本课程介绍了半监督学习以及深度学习和神经网络的初步介绍内容。

少数的标注数据使用过于复杂的模型很容易过拟合，大量的数据是没有标签的，因此，生活中的大部分问题是半监督学习，主要分为直推式(transductive learning)和纯粹半监督形式(inductive learning)。

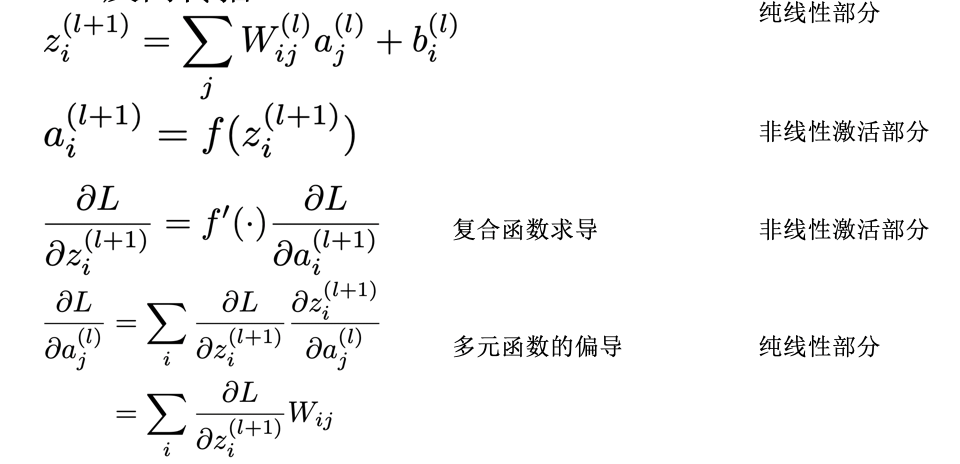
半监督学习也可以分为自学习和系统学习。自学习的算法过程是通过打上标签的数据进行训练，然后对无标签数据进行预测，对最可信的预测结果对应的无标签数据打上标签，然后继续对扩展后的有标签数据进行训练，如此循环往复。

系统学习的算法过程是对每个实例切分成两个view，对每个view都进行训练生成标签，拿出最可信的部分给对方view作为训练集合，采用两种模型相互印证。

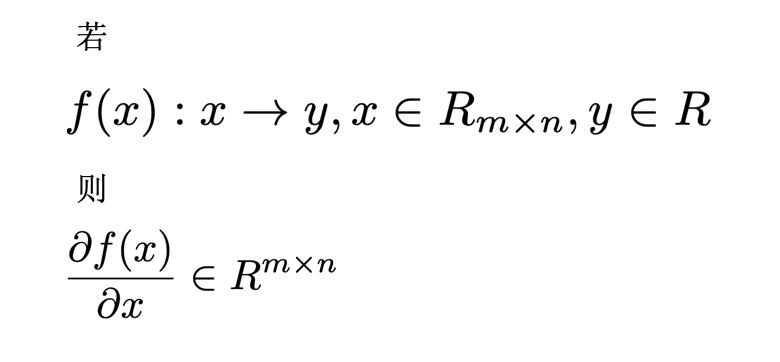
深度学习和神经网络的内容分为最大似然估计、神经网络、神经网络实践分析和实战部分。

最大似然估计的原理若D是离散分布，f即是在参数为theta 时观测到这一采样的概率。若其是连续分布f\_{\theta }}则为X\_{1},X\_{2},\ldots ,X\_{n}}联合分布的概率密度函数在观测值处的取值。一旦我们获得X\_1, X\_2,\ldots, X\_n，我们就能求得一个关于\theta 的估计。最大似然估计会寻找关于\theta 的最可能的值（即，在所有可能的\theta 取值中，寻找一个值使这个采样的“可能性”最大化）。从数学上来说，我们可以在\theta 的所有可能取值中寻找一个值使得似然函数取到最大值。这个使可能性最大的{\theta}^{hat}值即称为\theta 的最大似然估计。由定义，最大似然估计是样本的函数。线性回归的最小二乘法从本质上也是基于残差为高斯分布假设的最大似然估计。

神经网络是通过多层非线性复合实现“无限拟合”性质，主要概念为线性层、激活层和隐层。纯粹的线性层拟合能力是有限的。神经网络的更高层神经元是低层神经元的线性组合再经过非线性激活，非线性层逐点激活，遍历每层并求解loss对每个参数的导数，神经网络的反向传播是通过是通过链式求导实现的。如下图所示：



反向传播是神经网络的核心部分，其每层神经网络的任务是传递loss到本层梯度到上层，计算loss到本层参数的导数。在求导过程中，需要满足维数相容原则，即



本课程最后部分介绍了基于numpy的简单网络实现。