**Week1**

设计一个神经网络时，输入层与输出层的节点数往往是固定的，中间层则可以自由指定；

神经网络结构图中的拓扑与箭头代表着预测过程时数据的流向，跟训练时的数据流有一定的区别；

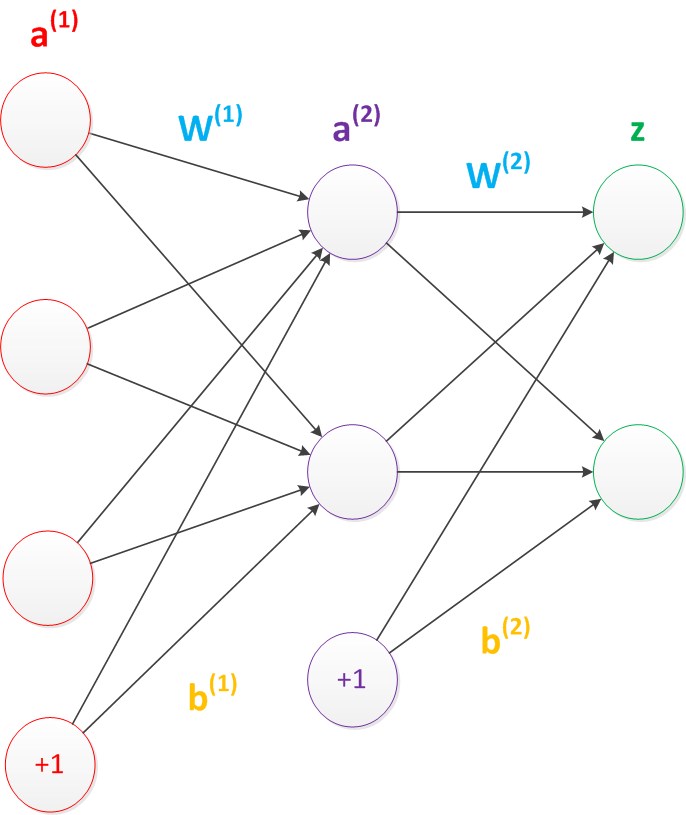
结构图里的关键不是圆圈（代表“神经元”），而是连接线（代表“神经元”之间的连接）。每个连接线对应一个不同的权重（其值称为权值），这是需要训练得到的。

从左到右的表达形式以Andrew Ng和LeCun的文献使用较多，Caffe里使用的则是从下到上的表达。

在“感知器”中，有两个层次。分别是输入层和输出层。输入层里的“输入单元”只负责传输数据，不做计算。输出层里的“输出单元”则需要对前面一层的输入进行计算。我们把需要计算的层次称之为“计算层”，并把拥有一个计算层的网络称之为“单层神经网络”。

w1,2代表后一层的第1个神经元与前一层的第2个神经元的连接的权值

ax(y)代表第y层的第x个节点



偏置单元与后一层的所有节点都有连接，我们设这些参数值为向量b，称之为偏置，其没有输入。

在两层神经网络中，我们不再使用sgn函数作为函数g，而是使用平滑函数sigmoid作为函数g。我们把函数g也称作激活函数（active function）。

从输入层到隐藏层时，数据发生了空间变换。也就是说，两层神经网络中，隐藏层对原始的数据进行了一个空间变换，使其可以被线性分类，然后输出层的决策分界划出了一个线性分类分界线，对其进行分类。多层的神经网络的本质就是复杂函数拟合。

在设计一个神经网络时，输入层的节点数需要与特征的维度匹配，输出层的节点数要与目标的维度匹配。而中间层的节点数，却是由设计者指定的。

机器学习模型训练的目的，就是使得参数尽可能的与真实的模型逼近。具体做法是这样的。首先给所有参数赋上随机值。我们使用这些随机生成的参数值，来预测训练数据中的样本。样本的预测目标为Yp，真实目标为Y。那么，定义一个值loss，计算公式如下。

loss = (Yp - Y)²

这个值称之为**损失**（loss），我们的目标就是使对所有训练数据的损失和尽可能的小。

**梯度下降**算法每次计算参数在当前的梯度，然后让参数向着梯度的反方向前进一段距离，不断重复，直到梯度接近零时截止。一般这个时候，所有的参数恰好达到使损失函数达到一个最低值的状态。

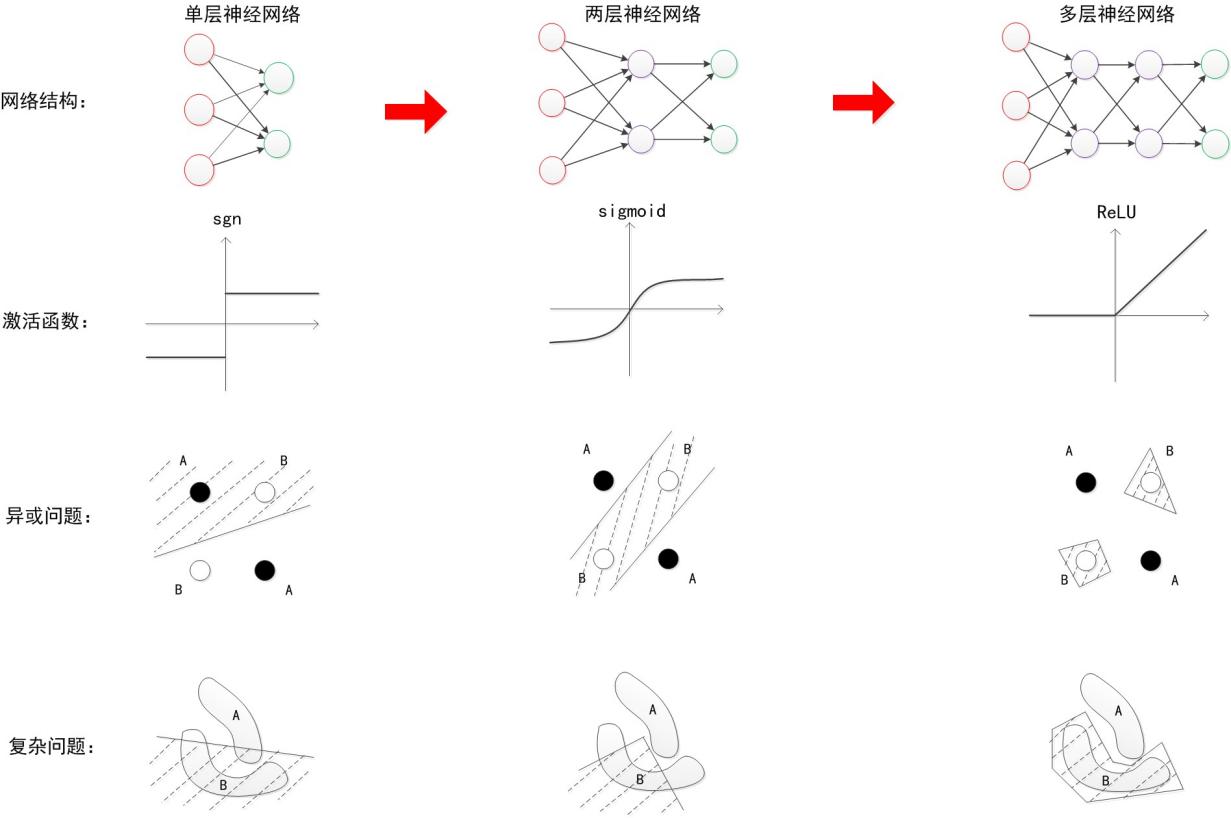
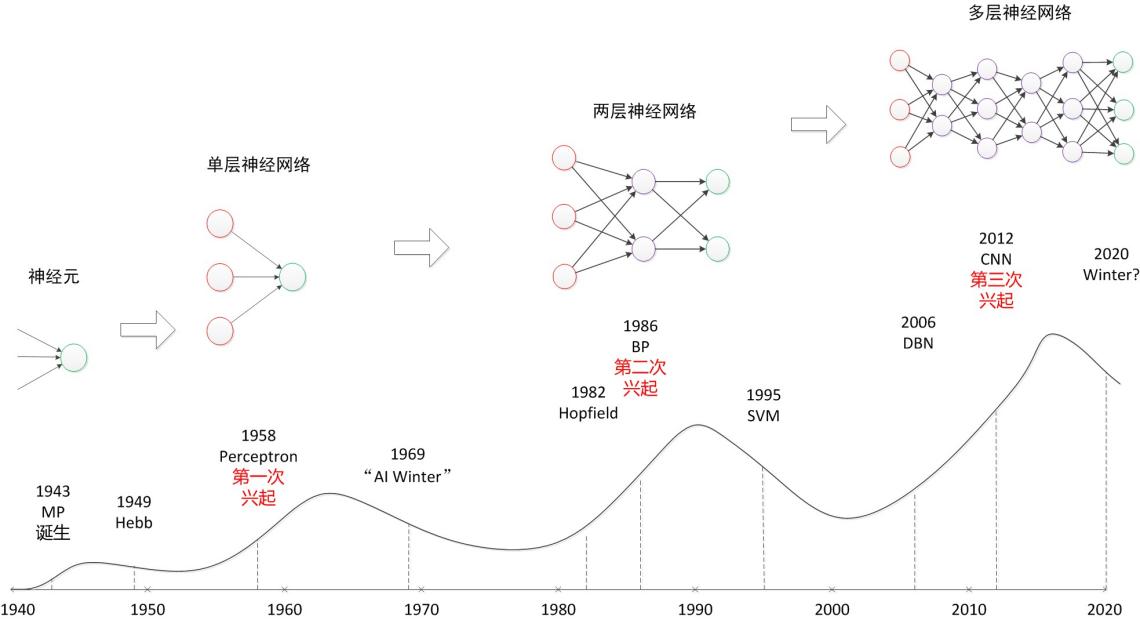
由于结构复杂，每次计算梯度的代价很大。因此还需要使用反向传播算法。反向传播算法是利用了神经网络的结构进行的计算。不一次计算所有参数的梯度，而是从后往前。首先计算输出层的梯度，然后是第二个参数矩阵的梯度，接着是中间层的梯度，再然后是第一个参数矩阵的梯度，最后是输入层的梯度。

**反向传播算法**的启示是数学中的链式法则。提升模型在测试集上的预测效果的主题叫做**泛化**，相关方法被称作**正则化**。神经网络中常用的泛化技术有权重衰减等。

增加更多的层次有什么好处？更深入的表示特征，以及更强的函数模拟能力。更深入的表示特征可以这样理解，随着网络的层数增加，每一层对于前一层次的抽象表示更深入。在神经网络中，每一层神经元学习到的是前一层神经元值的更抽象的表示。例如第一个隐藏层学习到的是“边缘”的特征，第二个隐藏层学习到的是由“边缘”组成的“形状”的特征，第三个隐藏层学习到的是由“形状”组成的“图案”的特征，最后的隐藏层学习到的是由“图案”组成的“目标”的特征。通过抽取更抽象的特征来对事物进行区分，从而获得更好的区分与分类能力。

在单层神经网络时，我们使用的激活函数是**sgn函数**。到了两层神经网络时，我们使用的最多的是**sigmoid函数**。而到了多层神经网络时，**ReLU函数**在训练多层神经网络时，更容易收敛，并且预测性能更好。ReLU函数不是传统的非线性函数，而是分段线性函数。其表达式非常简单，就是y=max(x,0)。

泛化技术变的比以往更加的重要。神经网络的层数增加了，参数也增加了，表示能力大幅度增强，很容易出现**过拟合现象**。因此正则化技术就显得十分重要。**Dropout技术**，以及**数据扩容技术**是目前使用的最多的正则化技术。



监督学习、回归、分类

监督学习：从标记的训练数据来推断一个功能的机器学习任务

非监督学习：利用一组没有标记响应的输入数据组成的数据集中进行推断

**监督学习模型：**

**线性回归模型**

**分类模型**

训练集的标准符号

监督学习和无监督学习是机器学习的两种主要方法。以下是如何判断一个事例是监督学习还是无监督学习的一些关键点：

监督学习：  
监督学习是一种通过已知输入和输出来训练模型的方法。在这种学习中，训练数据集中的每个输入都与一个已知的输出相关联，模型的任务是学习如何根据输入预测输出。常见的监督学习算法包括线性回归、决策树、支持向量机（SVM）和神经网络等。

例如，假设你有一个带有标记的数据集，标记为“猫”或“狗”，你的任务是训练一个模型来区分这两类图像。在这个例子中，输入是图像，输出是标记（猫或狗），因此这是一个监督学习问题。

无监督学习：  
无监督学习是一种不需要已知输出数据来训练模型的方法。在这种情况下，模型的任务是学习输入数据的内在结构和关系，而不是预测特定的输出。常见的无监督学习算法包括聚类、降维和自编码器等。

例如，假设你有一组未经标记的图像，你的任务是找出这些图像中的模式或集群。在这个例子中，输入是图像，输出是图像的集群，因此这是一个无监督学习问题。

总的来说，判断一个事例是监督学习还是无监督学习主要取决于你是否拥有输入数据的标签或者是否需要对数据进行某种形式的标记。如果你有标签数据，那么你可以使用监督学习算法来训练模型并预测输出。如果你没有标签数据，那么你可以使用无监督学习算法来探索数据的内在结构和关系。

以下是一些判断监督学习和无监督学习的试题：

监督学习是通过已知的输入和输出来训练模型，而无监督学习不需要已知输出数据来训练模型。

给定一组图像数据，如果你的任务是找出图像中的物体，那么这是一个监督学习问题，因为你需要使用标记好的数据集来训练模型进行物体检测。

给定一组社交媒体数据，如果你的任务是找出数据中的社区或群体，那么这是一个无监督学习问题，因为你需要使用聚类算法来找出数据中的模式和结构。

给定一组股票价格数据，如果你的任务是预测未来的股票价格，那么这是一个监督学习问题，因为你需要使用历史价格数据来训练模型进行股票价格预测。

给定一组用户数据，如果你的任务是进行用户聚类或用户画像，那么这是一个无监督学习问题，因为你需要使用降维或聚类算法来找出用户数据的结构和模式。

给定一组医学图像数据，如果你的任务是检测图像中的病变或异常区域，那么这是一个监督学习问题，因为你需要使用标记好的图像数据集来训练模型进行病变检测。

给定一组语音数据，如果你的任务是进行语音识别或语音转换，那么这是一个监督学习问题，因为你需要使用标注好的语音数据来训练模型识别语音信号。

给定一组文本数据，如果你的任务是进行文本分类或文本聚类，那么这是一个监督学习问题，因为你需要使用标注好的文本数据来训练模型进行文本分类或聚类。