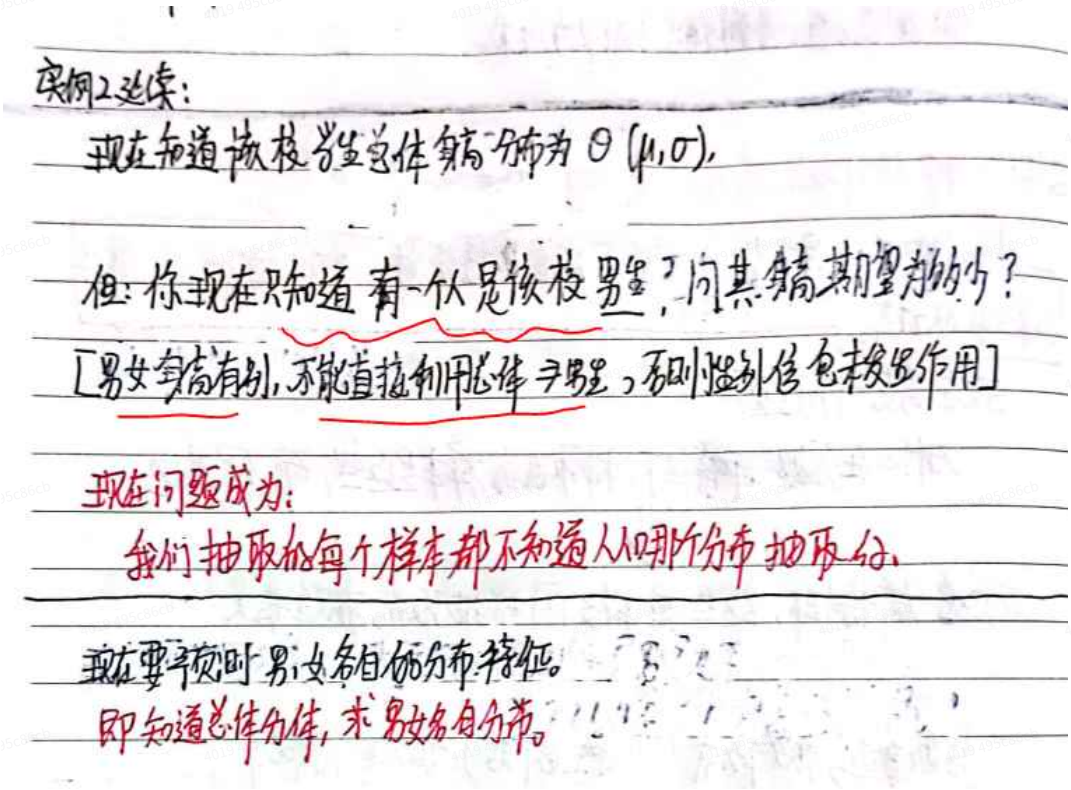
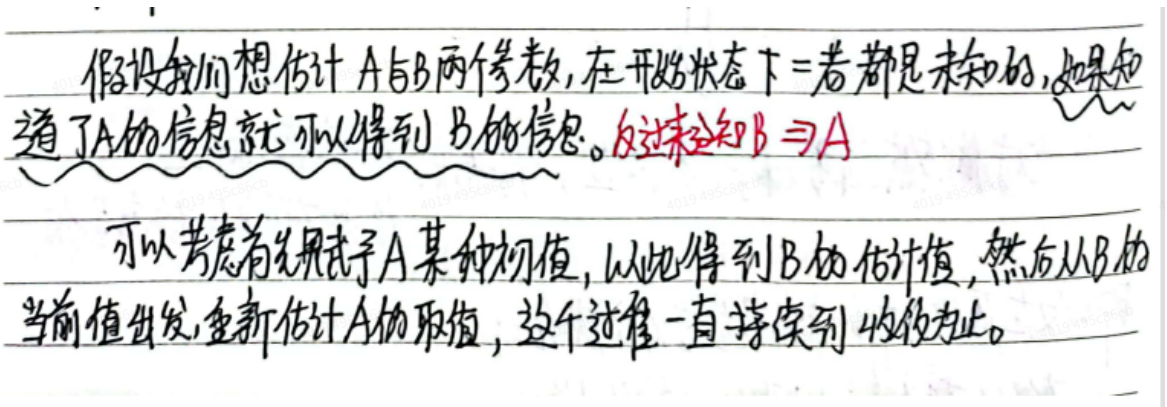
2.EM算法


实例2延续：
现在知道该校学生总体身高分布为 $\theta(\mu,\sigma)$.
但：你现在只知道有一个人是该校男生，问其身高期望为多少？
[男女身高有别，不能直接利用总体学生，否则增加信息未发挥作用]
现在问题成为：
我们抽取的每个样本都不知道人从哪个分布抽取出.
现在要预测男、女各自的分布特征。
即知道总体分体，求男女各自分布。

1.EM算法


假设我们想估计A与B两个参数，在开始状态下二者都是未知的，如果知道了A的信息就可以得到B的信息。反过来知B可A
可以考虑首先赋予A某种初值，以此得到B的估计值，然后从B的当前值出发，重新估计A的取值，这个过程一直持续到收敛为止。

Date: Page:

对于男女身高而言，已知每个人的身高，但不知性别。

要求男女身高各自的分布特征，则需要首先知道其性别，然后根据性别分开这些人，对男女分别使用极大似然估计法。

最大期望算法：

E步：先随便猜一下男生（身高）的正态分布的参数，如男生均值1.7m，方差0.1m。

然后测出每个人更属于这个分布还是 → 性别：男
                                  → 性别：女

M步：将上面每个人分为男女两部分后，

我们对其分别利用最大似然估计法，计算两分布的参数。

在更新完两分布参数后，每个样本属于这两个分布的概率再次改变，则继续E、M如此循环，直到参数不再发生变化为止。

2.算法推导

假设我们有一个样本集 $\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$，包含m个独立样本。

但每个样本 i 对应的类别 $z^{(i)}$ 是未知的（相当于聚类），即称为隐含变量。

由于有隐含变量限制，所以无法直接使用最大似然求解。

但是，对于参数估计，我们本质上还是想获得一个使似然函数最大的参数 θ，现在仅仅不过似然函数式中的了一个未知变量 z

$$L(\theta) = \sum_i p(x^{(i)}; \theta) = \sum_i \sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta)$$

目标：找到合适的 θ 与 z 让 $L(\theta)$ 最大。

若先估出一个参数，计算出 z，再利用求到的 z 更新参数，依次循环。

notebook