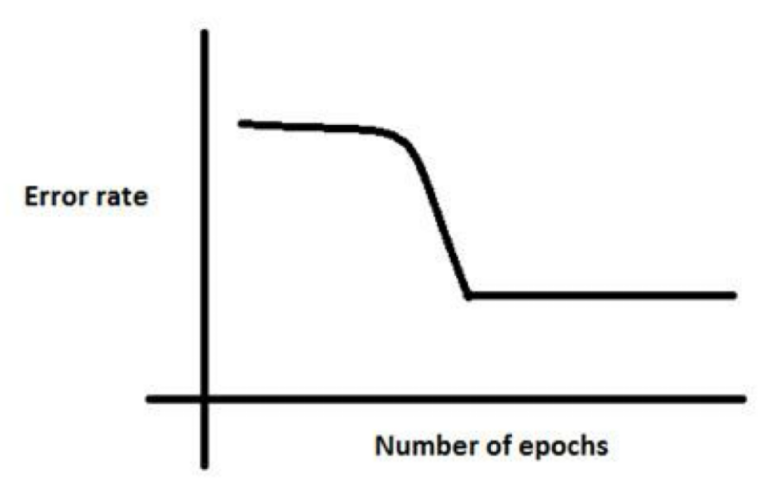
**ML\_神经网络优化方法**

**选择题&简答题**

1. 在训练神经网络时，损失函数（Loss）在最初的几个 epochs时没有下降，可能的原因是？



1. 批规范化（Batch Normalization）的好处有？

a. 让每一层的输入的范围都大致固定

b. 它将权重的归一化平均值和标准差

c. 它是一种非常有效的反向传播（BP）方法

d. 这些均不是

1. 在一个神经网络中，下面哪种方法可以用来处理过拟合？

a. Dropout

b. Batch Normalization

c. 正则化

d. 以上都可以

1. 考虑某个具体问题时，你可能只有少量数据来解决这个问题。不过幸运的是你有一个类似问题已经预先训练好的神经网络。可以用下面哪种方法来利用这个预先训练好的网络？

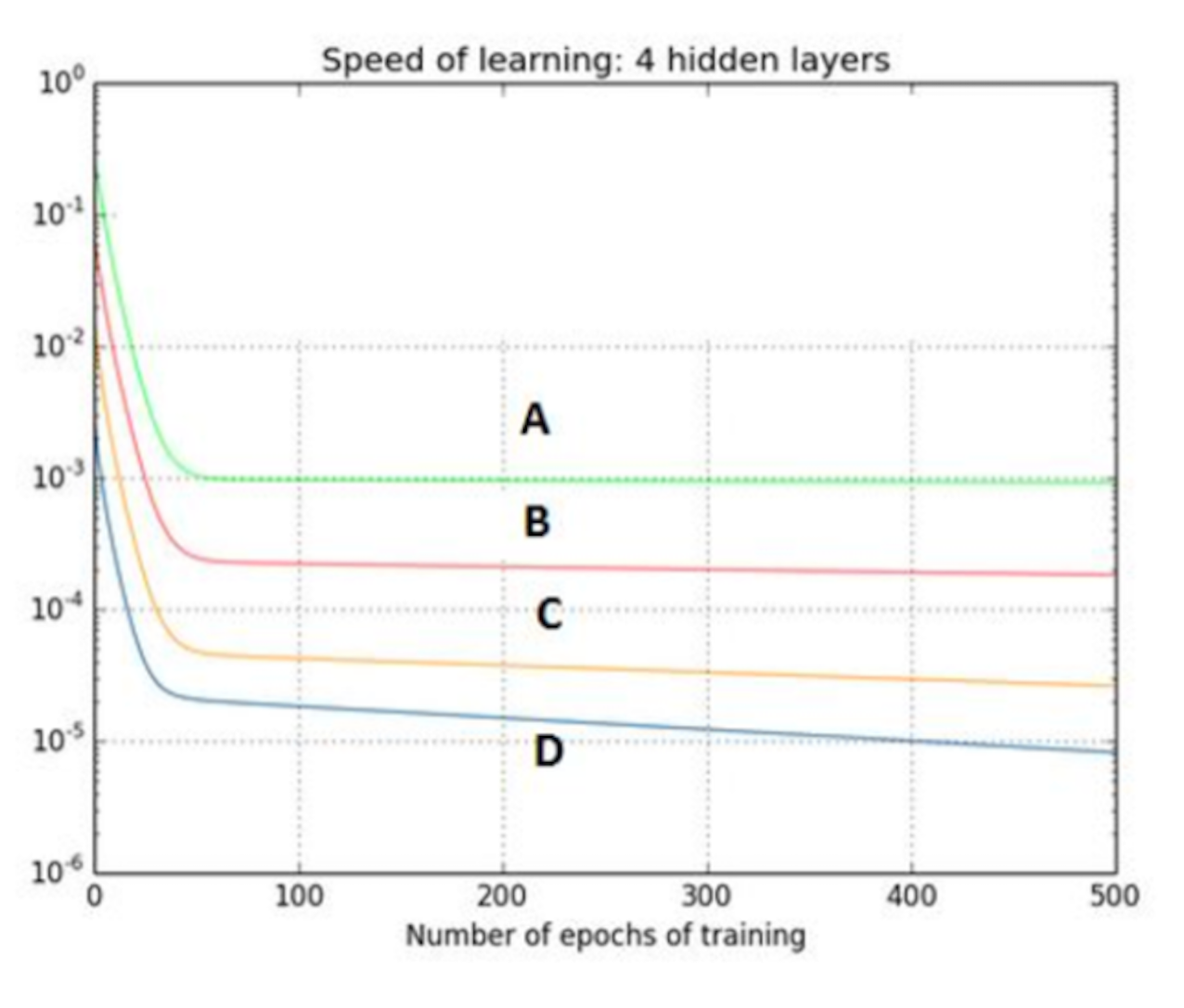
a. 把除了最后一层外所有的层都冻结，重新训练最后一层

b. 对新数据重新训练整个模型

c. 只对最后几层进行调参（fine tune）

d. 对每一层模型进行评估，选择其中的少数来使用

1. 下图是一个利用sigmoid函数作为激活函数的含四个隐藏层的神经网络训练的梯度下降图。这个神经网络遇到了梯度消失的问题。下面哪个叙述是正确的（纵坐标表示梯度大小）？



a. 第一隐藏层对应D，第二隐藏层对应C，第三隐藏层对应B，第四隐藏层对应A

b. 第一隐藏层对应A，第二隐藏层对应C，第三隐藏层对应B，第四隐藏层对应D

c. 第一隐藏层对应A，第二隐藏层对应B，第三隐藏层对应C，第四隐藏层对应D

d. 第一隐藏层对应B，第二隐藏层对应D，第三隐藏层对应C，第四隐藏层对应A

1. 对于一个分类任务，如果开始时神经网络的权重不是随机赋值的，而是都设成0，下面哪个叙述是正确的？

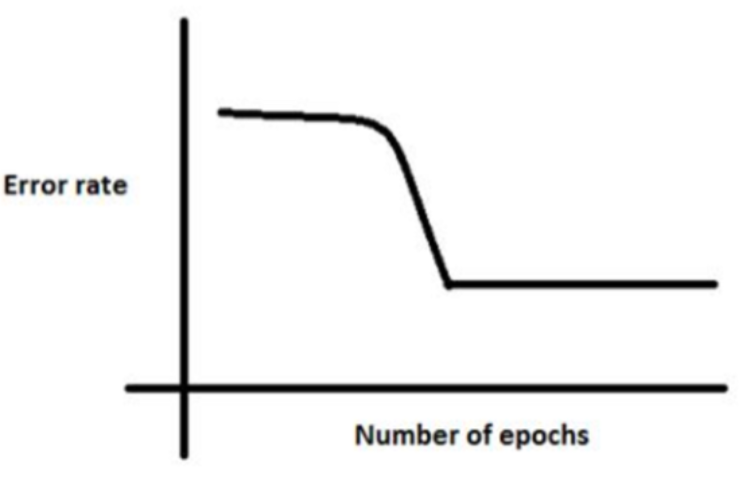
a. 其他选项都不对

b. 没啥问题，神经网络会正常开始训练

c. 神经网络可以训练，但是所有的神经元最后都会变成识别同样的东西

d. 神经网络不会开始训练，因为没有梯度改变

1. 下图显示，当开始训练时，误差一直很高，这是因为神经网络在往全局最小值前进之前一直被卡在局部最小值里。为了避免这种情况，我们可以采取下面哪种策略？



a. 改变学习速率，比如一开始的几个训练周期不断更改学习速率

b. 一开始将学习速率减小10倍，然后用动量项 （momentum）

c. 增加参数数目，这样神经网络就不会卡在局部最优处

d. 其他都不对

1. 下列的哪种方法可以用来降低深度学习模型的过拟合问题？ 1.增加更多的数据 2.使用数据扩增技术（data augmentation） 3.使用归纳性更好的架构 4.正则化数据 5.降低架构的复杂度

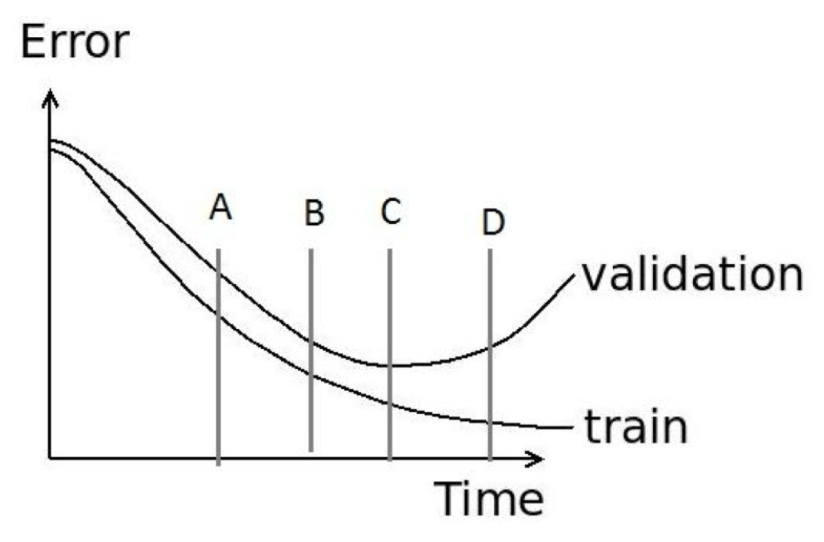
a. 145

b. 123

c. 1345

d. 都可以

1. 当训练一个神经网络来作图像识别任务时，通常会绘制一张训练集误差和交叉训练集误差图来进行调试。在图中，最好在哪个时间停止训练？



1. Xavier初始化是最为常用的神经网络权重初始化方法，下图是初始化的公式

Xavier初始化是用来帮助信号能够在神经网络中传递得更深，下面哪些叙述是对的？ 1 如果权重一开始很小，信号到达最后也会很小 2 如果权重一开始很大，信号到达最后也会很大 3 Xavier初始化是由高斯分布引出的 4 Xavier初始化可以帮助减少梯度弥散问题

a. 234

b. 1234

c. 124

d. 134

1. 声明1：可以通过将所有权重初始化为0来训练网络

声明2：可以通过将偏差初始化为0来很好地训练网络。

以上哪些陈述是真实的？

a. 1

b. 2

c. 12

d. 都是错的

1. 如果我们希望预测n个类（p1,p2,…,pk）的概率使得所有n个p的和等于1，那么下列哪个函数可以用作输出层中的激活函数？

a. Softmax

b. Relu

c. Sigmoid

d. Tanh

1. 使用批量归一化可以解决神经网络训练中的哪些问题？

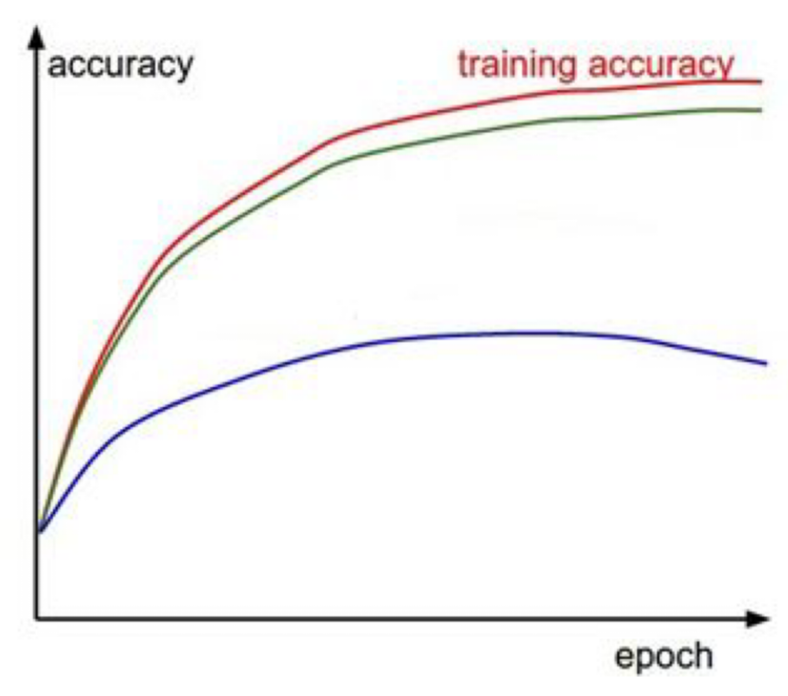
a. 过拟合

b. 限制输出过大或过小

c. 训练过慢

d. 以上所有

1. 上面的红色曲线表示关于深度学习算法中每个时期的训练精度。绿色和蓝色曲线都表示验证的准确性。 哪条曲线表示过拟合overfitting？



a. 绿色曲线

b. 蓝色曲线

1. 我们可以采取哪些措施来防止神经网络中的过拟合？

a. 数据增强

b. 权重共享

c. 提前停止

d. Dropout

e. 以上全部

1. 以下模型中，在数据预处理时，不需要考虑归一化处理的是

a. logistic回归

b. SVM

c. 树形模型

d. 神经网络

1. 假设用 表示一个来自小批量 的输入，批量归一化 层的参数为：拉伸参数（scale） 和偏移参数(shift) ，它们的形状与 相同。请写出经过批量归一化转换后的 表达式 。
2. 批归一化和Dropout一般不会同时使用，请简述原因。
3. 简述Dropout能够防止过拟合的原因。
4. 假设Dropout概率为 ，为了保证期望值不变，则在测试时，该层模型权重 应该变为？
5. 设模型初始参数 为 ，学习率为0.1，动量更新权重为1，若模型在两次训练过程中的梯度依次为 ，求两次训练后模型的参数 .