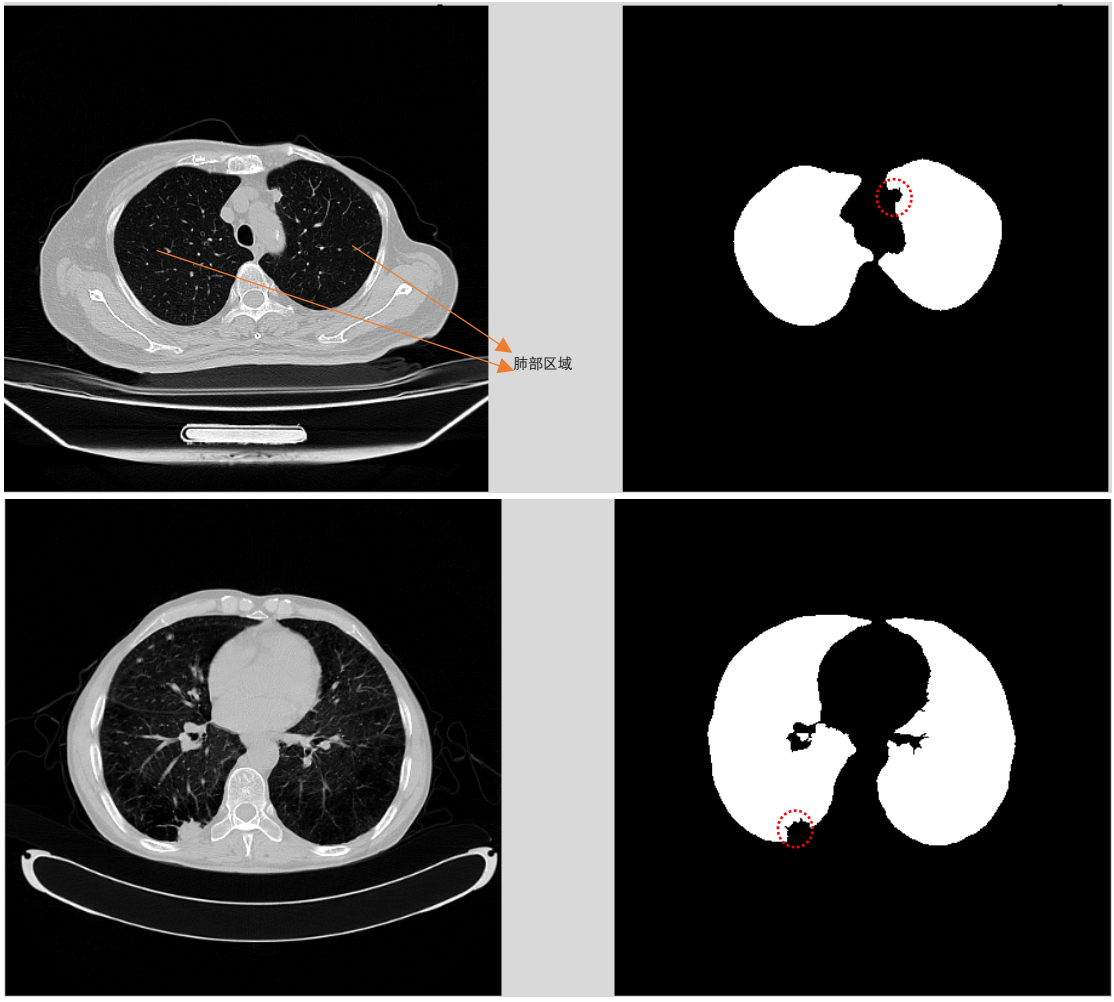
以下任选一题。

题目一：（图像算法题）实现下图中肺部区域的分割，

1. 规定时间 1 个小时，尽量实现满足要求的结果，若超时完成，由 HR 记录完成时间。
2. 请给出具体的实现方法的 code，并给出运行的结果截图。
3. 编程语言不限，方法不限，代码简洁、复杂度低、通用性强为优。

题目描述：图（左上，左下）为肺部 CT 扫描轴截面的一帧，对应附件图片 case1_1.png 和 case1_2.png，图（右上，右下）为对应的肺部区域分割 mask 效果示意图（待实现，给出通用的实现方法），请提供具体实现方法，并展示分割效果截图。(加分项：示意图中的分割 mask 存在 bug，如红色虚线圆圈所示，此处属于肺内病灶，不应缺失)



题目二：（机器学习题）聚类算法和分布式机器学习

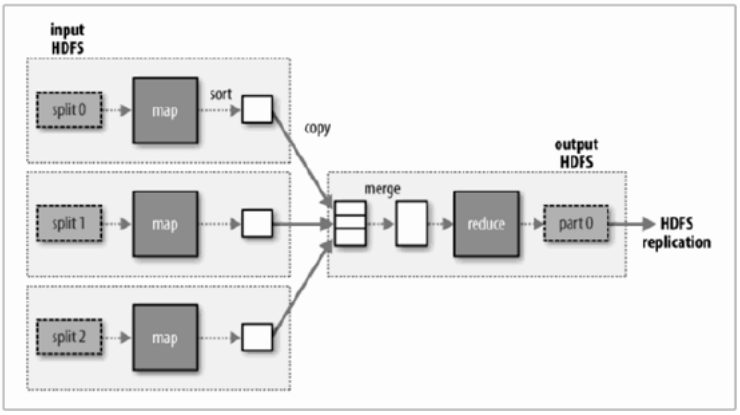
1. 规定时间 40 分钟，尽量实现满足要求的结果，若超时完成，由 HR 记录完成时间。
2. 可以给出伪代码，或者任意语言实现。
3. 第 c 题完整能跑通的代码加分。可以自己编写 user cases。推荐使用 jupyter notebook 反馈结果。若能给出完整代码可以加时 20 分钟。

K-Means 是一种常用的聚类算法。

a) 请简述 K-Means 对数据的基本假设和算法流程

b) 和高斯混合模型聚类算法（GMM）相比，K-Means 和 GMM 在算法上关系，他们各自有什么优点和缺点？

c) 在大数据时代，容易并行化的算法更受欢迎。在大规模分布式计算中，Map-Reduce（映射-归纳框架）是一种最受欢迎的计算模型，很多著名的分布式框架都采用该编程模型，例如 Hadoop 和 Spark。使用该编程模型时，程序员通常只需要定义 **Mapper** 和 **Reducer** 两个函数，其余都交给框架即可。



其中，Mapper 函数用于给各台 worker 机发出相同的指令（即“映射”），并得到相应的返回结果。Reducer 函数用于回收这些 worker 的返回结果（即“归纳”），并将这些分布式的返回结果汇总成一个单独的返回结果。以词频计数（word counting）为例，该任务是分布式计算中的“hello world”程序。举个例子，我们需要统计分布在 100 台计算机上的 1TB 的文本词频，即“machine”出现多少次，“learning”出现多少次等。在该任务中，Mapper 和 Reducer 的设计是非常自然的，Mapper 将各台计算机（worker）上的文本统计成词频字典（一个“单词”的哈希表），如{“machine”：1024，“learning”：666，…}，每台 worker 维持一个词频字典。之后，Reducer 将这些 worker 传来的词频字典汇总成一个总的词频字典，即各个单词词频对应相加。

事实上，K-Means 是一个非常容易利用 Map-Reduce 框架并行化的聚类算法。我们将数据分布在多台机器上，最终任务是将每个数据分配一个 1,…,K 的标签（K 是需要聚类的簇数）。如同标准的 K-Means，其并行版本同样是一个迭代算法。在迭代过程开始前，我们进行和标准 K-Means 相同的初始化过程。请指出该初始化过程是什么，并设计 K-Means 并行算法中**一个迭代过程中**的 Mapper 和 Reducer 函数，**分别说明两者**的计算过程或伪代码。