

第三周

Day1-Day2

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。

1. 回顾：请参考课本5.1节，总结K均值算法的步骤和怎么选取K值

参考答案：参考课本5.1节进行总结

2. 回顾：请参考课本5.2节以及视频，说明EM算法为什么收敛

参考答案：参考课本5.2节以及视频进行总结

3. 回顾：请参考课本5.2节以及视频，说明EM算法中E步和M步具体步骤，E中的期望是什么

参考答案：参考课本5.2节以及视频进行总结

4. 补充：EM为什么不用牛顿法或者梯度下降？

参考答案：EM求解的是带隐变量的参数估计问题，隐变量是隐含无法观察的，因此对似然函数进行求导是求不出来的

5. 补充：如何快速收敛数据量大的k-means算法？

参考答案：Mini Batch Kmeans方法：

(1) 从数据集中随机抽取一些数据形成小批量，把他们分配给最近的质心

(2) 更新质心：与k均值算法相比，数据的更新是在每一个小的样本集上。对于每一个小批量，通过计算平均值得到更新质心，并把小批量里的数据分配给该质心，随着迭代次数的增加，这些质心的变化是逐渐减小的，直到质心稳定或者达到指定的迭代次数，停止计算。