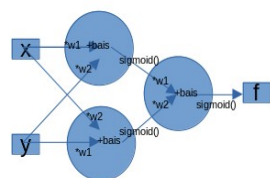


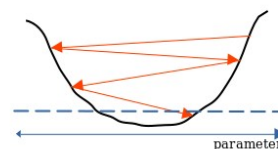
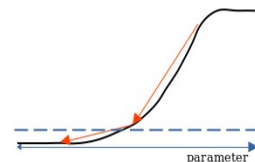
受到视频作者 Tsoding 的启发，使用三个线性神经元组成网络，并进行对逻辑



逻辑 1	$11 > 1, 10 > 0, 01 > 0, 00 > 0$
逻辑 2	$11 > 1, 10 > 1, 01 > 0, 00 > 1$
逻辑 3	$11 > 1, 10 > 0, 01 > 1, 00 > 0$

进行迭代。为了探究最佳学习率（RATE），现准备相同的神经网络，各神经元参数的权重初始值为 0，偏差为 10，将微分的极小量设为 $1e-3$ ，并将终止条件设置为模型与对应逻辑模型的损失函数小于 $1e-5$ 。在多次迭代中学习率（RATE）与所需迭代次数的相关关系如下图所示。

学习率决定了每次梯度下降的“步幅”，在一定范围内，更大的步幅有助于更早到达最低点。如果损失函数值与某一参数的关系呈现单调的关系，较大的学习率并无害处。但如果损失函数值对于某一参数的关系呈现



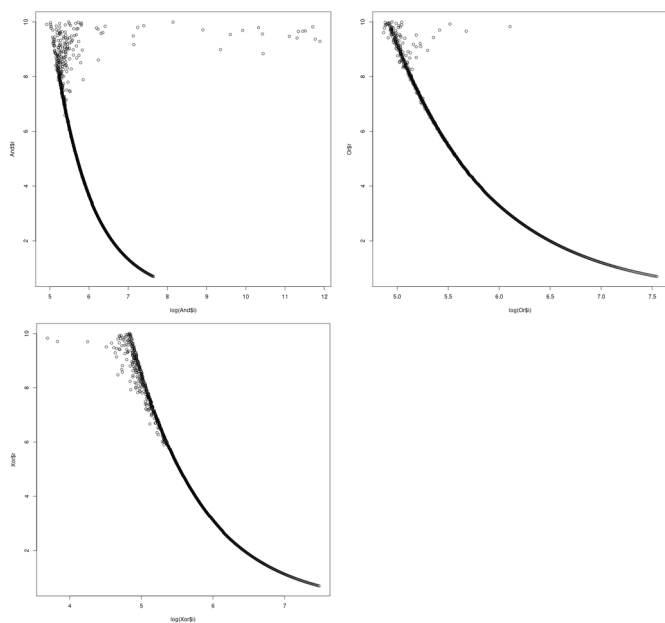
中间低、两边高的“山谷”形状，在学习率过大时，梯度下降有一定几率会跳过最小值，并让模型在最小值两侧弹跳，一定程度造成迭代次数的浪费，有时甚至会出现反复在两点间跳跃，无法达到最小值的困境。

除此之外，由于 double 的精度误差，有时会因为遇到较为平缓的点而在微分时得到 0，从而造成 Nan 或者 inf 使得参数完全失效。

回顾神经网络在学习率较大时对三种逻辑的表现，注意到以下规律：

- 网络对逻辑 1 拟合时，较大学习率造成的异常数据几乎总是表示较大的迭代次数，这可能表示在网络对逻辑 1 进行拟合时，较多的参数对与损失函数最小值有类似“山谷”的关系。
- 网络对逻辑 2 拟合时，较大的学习率造成了较少的异常点，这些异常点同时拥有大于或小于正常迭代次数的表现。说明各有相当数量的参数分别表现为上文所述的两种状态。
- 网络对逻辑 3 拟合时，较大的学习率几乎总是造成了较少迭代次数的异常点，表明较多的参数与损失函数值有单调关系。

衡量学习率的高效性的方法还有在指定迭代次数下损失函数值的大小，以及对下降中驻点的避让能力的比较，不过这需要更为复杂的程序进行评估，这为后续的探求提供了方向。



对比三幅图可知，当学习率小于某一接近 6 的值时，网络对三种模型的拟合有相似的表现，均呈相似的反比例关系，特别地，在学习率为 0.7 时，迭代次数均接近 10 万次，这表明目前各参数所设置的初始值使网络对三种逻辑的拟合过程具有可比性。当学习率大于某一值时，模型对三种逻辑的拟合表现互不相同，现分析其原因。