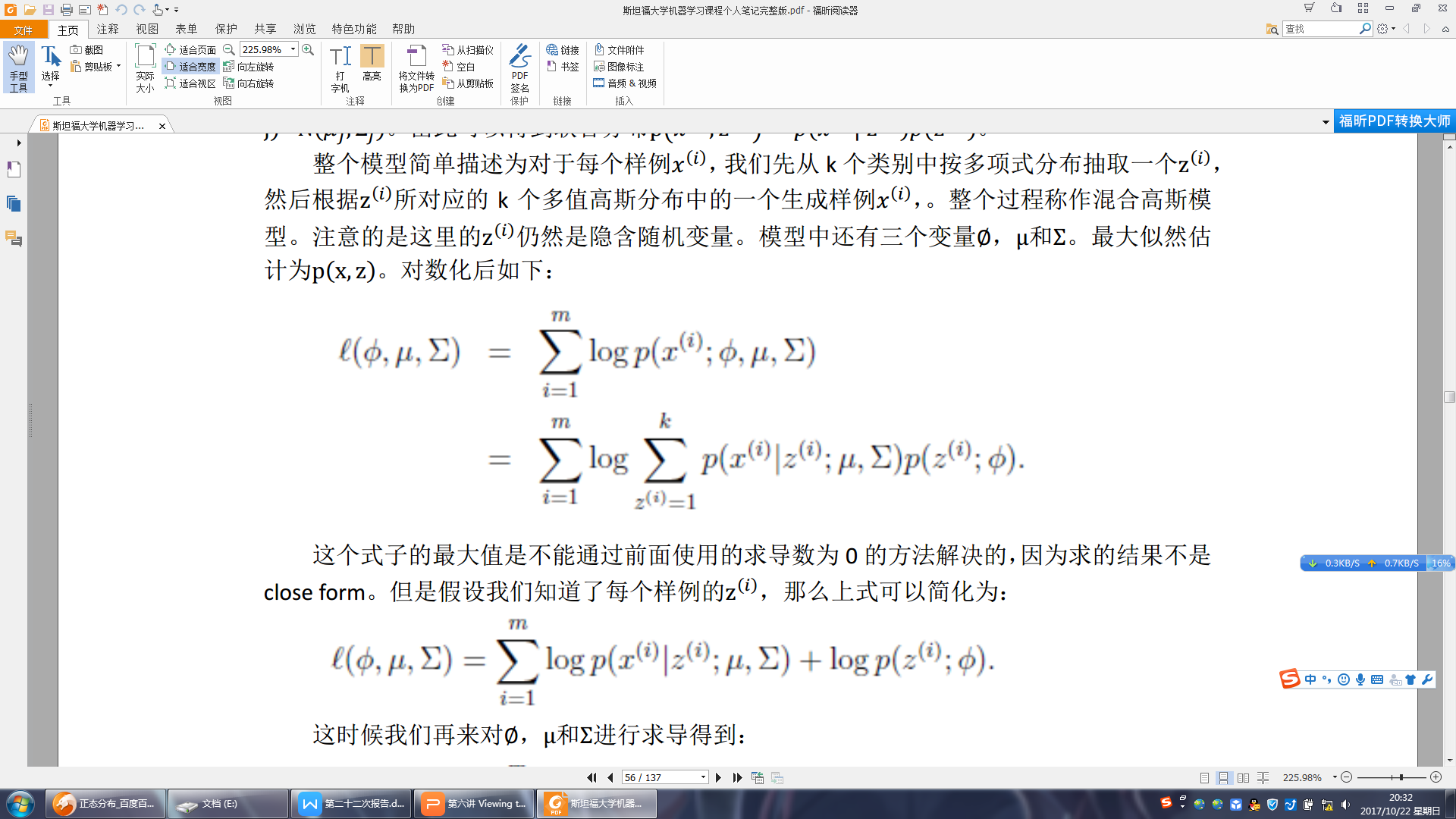
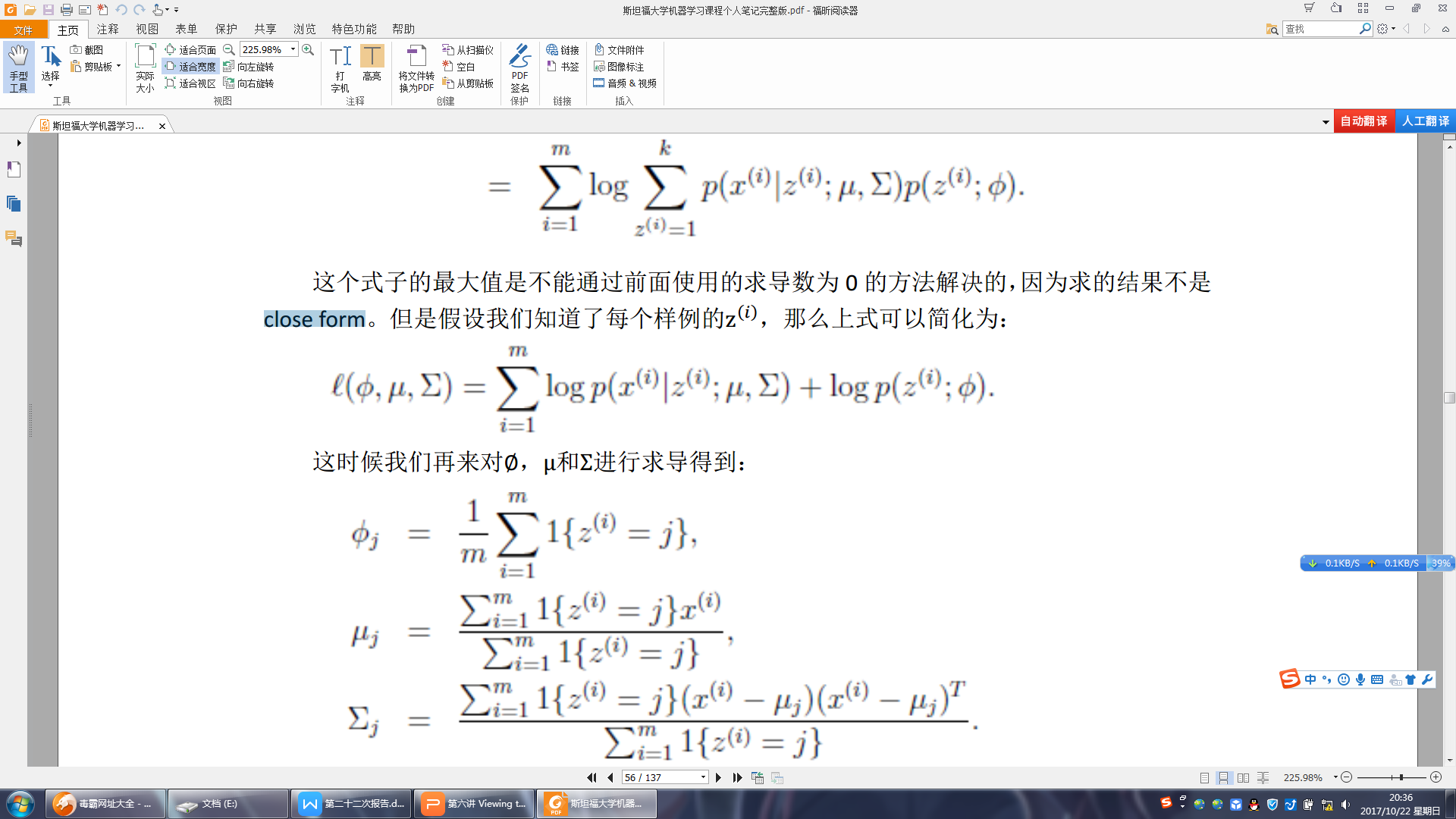
本周学习：混合高斯模型（Mixtures of Gaussians）和EM算法（Expectation-Maximization ）

给定样本{x1，x2，......xm}，将隐含类标签用zi表示。这里上一次的k-means不一样（那个是直接指定类别）。首先，假设zi是遵循一定的概率分布的，这里就认为是满多项式分布，即是zi~Multinomial（Ф），且p（zi=j）=Фj，1<=i<=k。在给定zi的时候，xi满足多值高斯分布，即（xi|zi=j）~N（uj，Σj）。那么求得联合密度分布为p（xi，zi）=p（xi|zi）p（xi）。

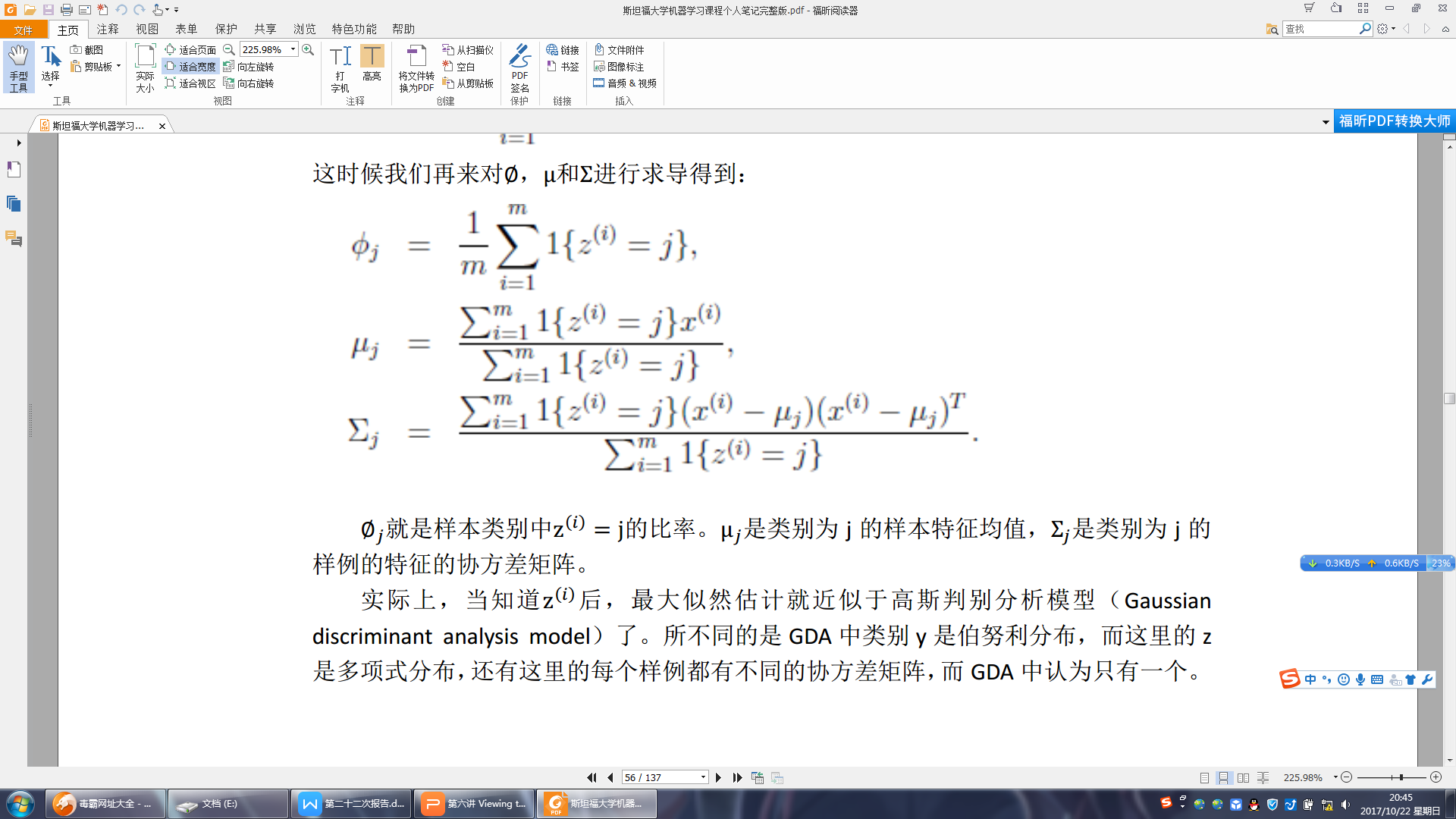
整个模型简单描述为对于每个样例xi，我们先从 k 个类别中按多项式分布抽取一个zi，然后根据zi所对应k个多值高斯分布中的一个生成样例xi。整个过程称作混合高斯模型。 注意的是这里的zi仍然是隐含随机变量。还有三个变量Ф，u，Σ。设它的最大似然估计为p（x，z），那么对数化后为



假设已知了zi，那么上式就可以化简为



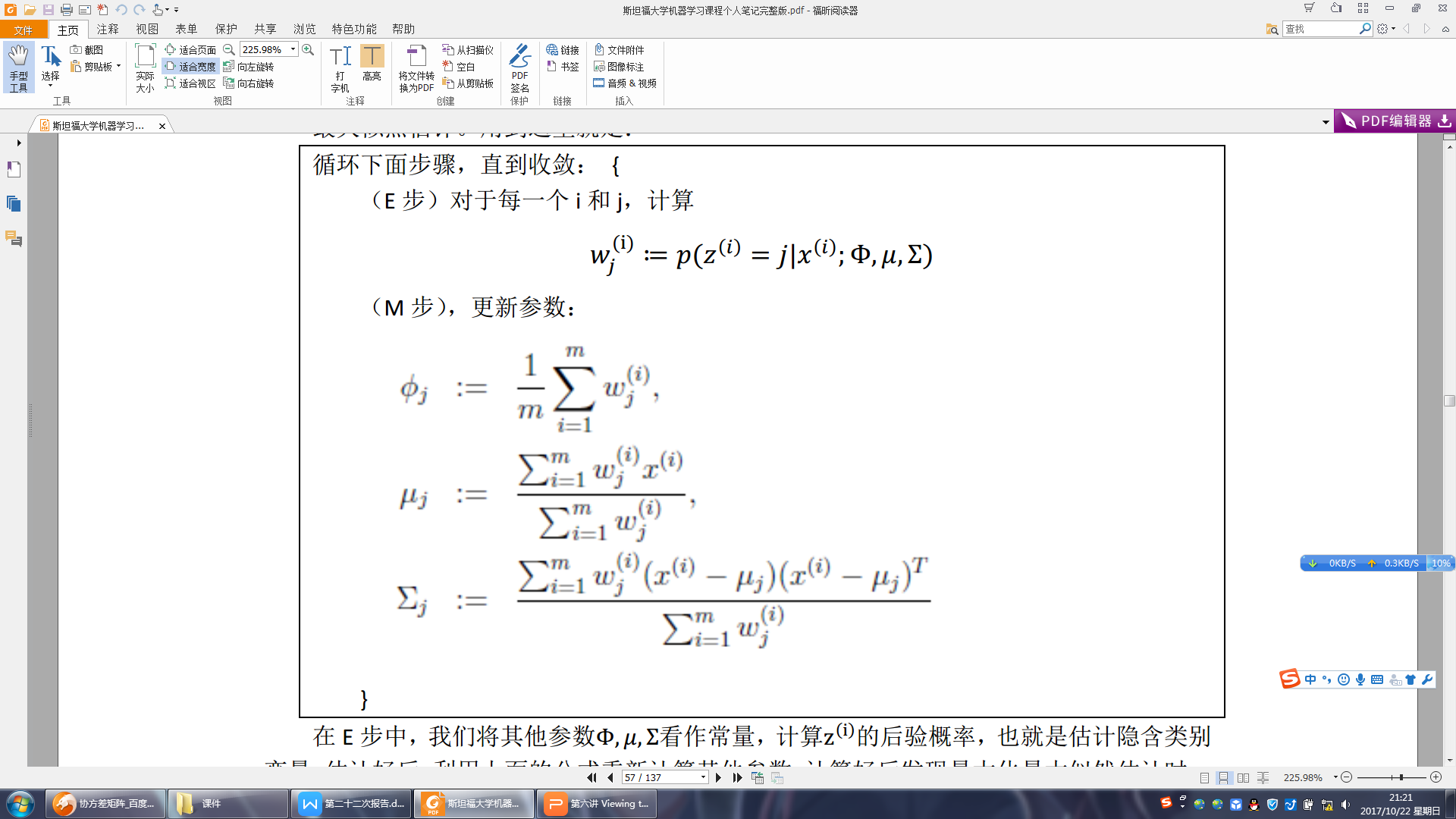
然后就对上面的式子进行对各个变量的求导并解方程，得到



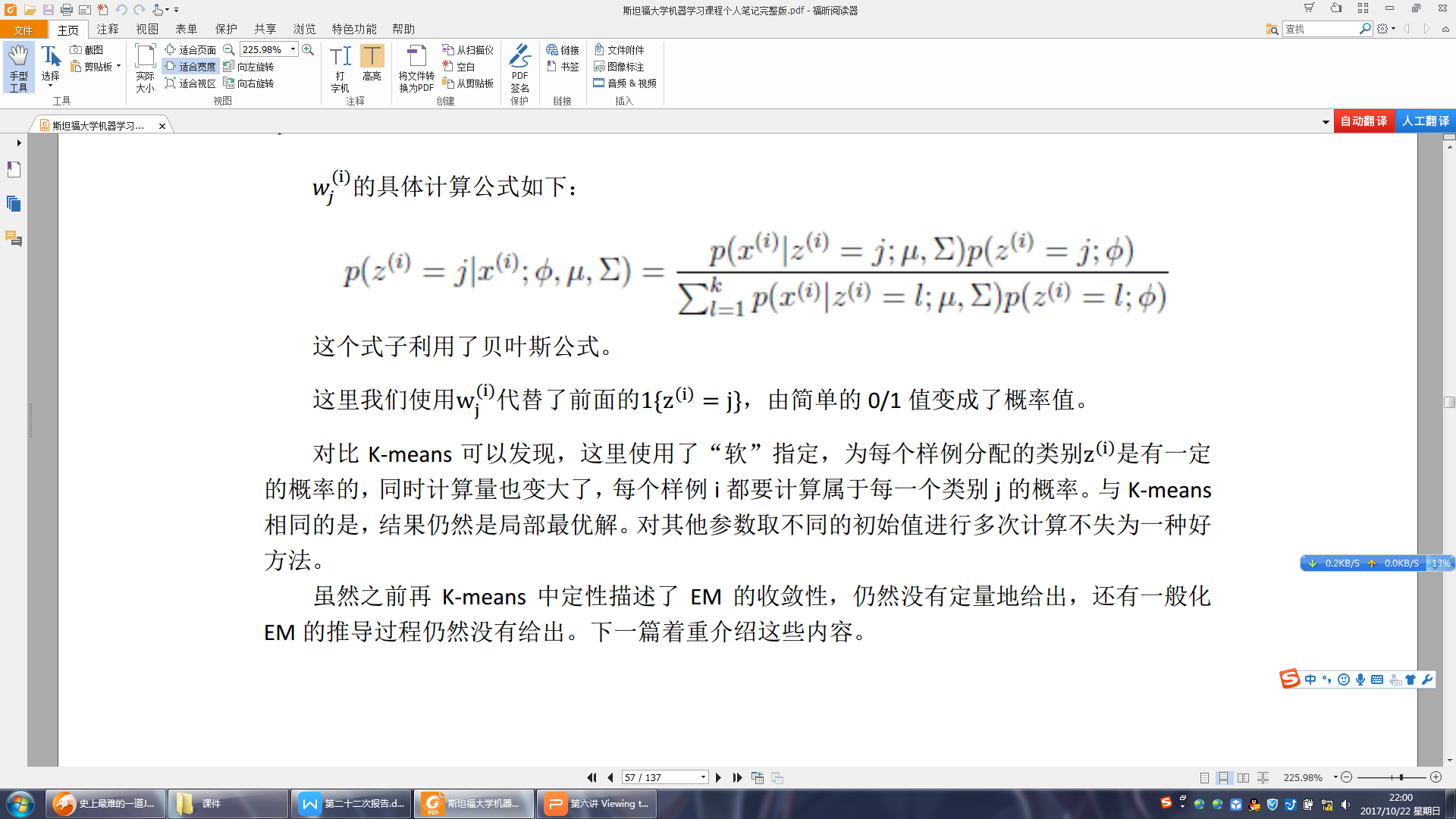
可以看出，Фj就是样本中zi=j的比值；uj是类别为j的样本的特征均值；Σj是类别为j的样例的特征的协方差矩阵。

有这样一个事实，当知道zi之后，最大似然估计就接近高斯判别模型了（Gaussian  
discriminant analysis model）（所不同的是 GDA 中类别 y 是伯努利分布，而这里的 z  
是多项式分布，还有这里的每个样例都有不同的协方差矩阵，而 GDA 中认为只有一个）。

回到问题开始时，实际上zi是未知的。这时候就要利用强大的EM思想。第一步猜测隐含类别变量z，第二步更新其他参数，以获得最大的最大似然估计。用到这里就是：



在E步中，将其他参数Ф，u，Σ看作常量，计算zi的后验概率，也就是估计隐含类别变量。估计好了之后，利用上面的公式重新计算其他参数，计算之后发现最大化最大似然估计时，wj的值又不对了，重新计算，周而复始，直至收敛。



这里，我们使用wj（i）代替了前面的1{zi=j}，由简单的0/1值变成了概率值。

优缺点：

这里给每个样例分配的类别zi是有一定概率的，同时计算量也增加了，每一个样例i都要计算属于每一个类别j的概率。与k-means相同，结果仍然是局部最优解。对其他参数选取不同的初始值进行多次计算也是可以的。

参考资料：

[1].斯坦福大学机器学习个人笔记