正则化思想出发，提供新的机器学习算法

冯子扬 201688035 电计1601

1. 前言 - 正则化思想介绍

> 正则化思想，主要用于机器学习算法的平衡模型简单化和训练数据的拟合度。

在模型过于复杂的情况下，模型会学习到很多特征，从而导致可能把所有训练样本都拟合到，这样就导致了过拟合。解决过拟合可以从两个方面入手，一是减少模型复杂度，一是增加训练集个数。而正则化就是减少模型复杂度的一个方法。即以最小化损失和复杂度为目标（结构风险最小化）：

$$J(w) = Loss(x, w) + \lambda Complexity(w)$$

比如在逻辑回归中，通常可以在目标函数(经验风险)中加上一个正则化项$\Phi(w)$，即

$$J(w) = -\frac{1}{m}[\sum_{i=1}^{m} y_i log h_w(x_i) + (1 - y_i) log(1 - h_w(x_i))] + \lambda \Phi(w)$$

而这个正则化项一般会采用L1范数或者L2范数。其形式分别为$\Phi(w) = || \ w \ ||_1$ 和 $\Phi(w) = || \ w \ ||_2^2$ 。
以 L2 正则化为例，

$$L_2 \quad regularization \quad term = || \ w \ ||_2^2 = w_1^2 + w_2^2 + \cdots + w_n^2$$

有如下特点：

- 复杂度等于权重的平方和
- 可以减少非常大的权重
- 对线性模型来说首选比较平缓的斜率
- 贝叶斯先验概率：权重应该以 0 为中心，并呈正态分布

上述目标函数中的标量 λ 为正则化率，用来调整正则化项的整体影响，平衡模型简单化和训练数据的拟合。增大 λ 将增强正则化的效果，但过高的 λ 也会导致欠拟合风险。$\lambda = 0$ 时可以取消正则化。

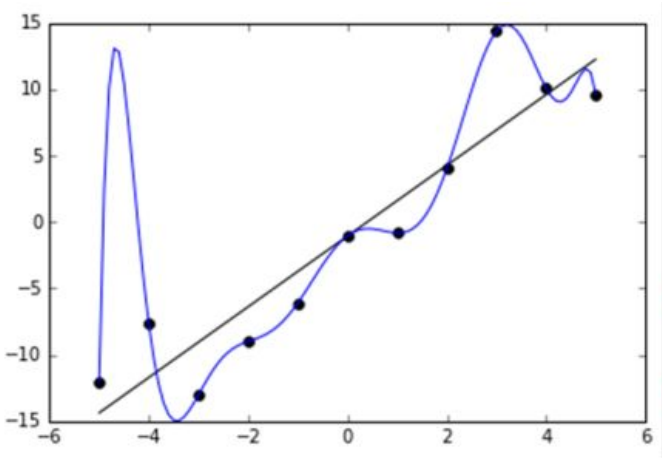
注意：较低的学习速率通常会产生和强 λ 类似的效果（都会产生较小的权重），因而不建议同时调整这两个参数。

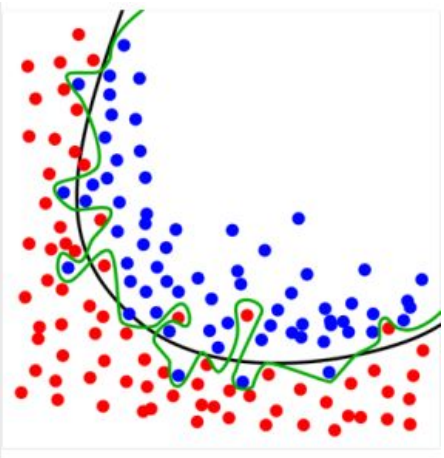
2. 从正则化处理过拟合的方向，来提出改进的机器学习算法

2.1 过拟合是什么?

过拟合（overfitting）是指在模型参数拟合过程中的问题，由于训练数据包含**抽样误差**，训练时，复杂的模型将抽样误差也考虑在内，将抽样误差也进行了很好的拟合。

具体表现就是最终模型在训练集上效果好；在测试集上效果差。模型泛化能力弱。





2.2 为什么要去解决过拟合问题?

emmmmm，我觉得这是个好问题！这是因为我们拟合的模型一般是用来预测未知的结果（不在训练集内），过拟合虽然在训练集上效果好，但是在实际使用时（测试集）效果差。同时，在很多问题上，我们无法穷尽所有状态，不可能将所有情况都包含在训练集上。所以，必须要解决过拟合问题。

> 继续提问！那为什么这个问题常出现在机器学习中呢?

这是因为机器学习算法为了满足尽可能复杂的任务，其模型的拟合能力一般远远高于问题复杂度，也就是说，机器学习算法有「拟合出正确规则的前提下，进一步拟合噪声」的能力。

而传统的函数拟合问题（如机器人系统辨识），一般都是通过经验、物理、数学等推导出一个含参模型，模型复杂度确定了，只需要调整个别参数即可。模型「无多余能力」拟合噪声。

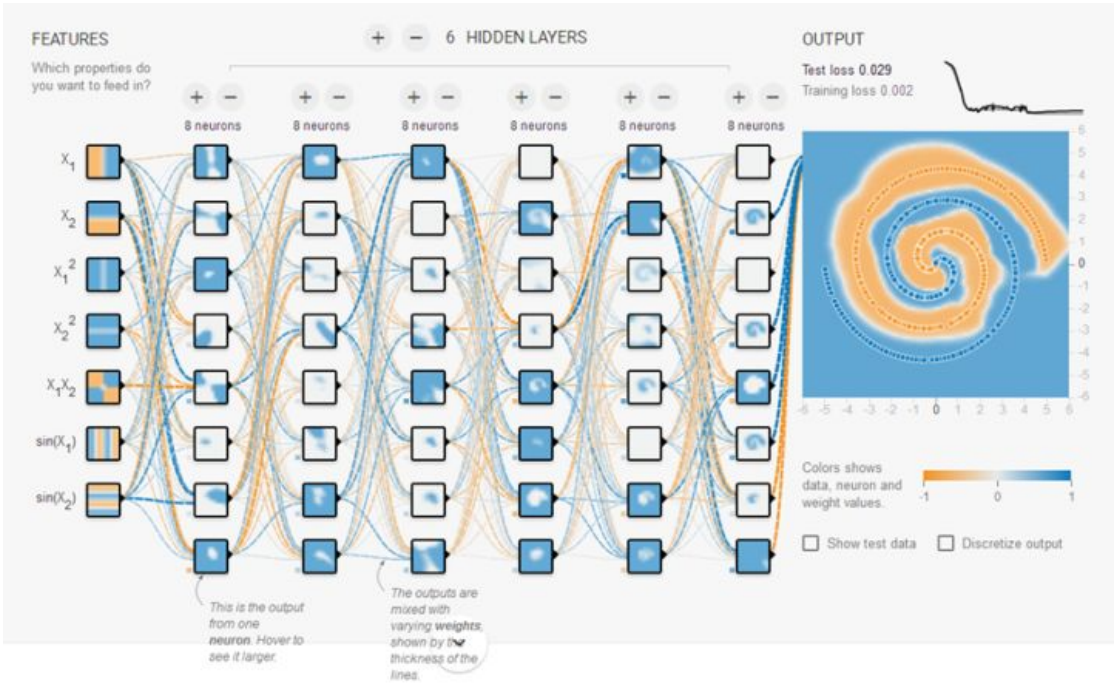
3. 那我们怎么“防止”过拟合?

防止过拟合？防止是不可能防止的，这辈子都不可能防止的，只要把模型复杂度减小一点，多增加训练集个数才能生活这样子。以神经网络为例：

上图是，过拟合的效果...输出的结果完全是没法看的...

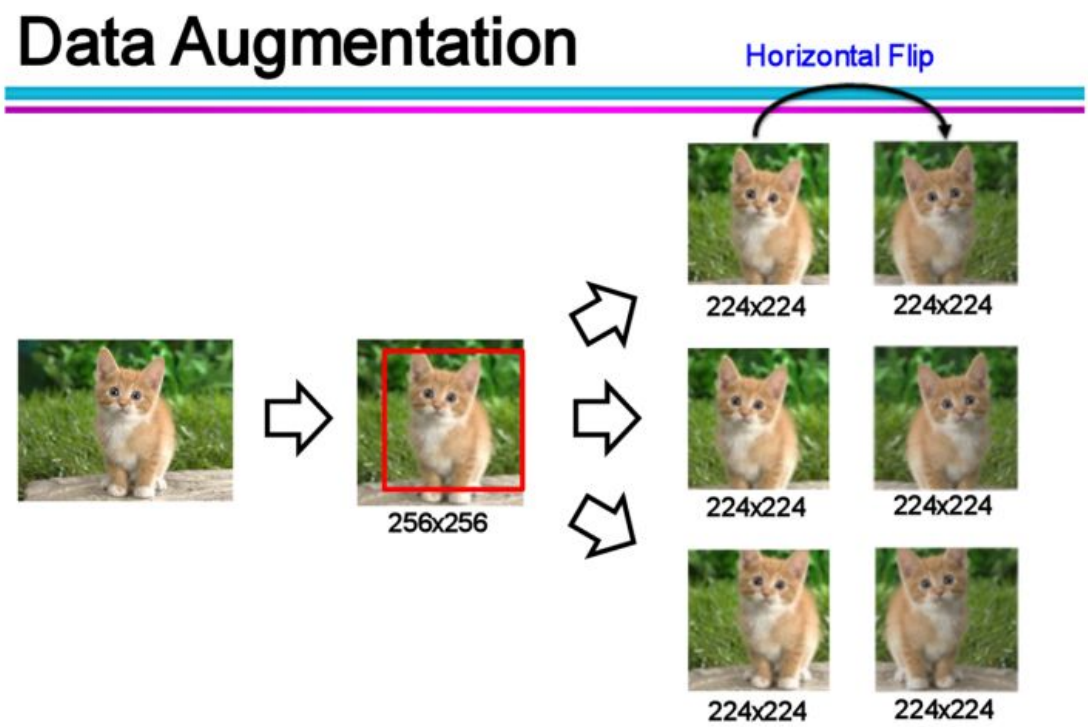
3.1 增加训练集个数

这是解决过拟合最有效的方法，只要给足够多的数据，让模型「看见」尽可能多的「例外情况」，它就会不断修正自己，从而得到更好的结果：



如何获取更多数据，可以有以下几个方法：

1. **从数据源头获取数据**：这个是容易想到的，例如物体分类，我就再多拍几张照片好了。但是，在很多情况下，大幅增加数据本身就不容易；另外，我们不清楚获取多少数据才算够。
2. **根据当前数据集估计数据分布参数，使用该分布产生更多数据**：这个一般不用，因为估计分布参数的过程也会代入抽样误差。
3. **数据增强（Data Augmentation）**：：通过一定规则扩充数据。如在物体分类问题里，物体在图像中的位置、姿态、尺度，整体图片明暗度等都不会影响分类结果。我们就可以通过图像平移、翻转、缩放、切割等手段将数据库成倍扩充；



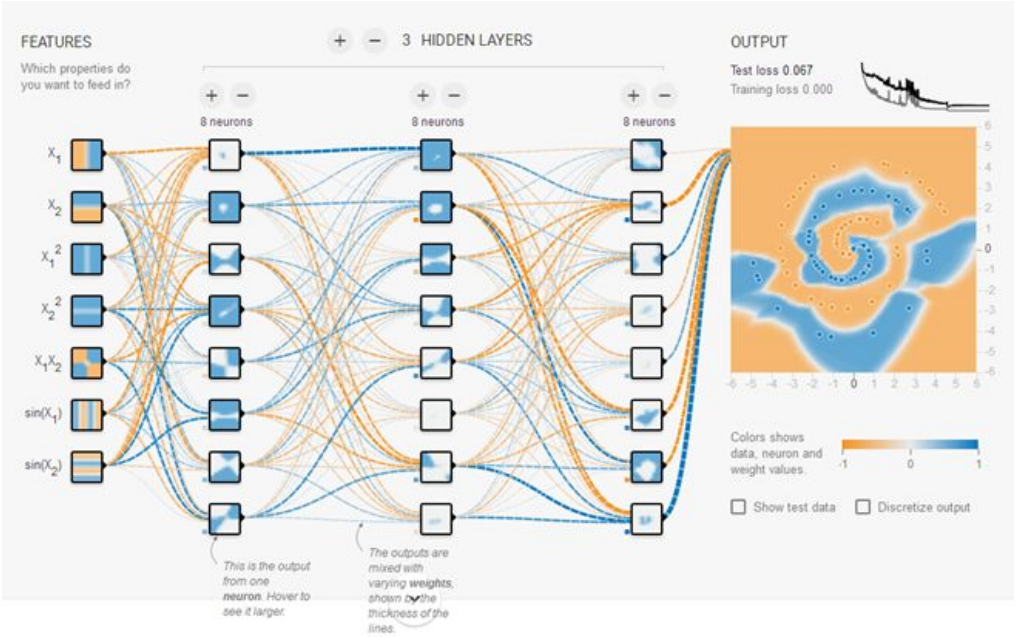
上图就是数据增强的效果。

3.2 使用合适复杂度的模型

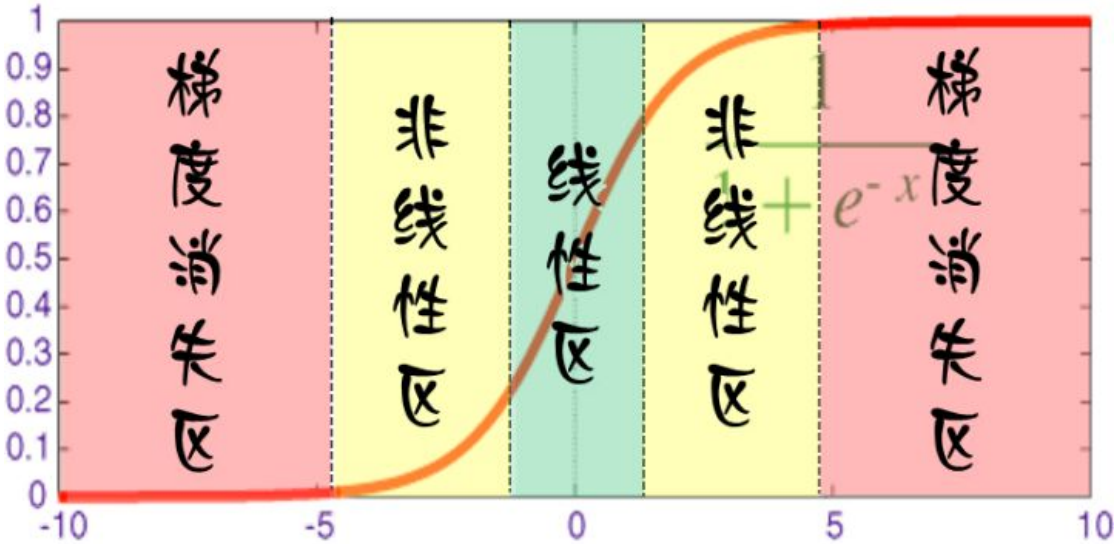
前面说了，过拟合主要是有两个原因造成的：数据太少+模型太复杂。所以，我们可以通过使用合适复杂度的模型来防止过拟合问题，让其足够拟合真正的规则，同时又不至于拟合太多抽样误差。

对于神经网络而言，我们可以从以下四个方面来：

1. 网络结构 Architecture
   这个很好理解，减少网络的层数、神经元个数等均可以限制网络的拟合能力；



2. 训练时间 Early stopping
   对于每个神经元而言，其激活函数在不同区间的性能是不同的：



   当网络权值较小时，神经元的激活函数工作在线性区，此时神经元的拟合能力较弱（类似线性神经元）。
   有了上述共识之后，我们就可以解释为什么限制训练时间（early stopping）有用：因为我们在初始化网络的时候一般都是初始为较小的权值。训练时间越长，部分网络权值可能越大。如果我们在合适时间停止训练，就可以将网络的能力限制在一定范围内。
3. 限制权值 Weight-decay，也叫正则化（regularization）
   原理同上，但是这类方法直接将权值的大小加入到 Cost 里，在训练的时候限制权值变大。以 L2 regularization为例：

$$C = C_0 + \frac{\lambda}{2n} \cdot \sum_i w_i^2$$

   训练过程需要降低整体的 Cost，这时候，一方面能降低实际输出与样本之间的误差 C_0，也能降低权值大小。
4. 在权值上加噪声

output on one case → $y^{noisy} = \sum_i w_i x_i + \sum_i w_i \varepsilon_i$ where ε_i is sampled from $N(0, \sigma_i^2)$

$$E\left[(y^{noisy} - t)^2\right] = E\left[\left(y + \sum_i w_i \varepsilon_i - t\right)^2\right] = E\left[\left((y-t) + \sum_i w_i \varepsilon_i\right)^2\right]$$

$$= (y-t)^2 + E\left[2(y-t)\sum_i w_i \varepsilon_i\right] + E\left[\left(\sum_i w_i \varepsilon_i\right)^2\right]$$

$$= (y-t)^2 + E\left[\sum_i w_i^2 \varepsilon_i^2\right]$$

because ε_i is independent of ε_j and ε_i is independent of $(y-t)$

$$= (y-t)^2 + \sum_i w_i^2 \sigma_i^2$$

So σ_i^2 is equivalent to an L2 penalty

在输入中加高斯噪声，会在输出中生成 $\sum_i \sigma_i^2 \cdot w_i^2$ 的干扰项。训练时，减小误差，同时也会对噪声产生的干扰项进行惩罚，达到减小权值的平方的目的，达到与 L2 regularization 类似的效果（对比公式）。

结合多种模型

简而言之，训练多个模型，以每个模型的平均输出作为结果。

从 N 个模型里随机选择一个作为输出的期望误差 $< [t - y_i]^2 >$，会比所有模型的平均输出的误差 $< [t - \overline{y}]^2 >$ 大

$$\overline{y} = < y_i >_i = \frac{1}{N}\sum_{i=1}^{N} y_i$$

i is an index over the N models

$$< (t - y_i)^2 >_i = < \left((t - \overline{y}) - (y_i - \overline{y})\right)^2 >_i$$

$$= < (t - \overline{y})^2 + (y_i - \overline{y})^2 - 2(t - \overline{y})(y_i - \overline{y}) >_i$$

this term vanishes

$$= (t - \overline{y})^2 + < (y_i - \overline{y})^2 >_i \boxed{-2(t - \overline{y}) < (y_i - \overline{y}) >_i}$$

感受：

这个题目“**试叙述迭代法思想或正则化思想的优点，并尝试在此基础上提出一种新的机器学习方法（最好能写到模型部分**”，我觉得蛮有难度的，所以只是从正则化思想解决过拟合这个方向，来阐述了一下优化的机器学习算法，要我提出一个新的算法实在太难。我认为人工智能是十分有趣的，以后可以继续边听刘老师的课，边实践做一些科研，帮助自己理解算法。谢谢刘胜蓝老师一学期来的辛苦教学工作～!