

第二周

注意：回顾题一般为课本上的题目，需要用自己的话进行总结打卡。补充题则为课本上没有另行补充的题目，会提供一些参考答案。

Day4-Day5

1. 回顾：请参考课本 4.1 节以及视频，从最大方差角度推导 PCA 算法。

参考答案：参考课本 4.1 节以及视频进行总结。

2. 回顾：请参考课本 4.2 节以及视频，从最小平方误差角度推导 PCA 算法。

参考答案：参考课本 4.2 节以及视频进行总结。

3. 回顾：请参考课本 4.4 节，对比 PCA 和 LDA 各自的特点。

参考答案：参考课本 4.4 节进行总结。

4. 补充：请从你的角度阐述 PCA 的一些优缺点。

参考答案：作为一个非监督学习的降维方法，它只需要特征值分解，就可以对数据进行压缩，去噪，因此在实际场景应用很广泛。为了克服 PCA 的一些缺点，出现了很多 PCA 的变种，比如为解决非线性降维的 KPCA，还有解决内存限制的增量 PCA 方法 Incremental PCA，以及解决稀疏数据降维的 PCA 方法 Sparse PCA 等。

PCA 算法的主要优点有：

- (1) 仅仅需要以方差衡量信息量，不受数据集以外的因素影响。
- (2) 各主成分之间正交，可消除原始数据成分间的相互影响的因素。
- (3) 计算方法简单，主要运算是特征值分解，易于实现。

PCA 算法的主要缺点有：

- (1) 主成分各个特征维度的含义具有一定的模糊性，不如原始样本特征的解释性强。
- (2) 方差小的非主成分也可能含有对样本差异的重要信息，因降维丢弃可能对后续数据处理有影响。