1. 验证z = sum(wx)+b标准差为sqrt(3/2)，假设Var(w)=1/1000, n=500。

提示：权重初始化目的是为了让每一层的输出z(神经元前的那一项)和该层输入 x 保持一致的数据分布。权重初始化的方法是对权重的数据除以输入权重神经元个数的根号，注意 w 和 x 和 b 三个变量相互独立。

1. 什么是训练集、验证集、测试集？在机器学习中为什么需要将样本分成独立的三部分：训练集（train set），验证集（validation set ) 和测试集（test set）？

训练集：训练集是机器学习模型用于训练和学习的数据集。通常情况下，训练集是原始数据集的一部分，用于训练模型的参数。模型通过训练集来学习数据的特征，并产生一个模型，以便在之后的预测中使用。

验证集：验证集是用于评估模型性能的数据集。它通常是从原始数据集中划分出来的，用于在训练过程中调整模型的参数和超参数，以提高模型的性能。验证集的作用是帮助开发人员调整模型，避免模型过拟合或欠拟合。

测试集：测试集是用于评估模型最终性能的数据集。它通常是从原始数据集中划分出来的，与训练集和验证集互不重叠。测试集的作用是评估模型在未见过的数据上的性能，并判断模型是否足够准确和鲁棒。

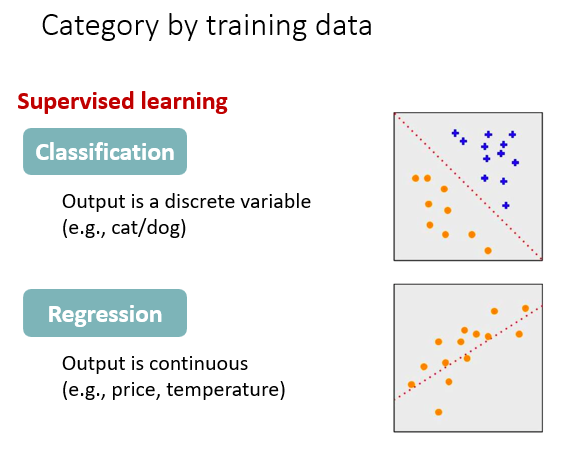
为什么这么分？

训练集：这是最容易理解的部分，即用来训练模型参数的那部分数据。这里就不过多解释了。

测试集：用来验证模型效果的数据，用来评价模型学得好还是不好。

验证集：用来在训练阶段验证模型训练效果，用于选择超参数，而不用测试集验证，避免作弊风险。

1. 什么是分类和回归？两者的联系和区别分别是什么？可结合图详细说明。



1. 为什么对于一个神经元，并使用梯度下降优化参数w时，如果输入x恒大于0，其收敛速度会比零均值化的输入更慢？

图示, 文本

描述已自动生成

1. 什么是梯度消失和梯度爆炸？详细说明梯度消失问题是否可以通过增加学习率来缓解？

梯度反传时，网络层之间的梯度（值大于 1.0）重复相乘导致的指数级增长会产生梯度爆炸。

梯度反传时，网络层之间的梯度（值小于 1.0）重复相乘导致的指数级减小会产生梯度消失。

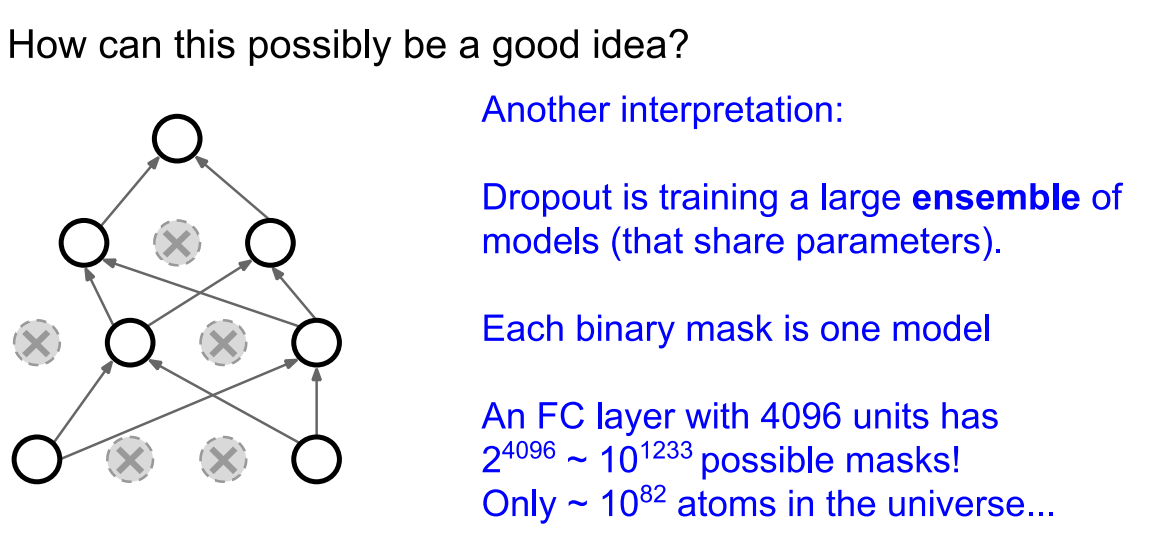
梯度消失问题不可以通过增加学习率来缓解，因为梯度消失不是指梯度太小，而是指梯度在最后一层往前传的过程中不断减小，这个消失，指的是最底层的梯度和最高层的梯度量级的差别。

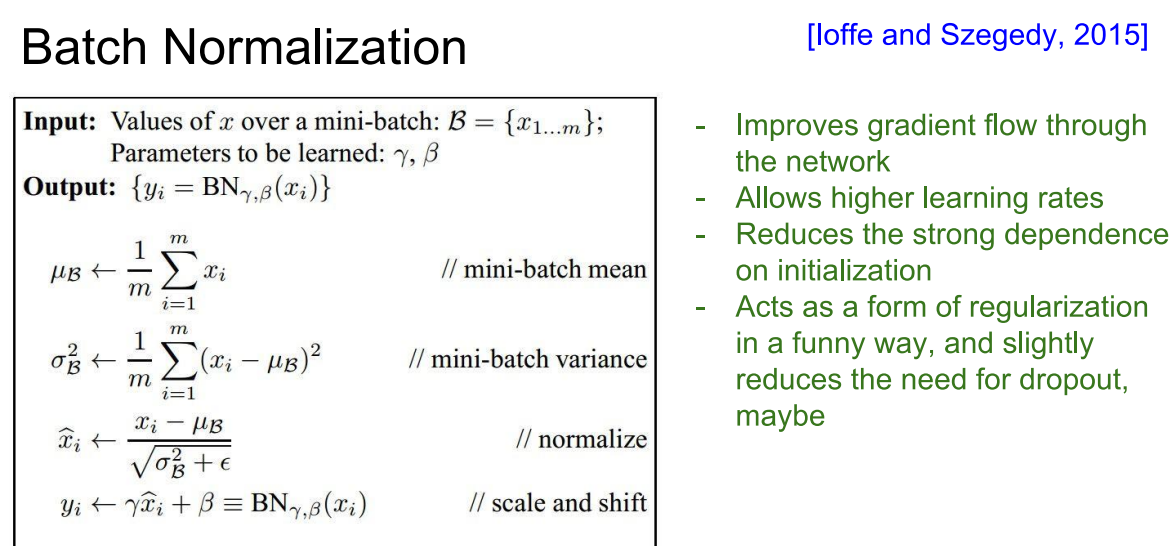
1. 什么是过拟合？Dropout为什么有助于防止过拟合？请叙述批归一化(Batch Normalization)的4个步骤。

过拟合是模型在训练的时候效果好，但是测试的时候效果差。说明模型泛化性能差。

图表

中度可信度描述已自动生成

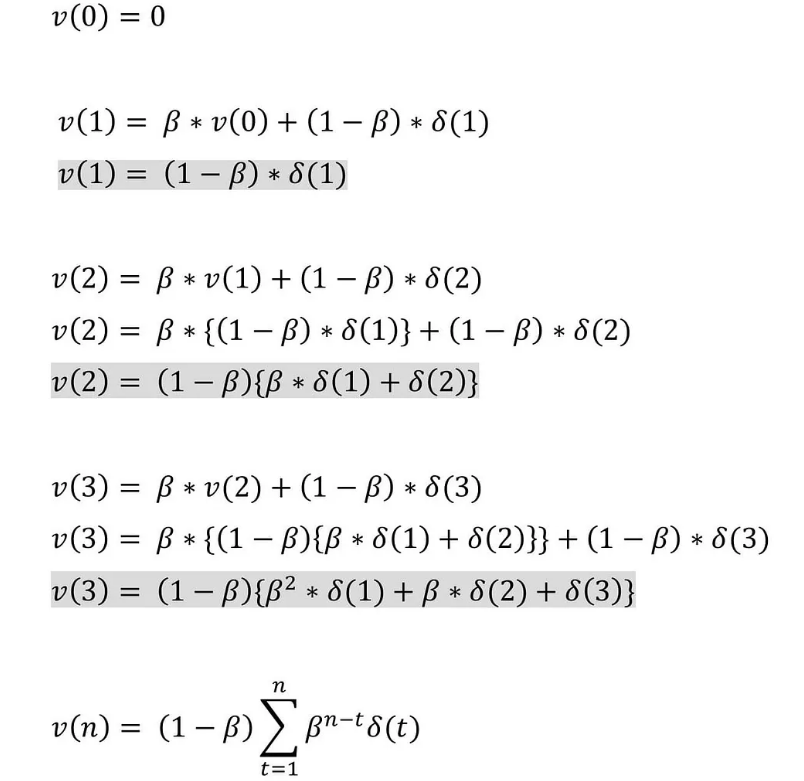




1. 梯度下降优化算法SGD with momentum中，参数更新时通过积累之前的动量来加速当前的梯度，其中一阶动量使用指数加权平均来计算：

一阶动量是各个时刻梯度方向的指数加权平均数，约等于最近个时刻的梯度向量和的平均值。请简要推导为什么是最近个时刻？SGD with momentum与标准SGD相比，它的优点和原理是什么？（可结合图论述）

假设v是momentum；是梯度。



次数越高，该项越接近0。

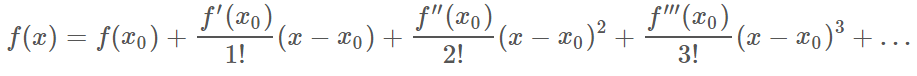
因此，一阶动量是各个时刻梯度方向的指数加权平均数，约等于最近个时刻的梯度向量和的平均值。

SGD with momentum与标准SGD相比，它的优点是可能逃离局部最优点。由于SGD只考虑当前时刻的梯度，在局部最优点的当前梯度为0。由计算公式可知，此时参数不再进行更新，故陷入局部最优的状态。

1. 梯度的定义：某一函数沿着某点处的方向导数可以以最快速度达到极大值，该方向导数定义为该函数的梯度

其中是自变量，是关于的函数，表示梯度，推导以下公式：

一元函数的泰勒展开大家应该都知道，公式如下：



不妨只取右边式子的前两项，也就是一个“约等于”的结果：

图表, 箱线图

描述已自动生成

文本

中度可信度描述已自动生成手表上有字

中度可信度描述已自动生成

我们的目标是在迭代过程中让下式恒成立，也就是说希望迭代过程中函数值会逐渐减小，用数学语言描述就是：容易想到，应该构造：此时：文本

描述已自动生成写成迭代形式：文本

描述已自动生成由于上式最右项恒小于等于0，我们就完成了对于梯度下降有效性的证明。从上述步骤归纳出来的参数迭代更新的公式如下：