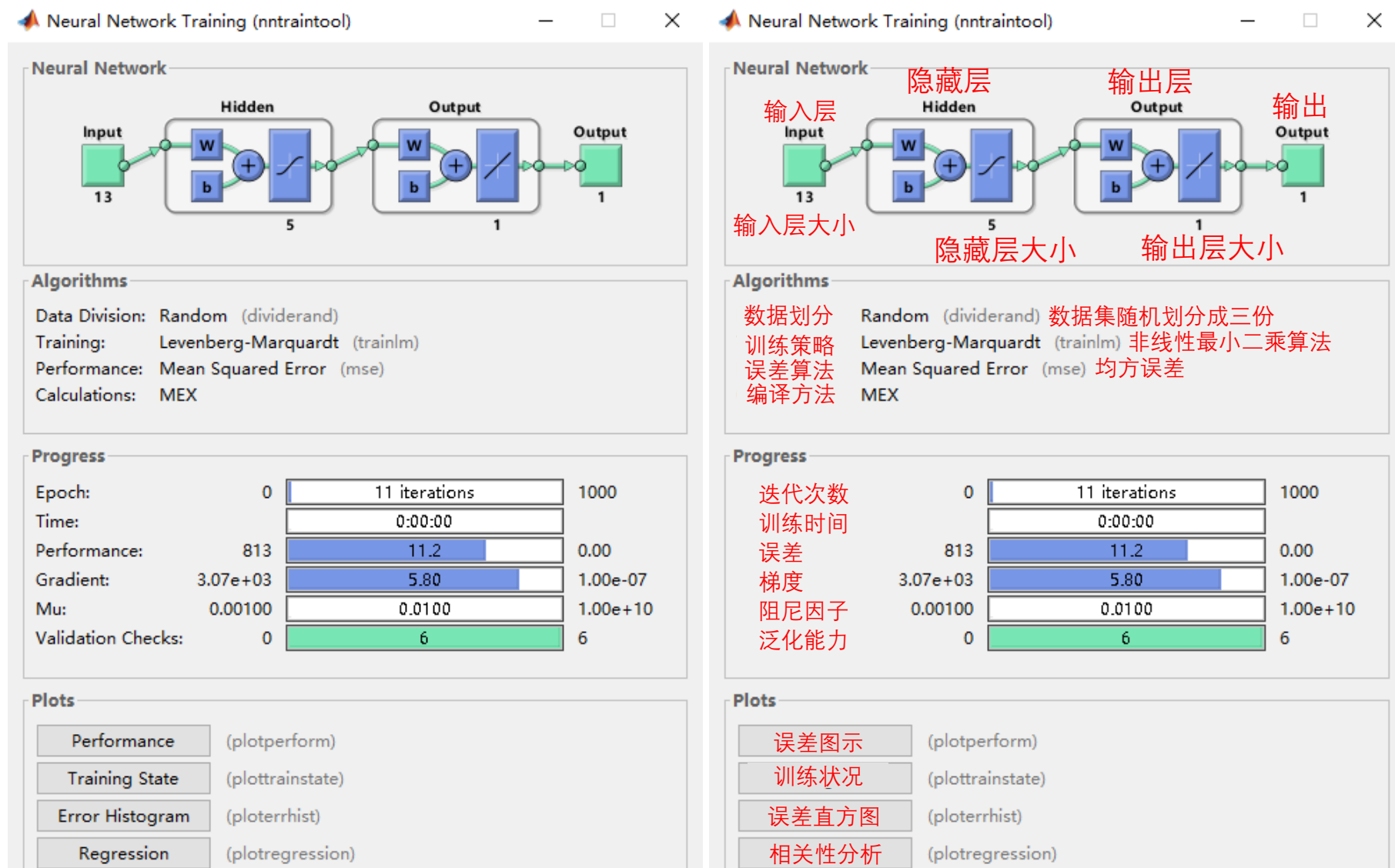
□ 体脂率估计

- 现有252人的年龄、身高、体重等13项身体指标的数据和他们的体脂率数据
- 现在已知张三的这13项身体指标，请预测张三的体脂率

年龄(岁)	重量(磅)	身高(英寸)	颈围 (cm)	胸围 (cm)	腹围 (cm)	臀围 (cm)	大腿围 (cm)	膝围 (cm)	踝围 (cm)	上臂(伸展)围 (cm)	前臂围 (cm)	腕围 (cm)
26	159.25	77.75	33.2	88.1	87.5	89.1	56.7	33.2	19.2	32.1	25.4	16.7

- 数据集已有，导入MATLAB，调用工具箱建立神经网络
- 再把张三的数据代入模型，求得结果
- 注意需要把建好的模型的各项指标写进论文并分析
- 代码文件NNs.mlx，需要与数据NNs.xlsx放在同一文件夹下
- 视频讲解了使用神经网络工具箱app的可视化操作

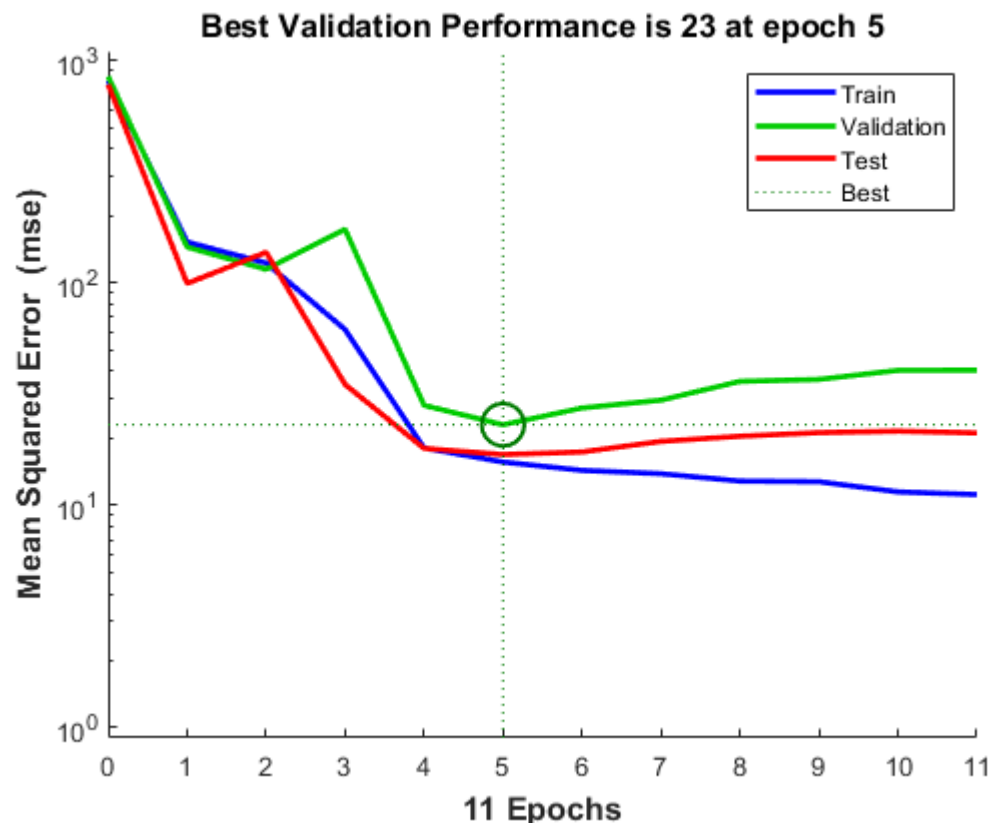
□代码求解



□ 代码求解

□ 均方误差

- 蓝绿红分别是训练集、验证集和测试集的均方误差
- Best虚线说明训练到第5代时结果最理想



学建模BOOM

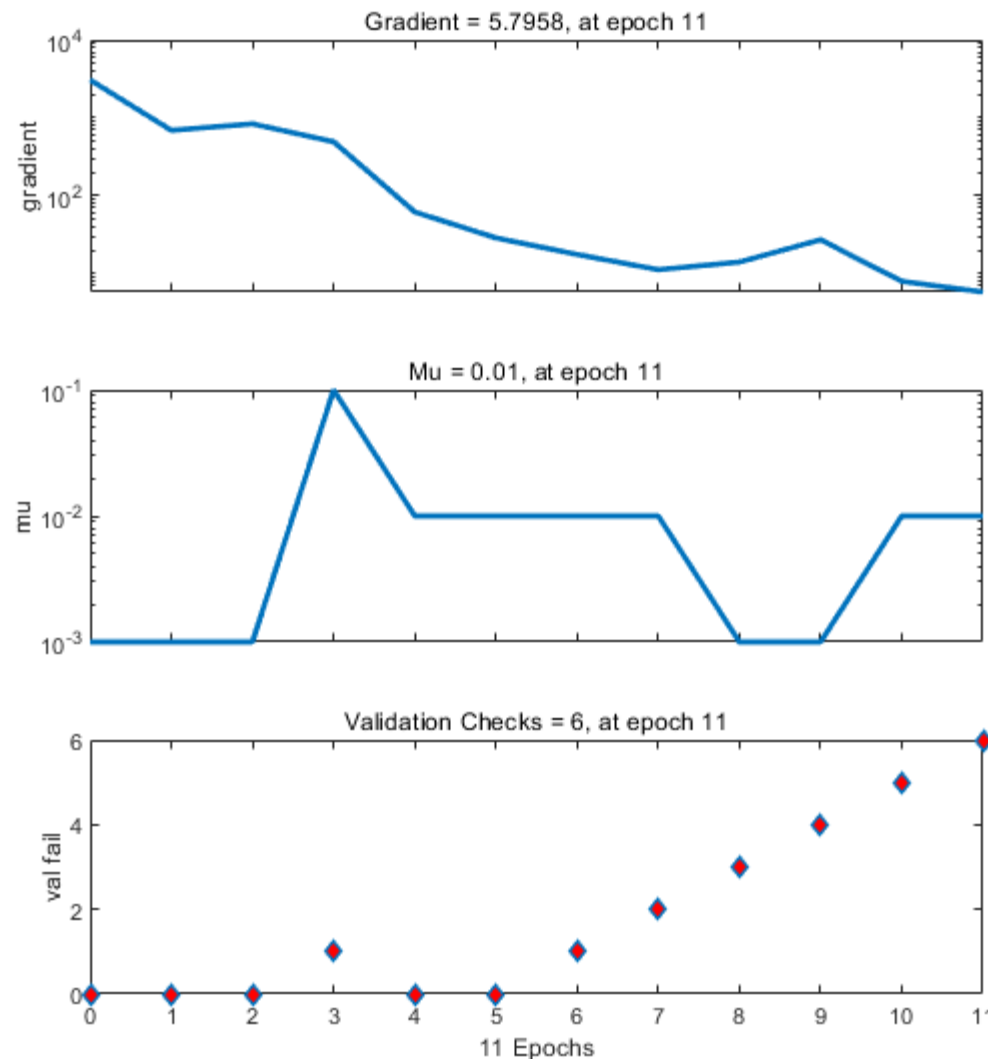
□ 代码求解

□ 梯度、阻尼因子和泛化能力

- 图一：**梯度**整体上在不断下降
- 图二：该神经网络采用**非线性最小二乘**策略，阻尼因子值越大意味着算法收敛效果越好
- 图三：泛化检验，训练集每完成一代训练，就用验证集检验误差。**图中纵轴值为0表示在训练中误差持续降低**。默认当连续6次误差不降低时终止训练

□ 防止过拟合

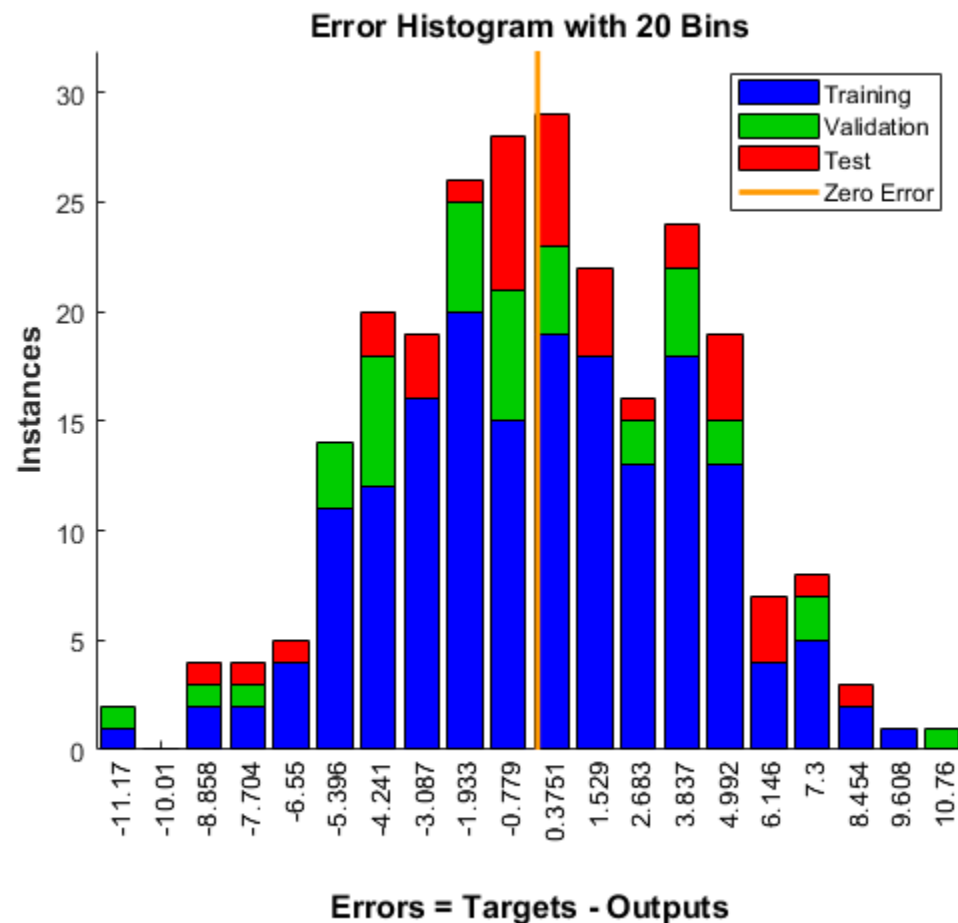
- 图三中，从第6代训练开始，验证集的误差连续多次都没有下降，说明训练得已经足够了
- 不及时停止的话就可能陷入过拟合



□ 代码求解

□ 误差直方图

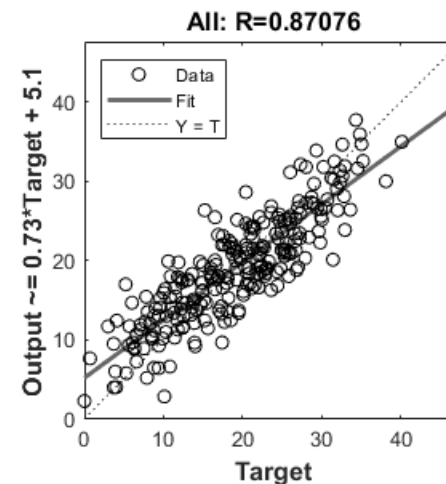
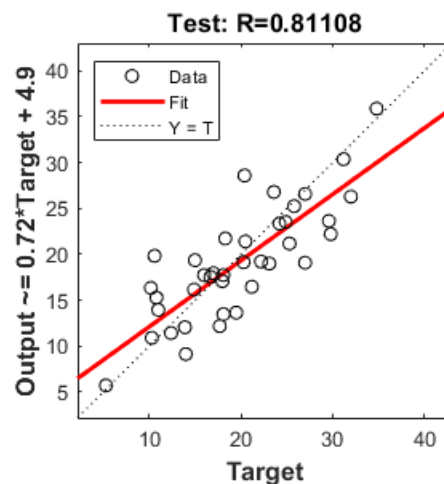
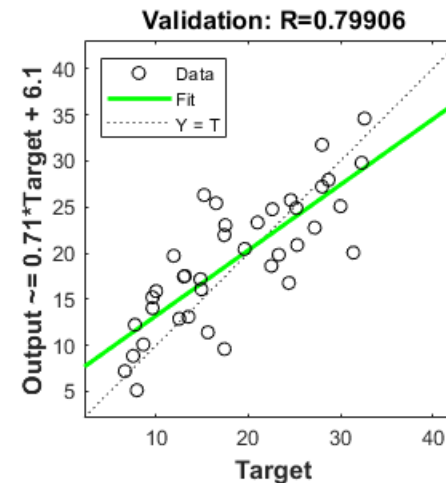
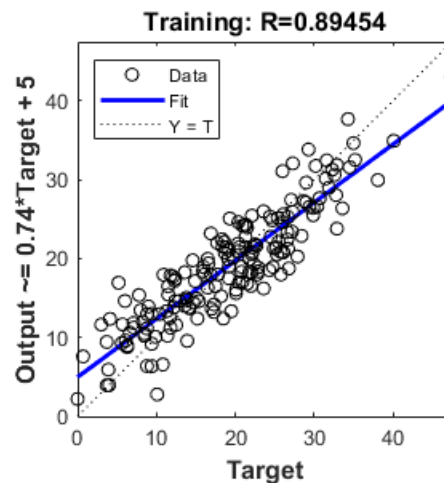
- 蓝色、绿色和红色分别代表训练集、验证集，和测试集
- 横坐标：误差区间的中位数
- 纵坐标：位于该误差区间的样本个数
- 完美情况是误差都为0，只有一根柱在中间0点
- 正常情况是0附近的样本最多，绝对值越大样本越少



□代码求解

□ 目标相对于输出的回归图

- 4张图分别是训练集，验证集，测试集和所有数据的回归图
- 横坐标：样本的实际值；
- 纵坐标：模型算出来的估计值；
- R为目标值和预测值的相关度
- R越接近1，表示线性化程度越高，结果越好
- 具体R多高算“好结果”，需要看题目本身的难度



□ 写出你的笔记

□ 费曼学习法



费曼学习法

- 费曼学习法：以教代学
- 只有当你能够教会别人，才代表你真正学会了！

□ 有奖征集：每学完一期课程，整理笔记，发布在各平台

- 将你每节课所学到的，整理出一套笔记
- 尽量不要照搬或截图课程的内容
- 可自行发布在知乎/CSDN等等各类平台

① 确定主题开始学习

② 理解所学内容

③ 把所学内容讲给别人

④ 把讲不清楚的地方去学明白

- 符合以下要求的文章，且文章点赞超过100或浏览量超1万的，可获取半价退款奖励（联系北海的QQ：1980654305）
- 1、标题设为：XXXX（模型或算法）——北海数学建模课程笔记
- 2、文章首行写：本文为北海的数模课程学习笔记，课程出自微信公众号：数学建模BOOM。

□ “从零开始学数学建模” 系列课程

- 本期课程视频出自**b站up**：数学建模BOOM
- 全套课程请关注**微信公众号**：数学建模BOOM，回复“课程”

END

微信公众号：数学建模BOOM