一、讨论的问题：

1.在无监督学习中，降维的主要作用是什么？

主要有两个作用：

一是为了更好地适应模型。特别是在样本较少的数据集上，如果维度过高导致样本在空间上分布稀疏，很容易造成过拟合。解决这一问题的方法有两个：一个是借助正则来限制模型的复杂度；另一个就是降维来改变空间维度上的稀疏分布。

二是避免维度灾难的问题。在维度过高的时候，如果采用常见的距离度量方法度量空间样本上的距离，就会陷入维度灾难，也即大部分样本间的距离会被压缩到很小的范围内，以致无法区分。解决这一问题的方法也有两个：一个是优化距离度量方法，在计算过程中压缩；另一个就是降维，通过损失部分信息以达到规避维度过高的问题。

2.聚类和降维有什么区别和联系？

联系：聚类是对类别变量做降维的方法之一。

区别：降维是在数据预处理时为了防止过拟合时常用的方法，聚类是在整个数据集上按一个定义的拓扑距离进行分组的过程。

3. 提升方法中，为什么负梯度可以近似残差？

理解：残差是目前值与最终值的差距，也可以认为是目前损失距离最低点的距离。负梯度在当前的值是沿着此方向走一步在以后梯度不变的情况下目标下降的距离。它们在数值上相似的原因是二者都是在朝着最优点走，所以差距不会很大。

4. 有什么比较好的降维方法？

讨论后的理解：主成分分析算法PCA：通过某种线性投影，将高维的数据映射到低维的空间中表示，并期望在所投影的维度上数据的方差最大，以此使用较少的数据维度，同时保留住较多的原始数据点的特性。

二、读书计划：

本周第十三章

下周第十四章

三、读书摘要

常用的无监督学习算法主要有主成分分析方法PCA等，等距映射方法、局部线性嵌入方法、拉普拉斯特征映射方法、黑塞局部线性嵌入方法和局部切空间排列方法等。

从原理上来说PCA等数据降维算法同样适用于深度学习，但是这些数据降维方法复杂度较高，并且其算法的目标太明确，使得抽象后的低维数据中没有次要信息，而这些次要信息可能在更高层看来是区分数据的主要因素。所以现在深度学习中采用的无监督学习方法通常采用较为简单的算法和直观的评价标准。