

神经网络训练手册

对隐藏层的感性认识:

假设我们要区分几张图片是否为人脸。

1 → 是人脸 // 0 → 不是人脸。

则输入层的节点数与图片像素数相同，输出为1个节点(0或1)

那么隐藏层如何分析?

我们可以将这个问题分解为一些子问题，比如:

• 在上方有头发吗?

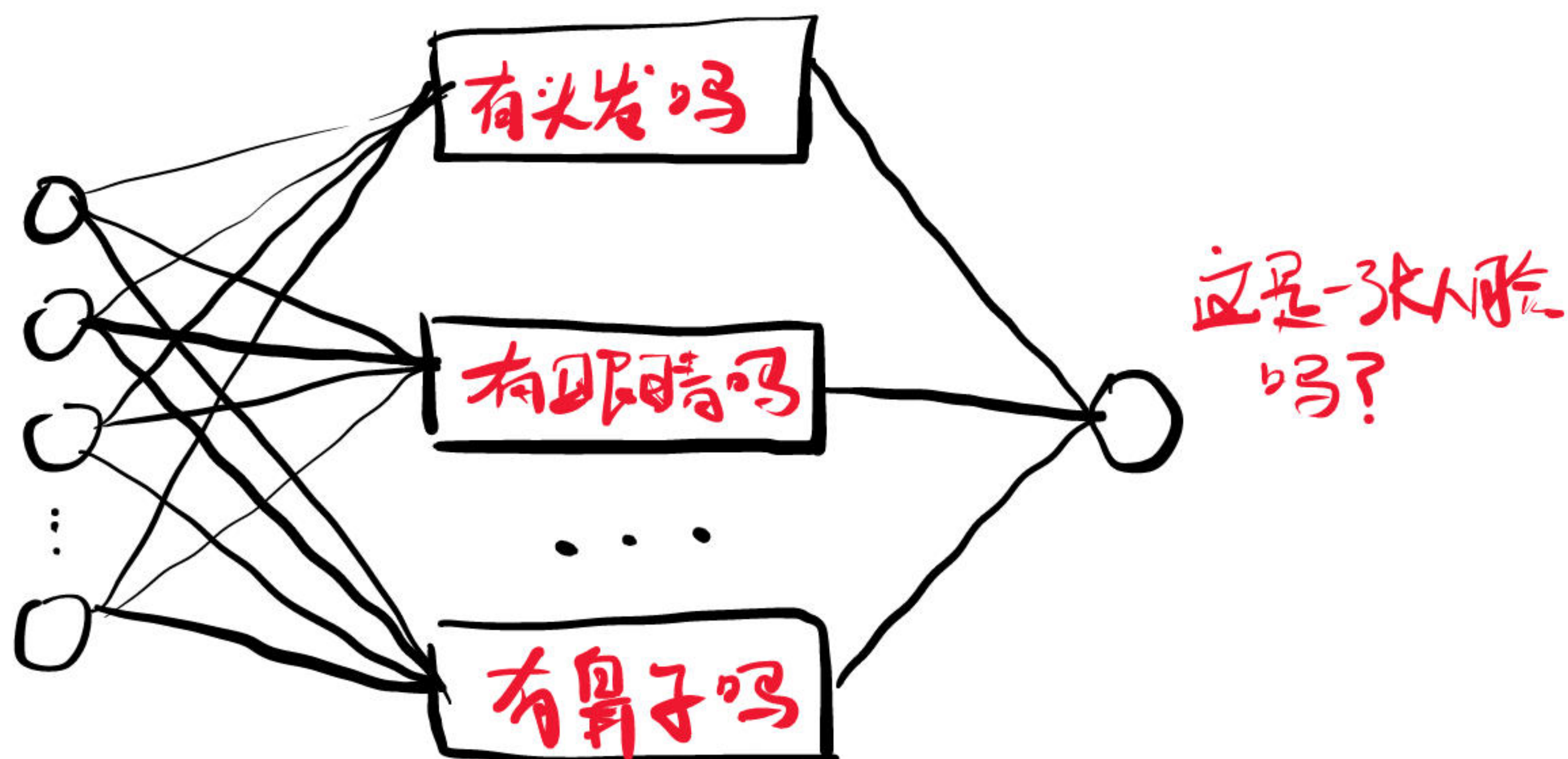
• 在左上、右上各有1个眼睛吗?

• 在中间有鼻子吗?

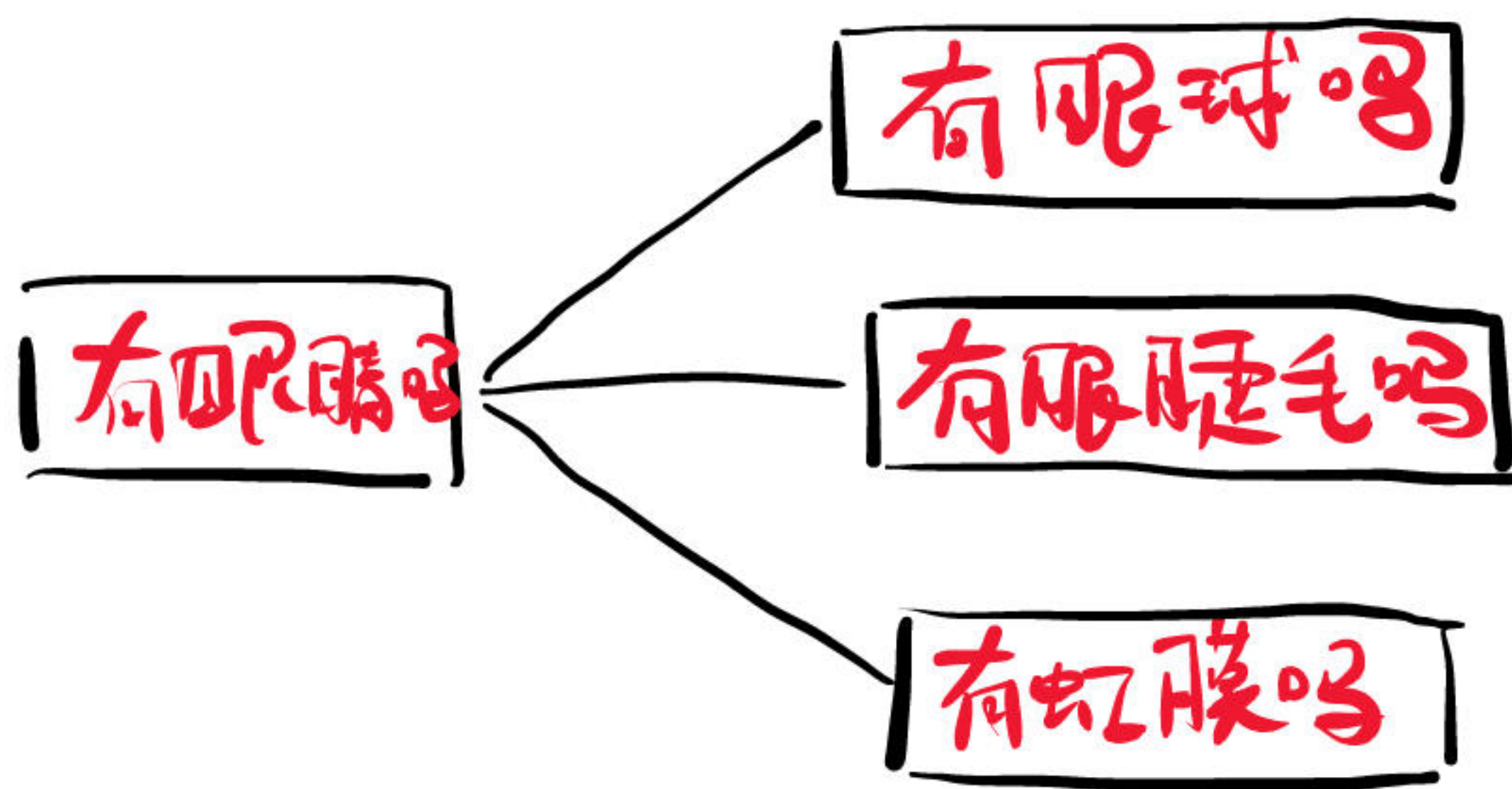
...

假如以上这些问题大部分或者全部为yes,

则认为是一张人脸



对于每个子问题,我们又可以进一步拆分



子问题是还将继续分解,直至最终的问题可以在一个神经元上被回答.

ML { Supervised Learning ✓
Unsupervised Learning

训练 (Training):

一个初始的神经网络通过不断优化自身参数,来让自己变得更准确的过程。

推理 (Inference):

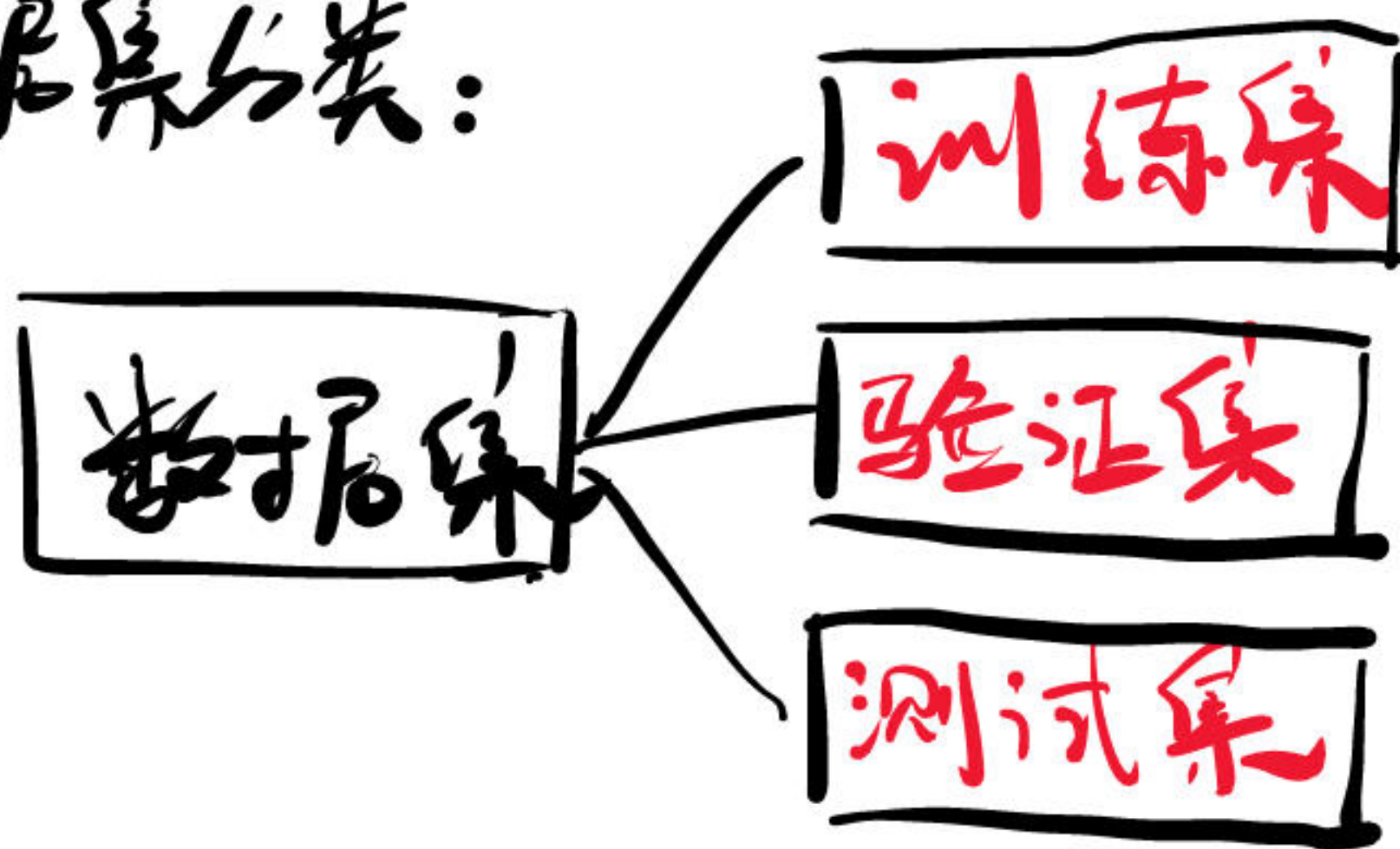
用训练好的神经网络模型对现场数据进行识别的过程。

根本问题:

深度学习最根本的问题是优化和泛化的对立。

- 优化: 指调节模型使其在训练数据上得到最佳性能。
- 泛化: 指训练好的模型在前所未见的数据上的性能表现。

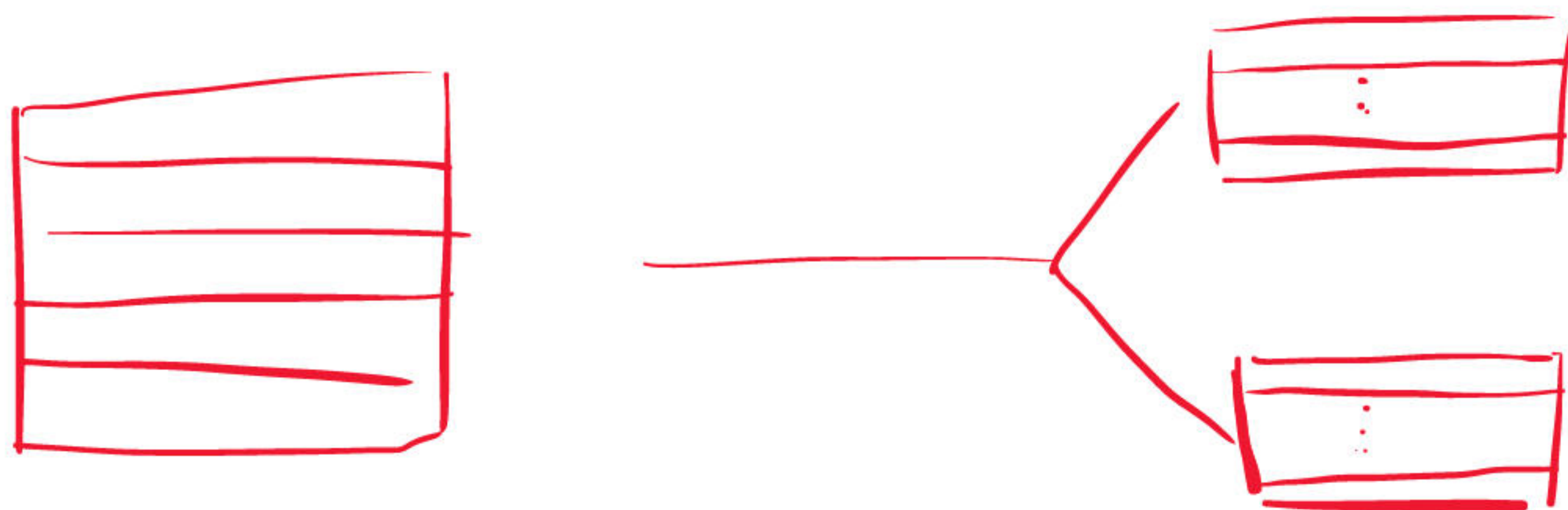
数据集分类:



- 训练集: 实际训练算法的数据集
- 验证集: 用于跟踪学习效果的数据集
- 测试集: 用于产生最终结果的数据集.

交叉验证:

即把数据多次随机分割作为训练集与测试集,最后算出多次分割准确度的平均值,作为最后的准确度.



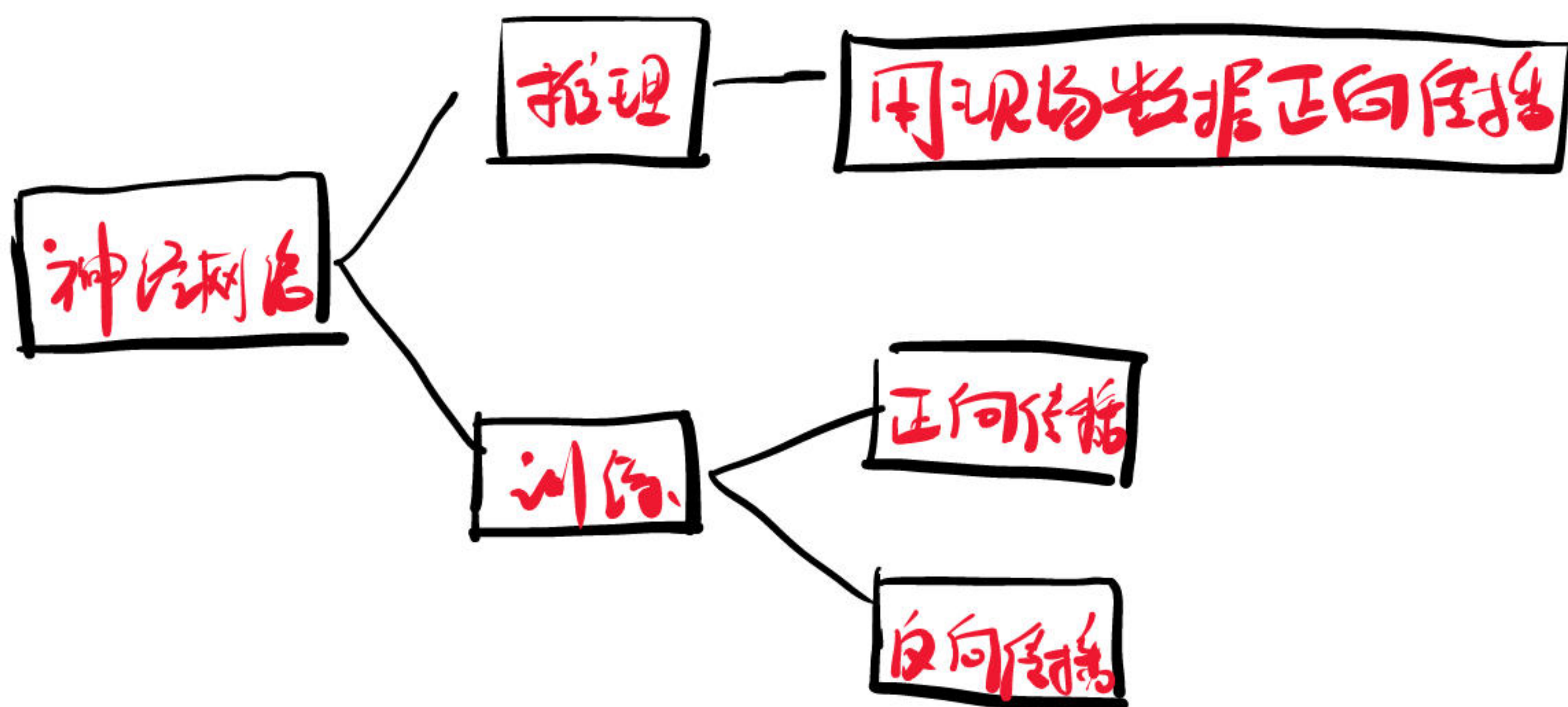
训练过程:

• 正向传播

输入信号从输入层经过各个隐藏层向输出层传播,在输出层得到实际的响应值

• 反向传播

按照梯度下降的方法从输出层经过各个隐含层并逐层不断地调整各神经元的连接权重与阈值.



简单概念名词:

- 代 (Epoch):

使用训练集的全部数据进行一次完整训练

- 批大小 (Batch size):

使用训练集的一小部分样本对模型权重进行一次反向传播,这一小部分样本数据被称为一批数据。

- 迭代 (Iteration):

使用一个 Batch 数据对模型进行一次参数更新的过程,被称为一次迭代。每一次迭代得到的结果都会被作为下一次迭代的初始值

$$\text{Number of Batches} = \frac{\text{Training Set Size}}{\text{Batch Size}}$$

比如有训练集有 500 个样本, Batch size = 10, 那么训练完整个样本集: iteration = 50, epoch = 1

神经网络训练过程

- ① 提取特征向量作为输入
- ② 定义神经网络结构
- ③ 通过训练,利用反向传播不断迭代权重
- ④ 使用预测好的神经网络来预测未知数据.

参数随机初始化

对于所有的参数我们必须要初始化他们的值,他们的初始化值不能相同,同时也不能初始化为0和1,例如:

- Xavier初始化 (sigmoid tanh)
- 权重初始化正态分布 (relu)
- MSRA Fliter (relu)
- 初始化为小的随机数

标准化 (Normalization)

由于输入数据范围较大, 同时样本中各数据可能量纲不一致。因此需要对其进行标准化处理

其中最典型的为数据的归一化处理, 即可将数据映射到 $[0, 1]$ 的区间上。

※ Z-score 标准化: (零均值归一化)

- 经过处理的数据均值为 0, 标准差为 1 (正态分布)
- 其中 μ 是样本均值, σ 为样本标准差。

损失函数

损失函数用于描述模型预测值与真实值的差距大小。一般有两种常见算法——均方平方差和交叉熵

• 均方平方差 (MSE):

$$MSE = \frac{1}{2} \sum_{i=1}^n \frac{1}{n} (f(x_i) - y_i)^2$$

• 交叉熵 (cross entropy):

$$C = -\frac{1}{n} \sum_x [y \ln a + (1-y) \ln (1-a)]$$

损失函数的选取取决于输入标签的数据的类型。

{ 输入实数 ——— 用 MSE
输入位向量 ——— 用 CE

梯度下降法

$$\theta_{t+1} = \theta_t - \alpha_t \nabla f(\theta_t)$$

学习率



学习率是一个超参数, 用于控制梯度下降的速率。

新权值 = 当前权值 - 学习率 × 梯度

泛化能力分类

- 拟合
- 欠拟合
- 不收敛
- 过拟合

Solution: 1. 减少特征

2. Early stopping

3. 更多训练样本

4. 重新清洗数据

5. Dropout

一. Early stopping

当 accuracy 不再提高时停止训练。

二. Drop out

随机删除一些隐藏层神经元, 同时保持输出层和输入层神经元个数不变。