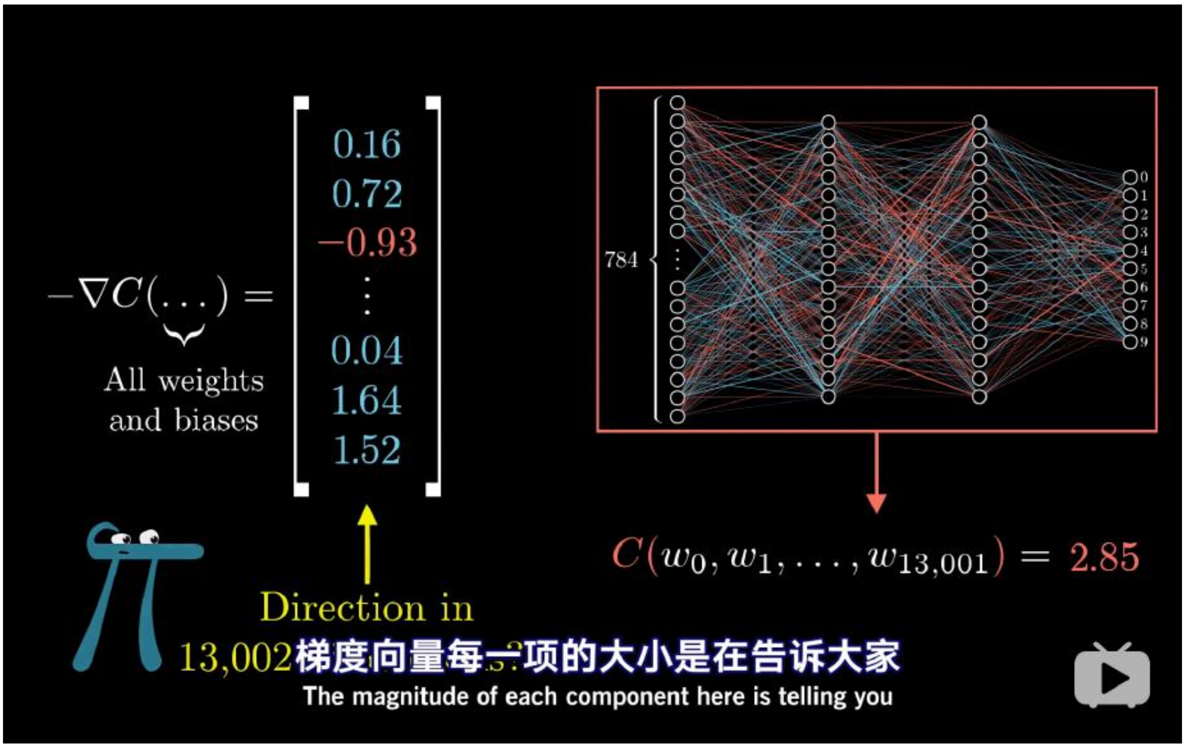
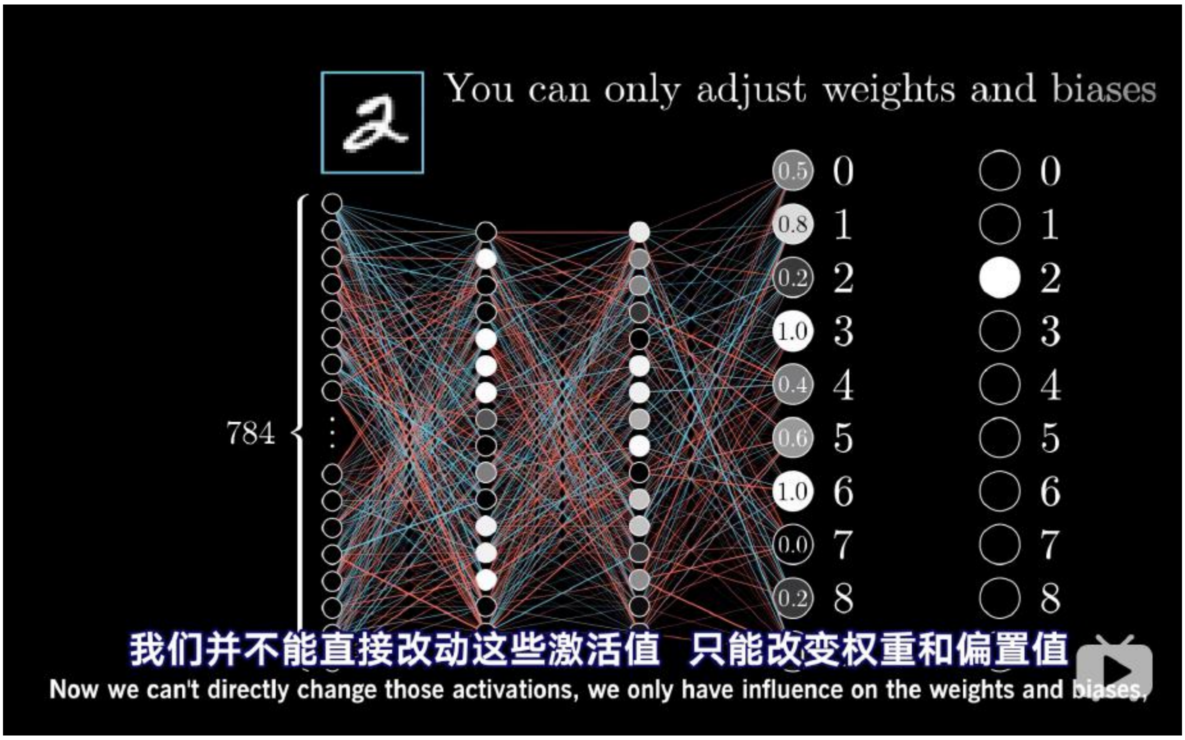
深度学习入门笔记

反向传播算法

梯度向量的含义：

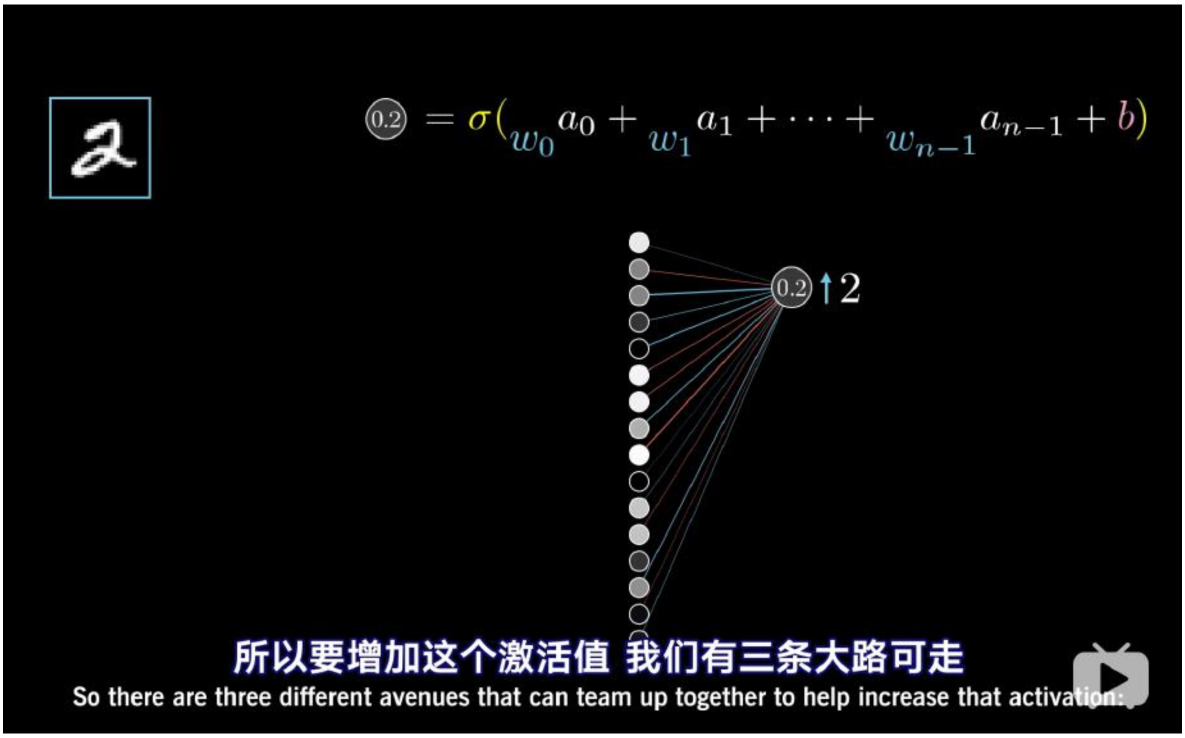


对于单个训练数据：

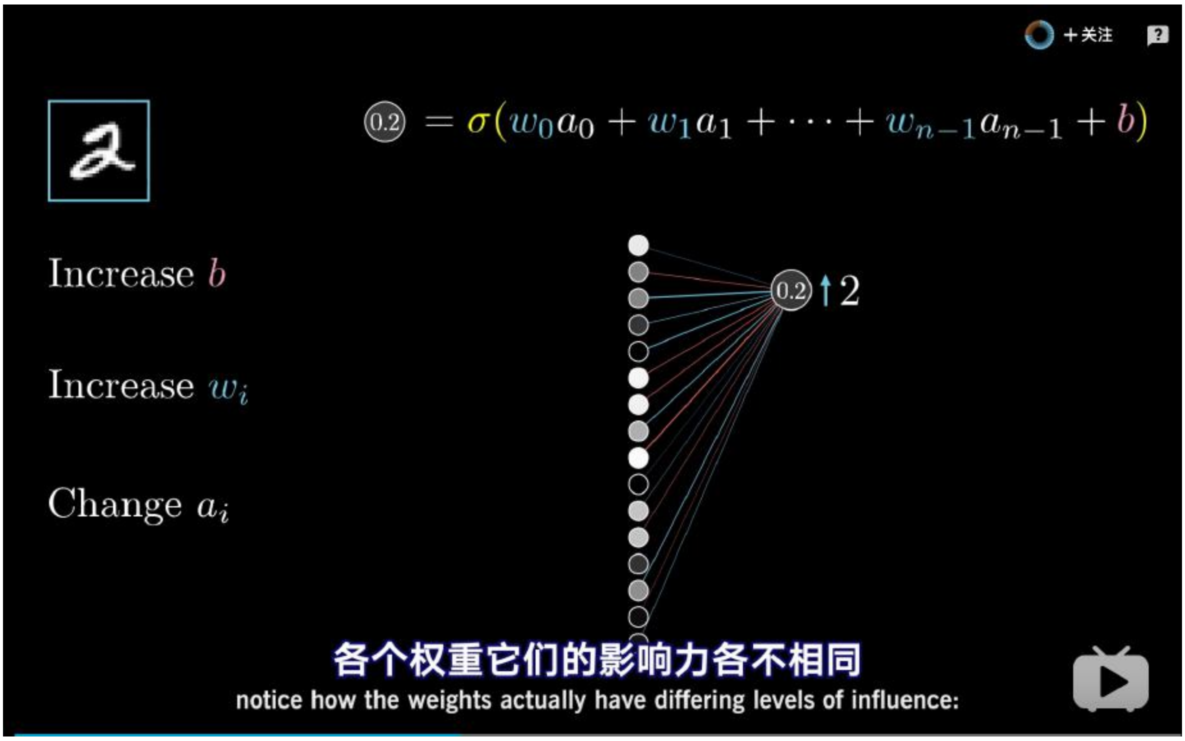


下一层的activation受到三个方面的值的影响：

- 与上一层连线的权重 weight
- 上一层的激活值 activation
- 当前节点的偏置值 bias



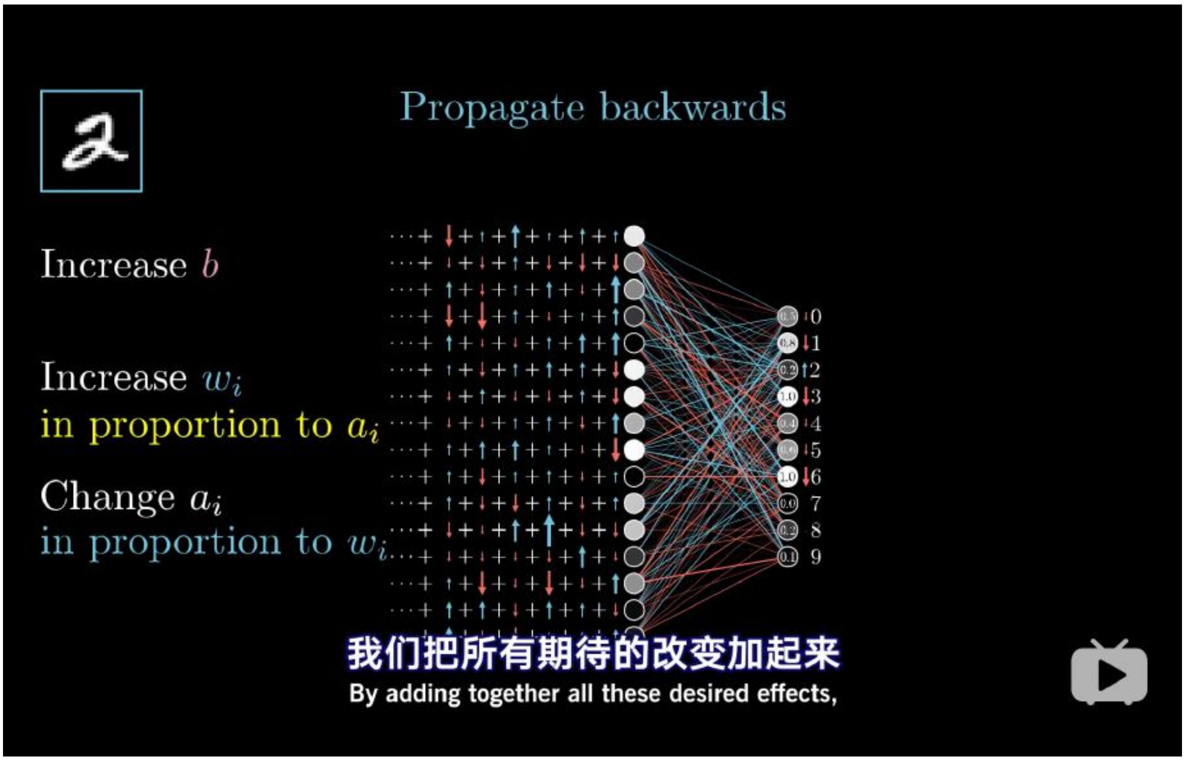
当调整权重 weight 的时候：

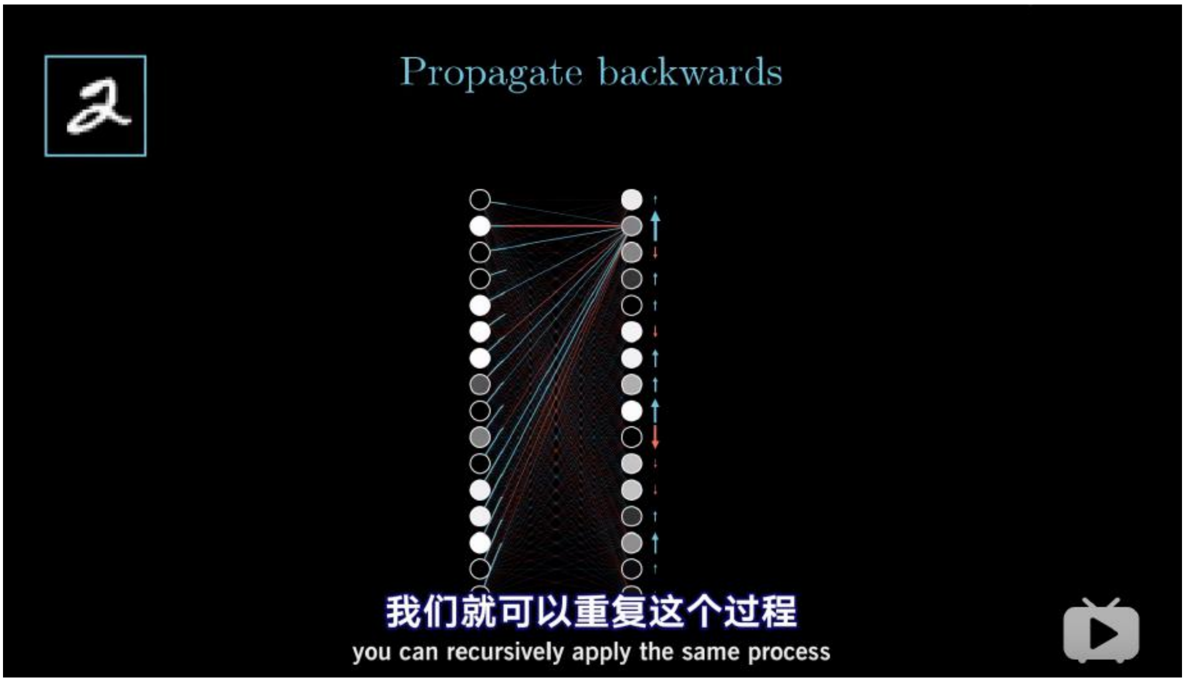


当调整上一层的激活值 activation 的时候：

（ps：但是记住，我们只能改变 weight 和 bias）

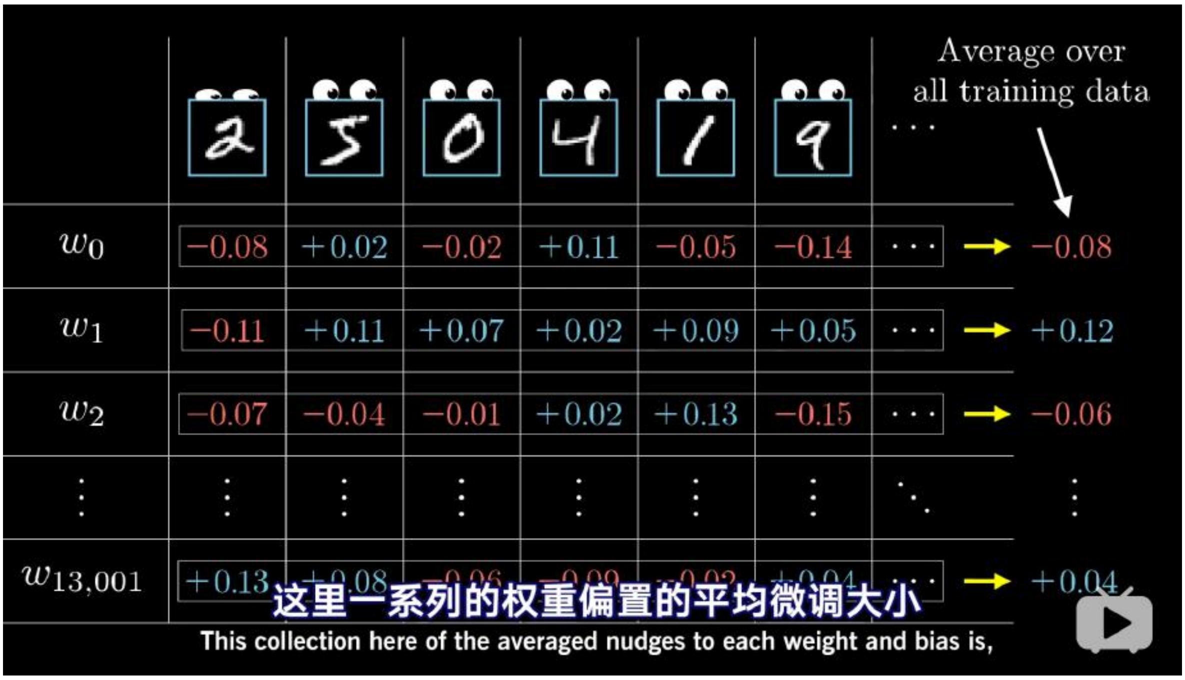
将最后一层的所有的神经元各自提出的变化量相加：



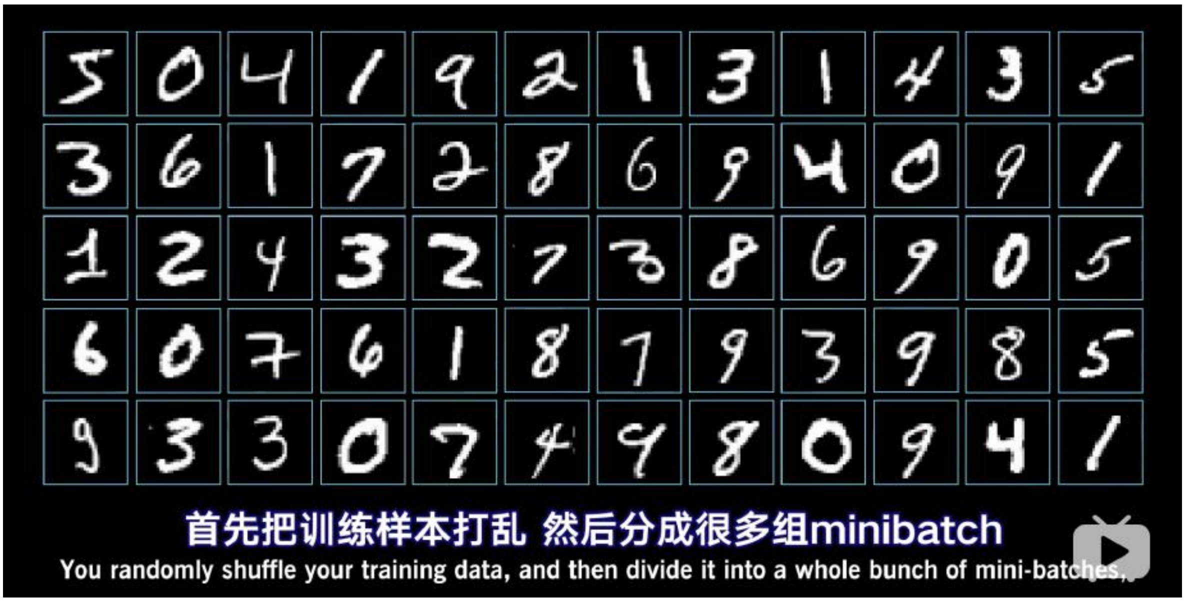


这其实就是反向传播算法(back propagation-BP)的主要思想，到此只考虑了一个训练样本“2”的反向传播对权重和偏置的影响

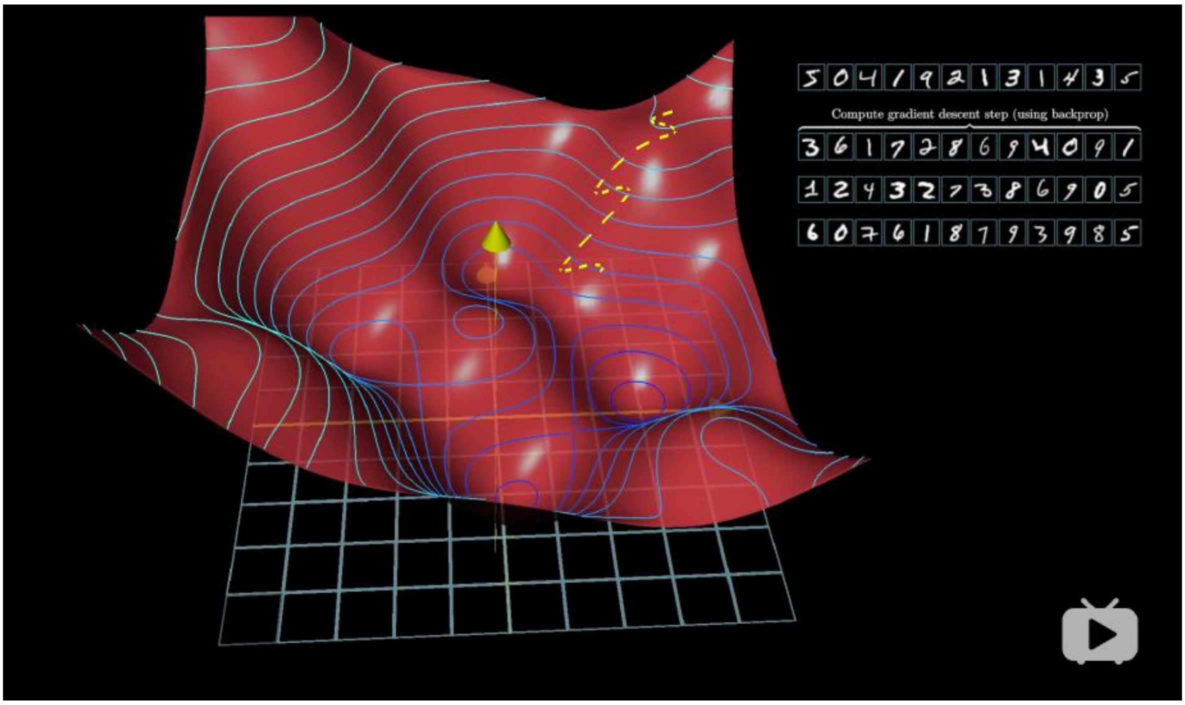
当我们考虑多个样本时：



实际的训练过程中，为了减少计算量。通常不会在每一次的梯度下降之前将所有的样本都计算一遍，而是将样本集划分成一个个字集 minibatch



每次梯度下降都计算一个 mini-batch 得到梯度下降向量，这样计算量会下降不少



这样梯度下降的过程就是曲折的，而非最快的。但最终还是能到达目标函数的局部最优值。这叫做“随机梯度下降(stochastic gradient descent)”

总结By RandomYane https://weibo.com/3229623314