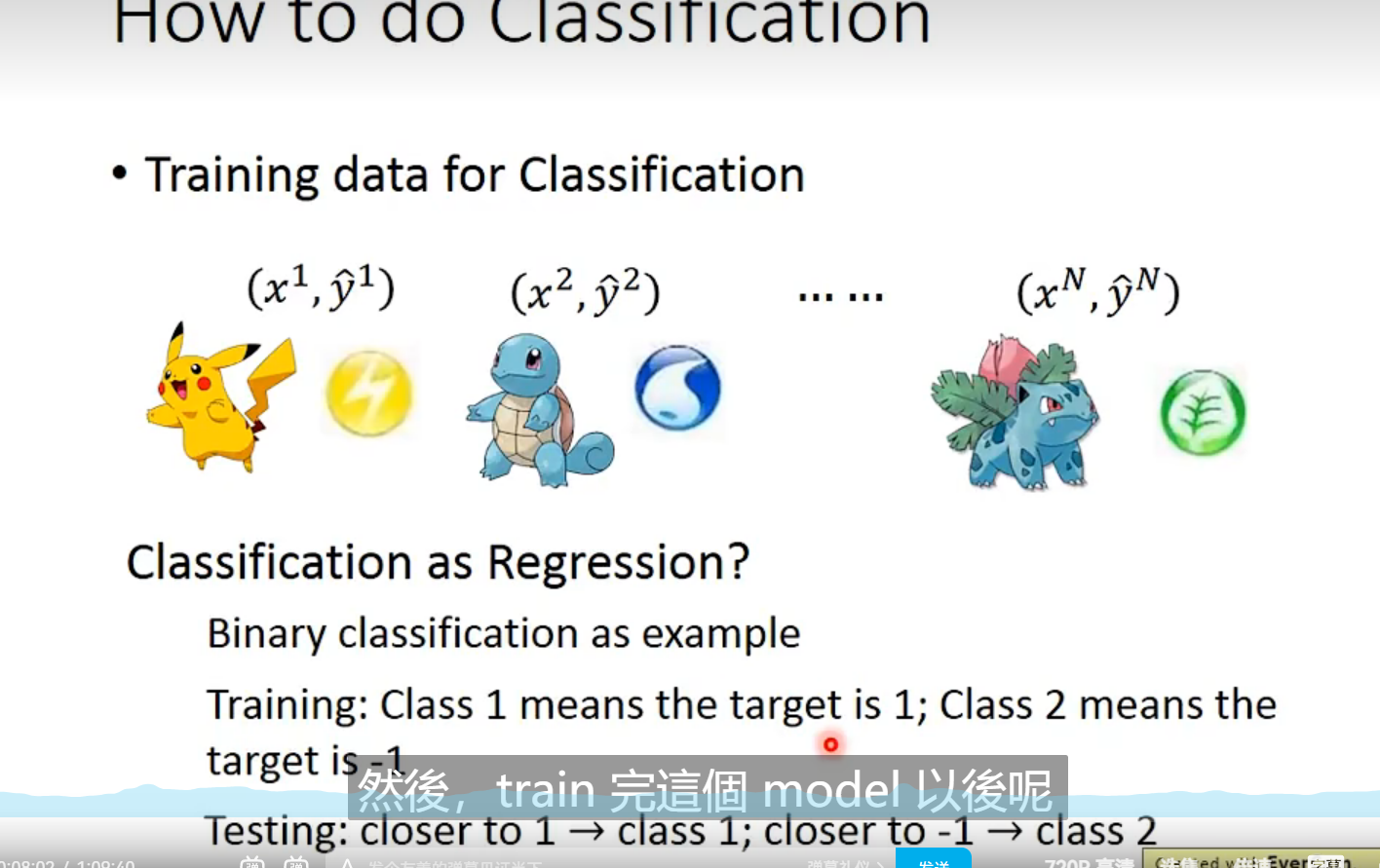
# 分类神奇宝贝

## Classification

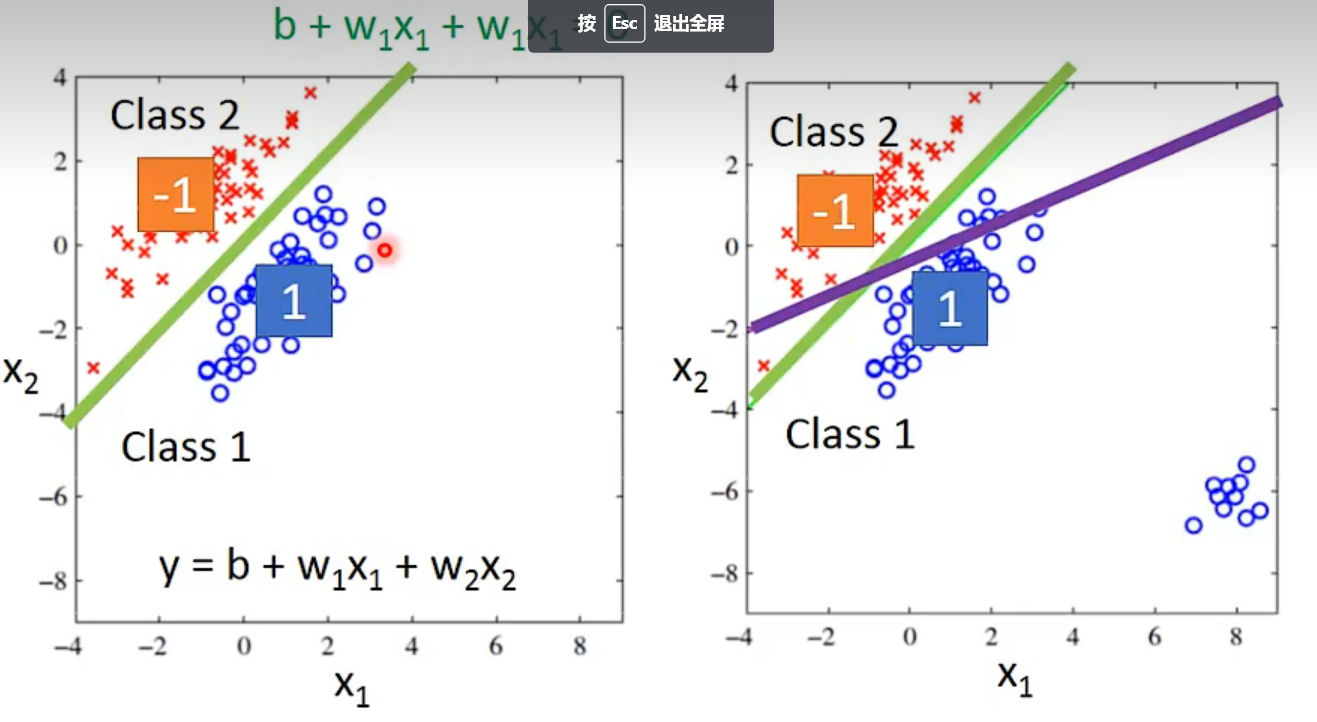
Input（vector），output=classification

How to do classification

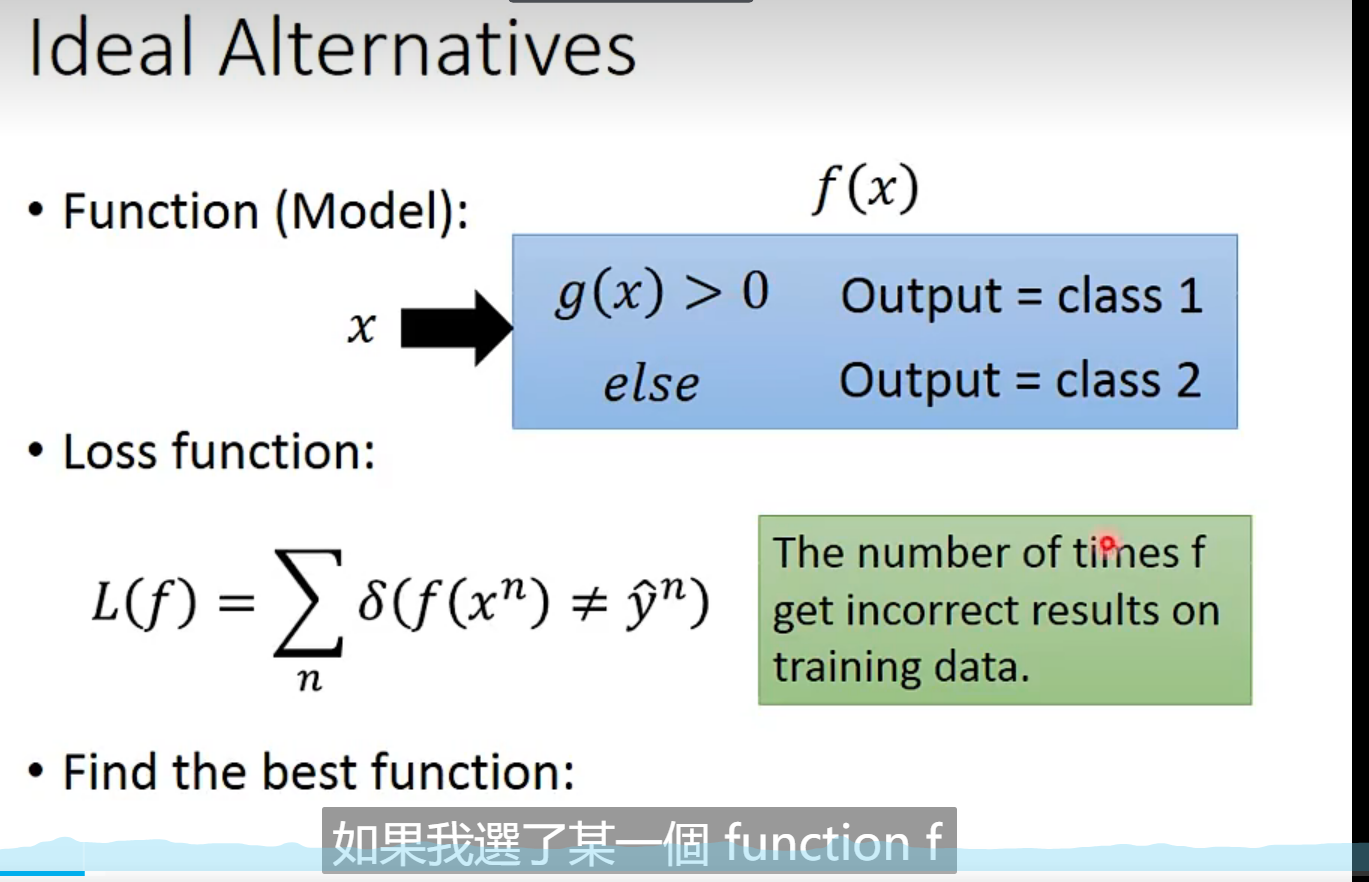
Training data for classification



找一个分界线

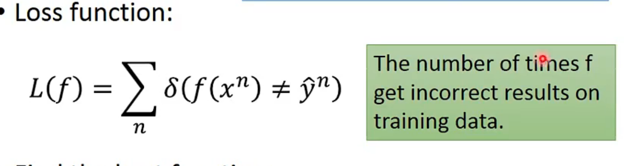


如果就是单纯的用regression（线性回归）会出现紫色的这条线（因为这种方法所得出的Loss越小，这样我们来进行测试的时候会出现分类错误。

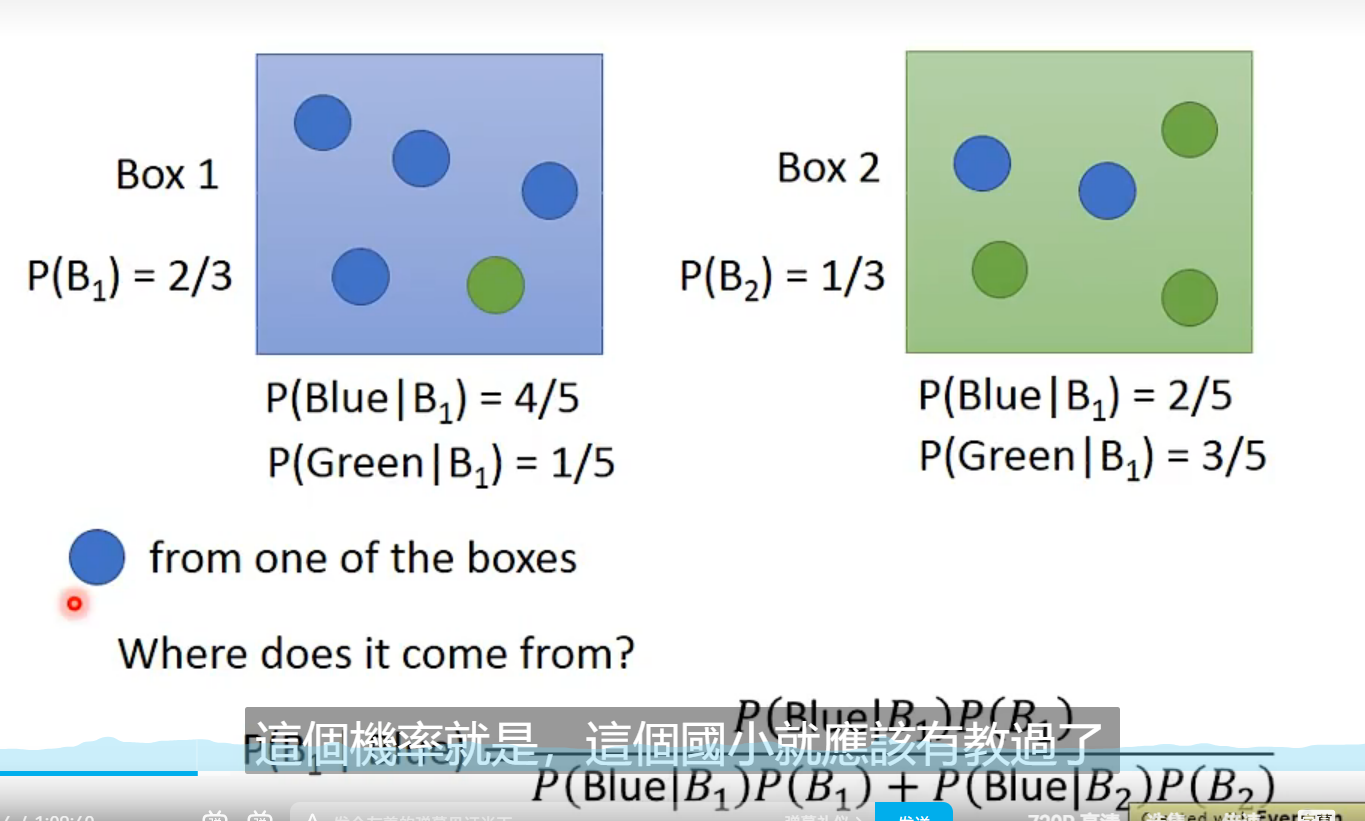


我们首先定义一个函数g（x）输入测试值x

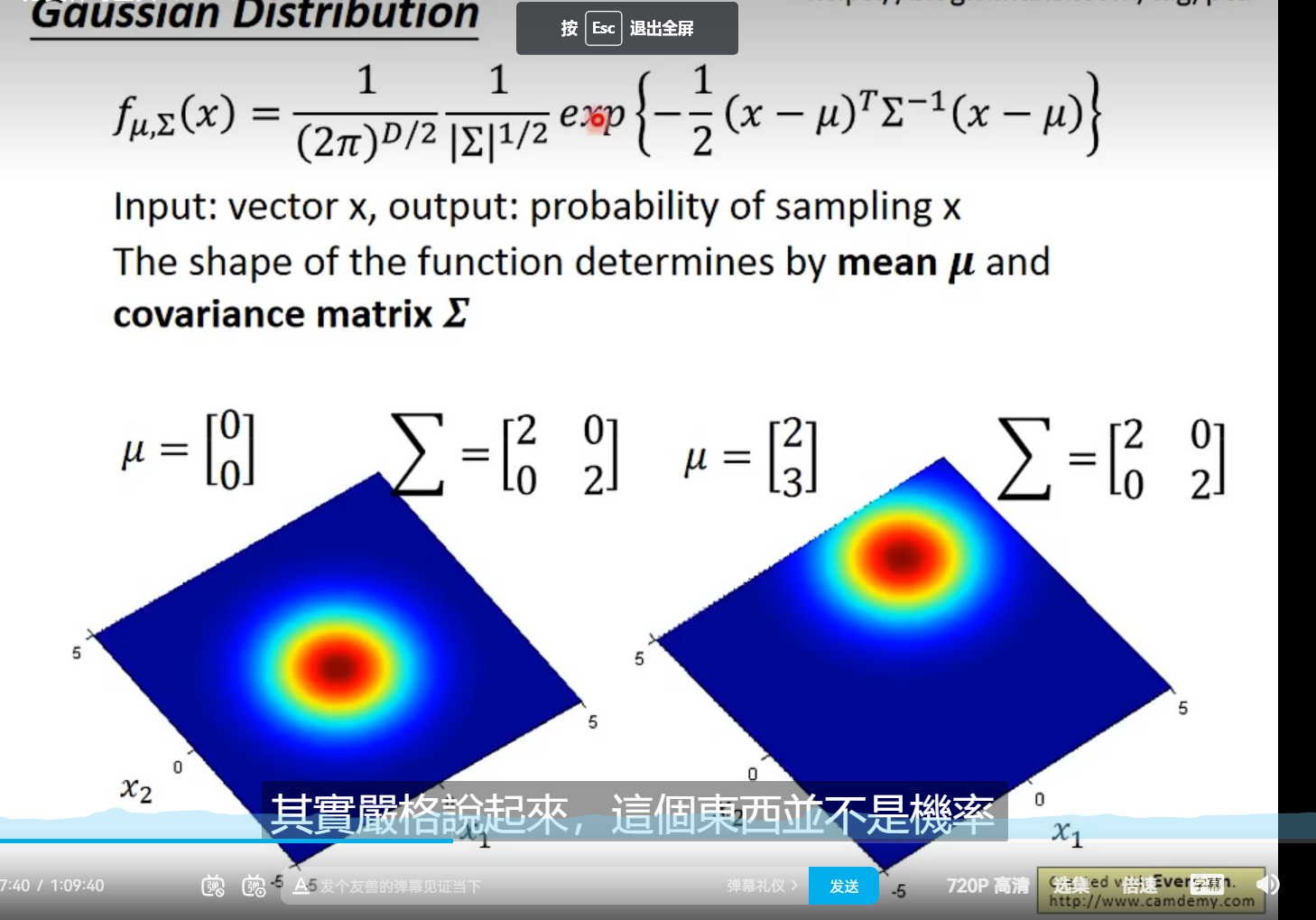
* If g（x）>0,定义为class1
* If g(x)<0,定义为class2
* If g（x）==0,正好与真实值相等



是代表分类错误的次数



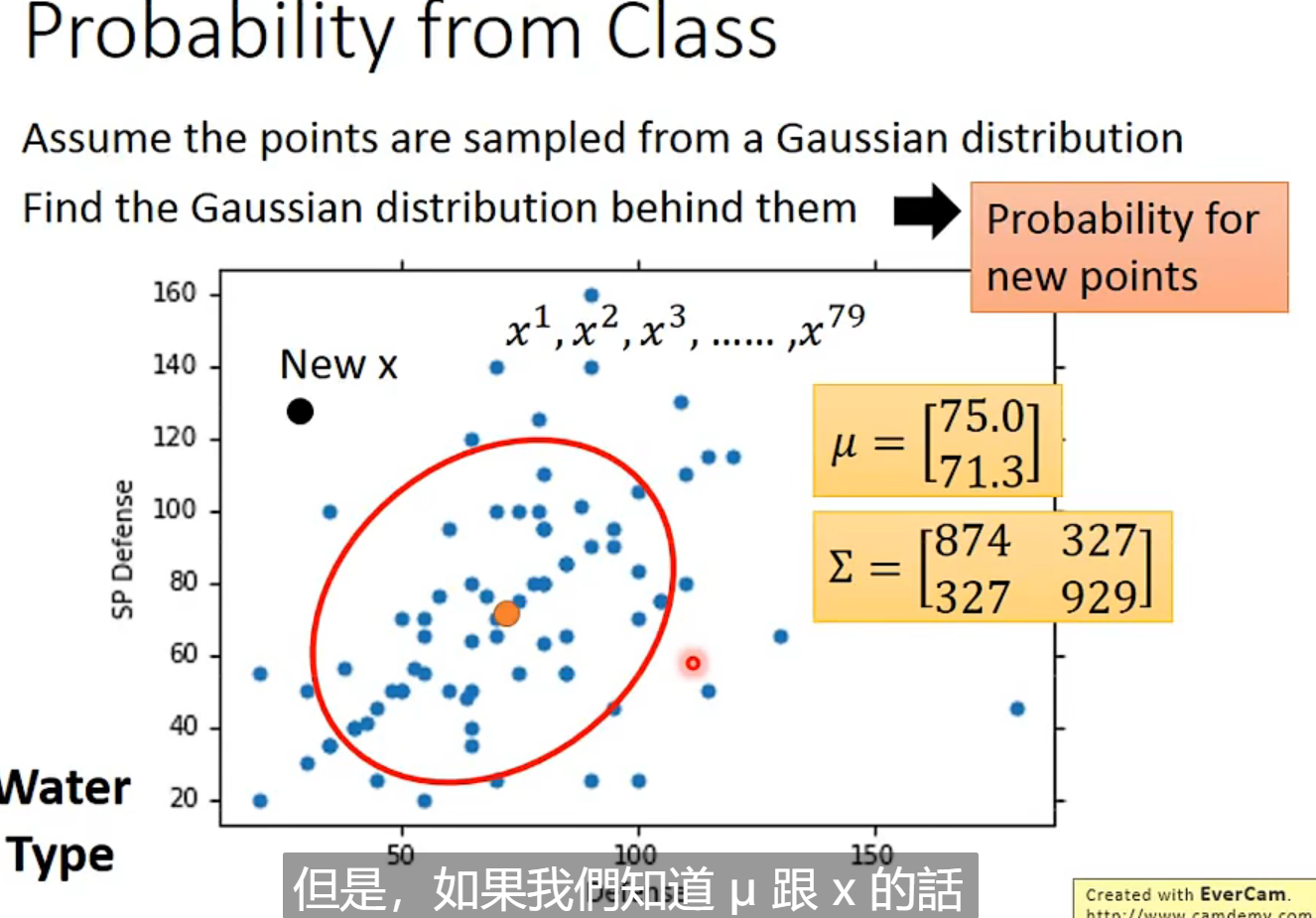
条件概率，全概率公式



* 表示分布的最高点

表示分布的散落程度

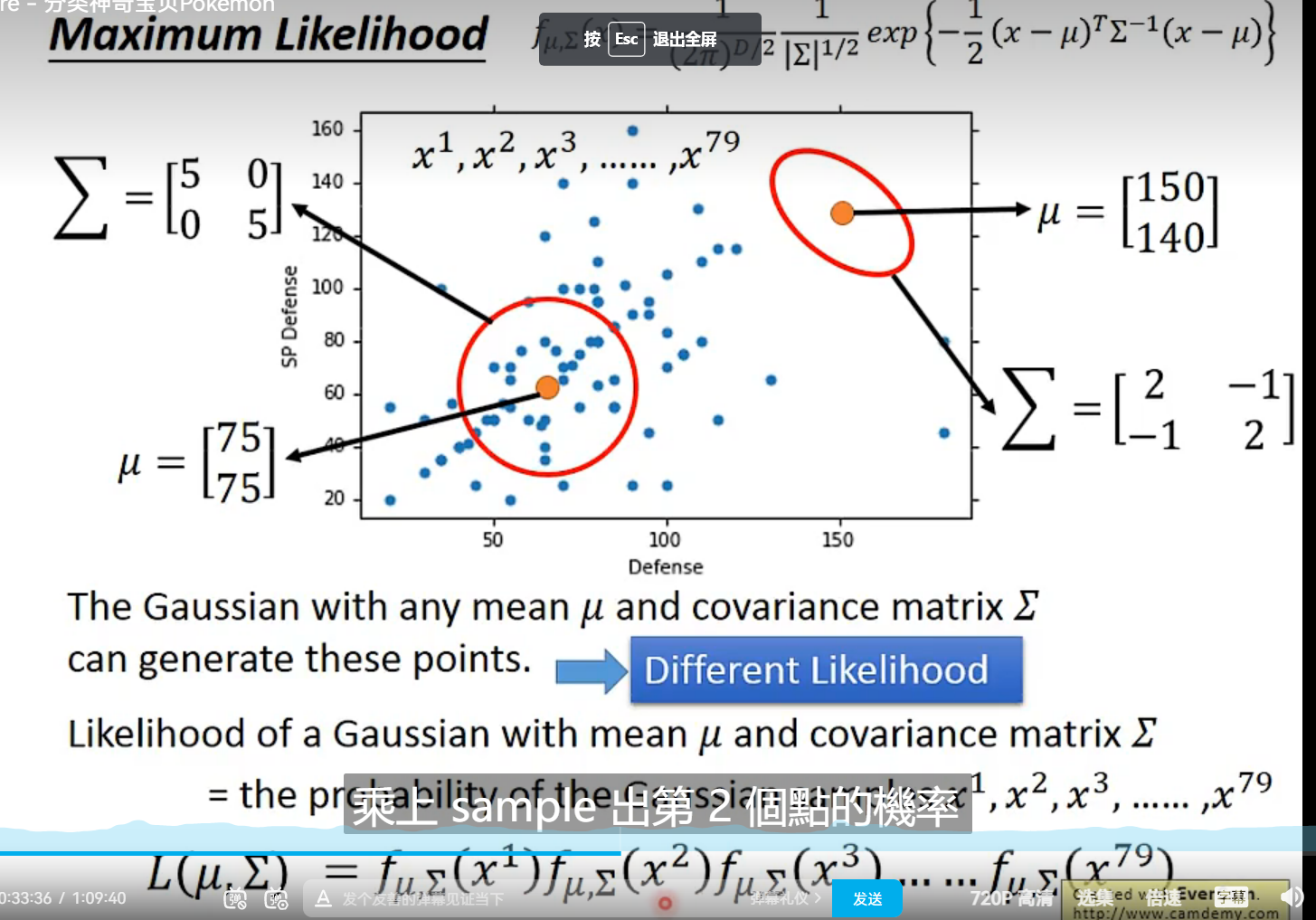
我们可以通过training data数据来得出一个上面这样的表达式



New点的值代入这个表达式，我们就可以得出他是这个样本的概率，尽管这个点不在这个区域内，

离中心点u越远，概率越低

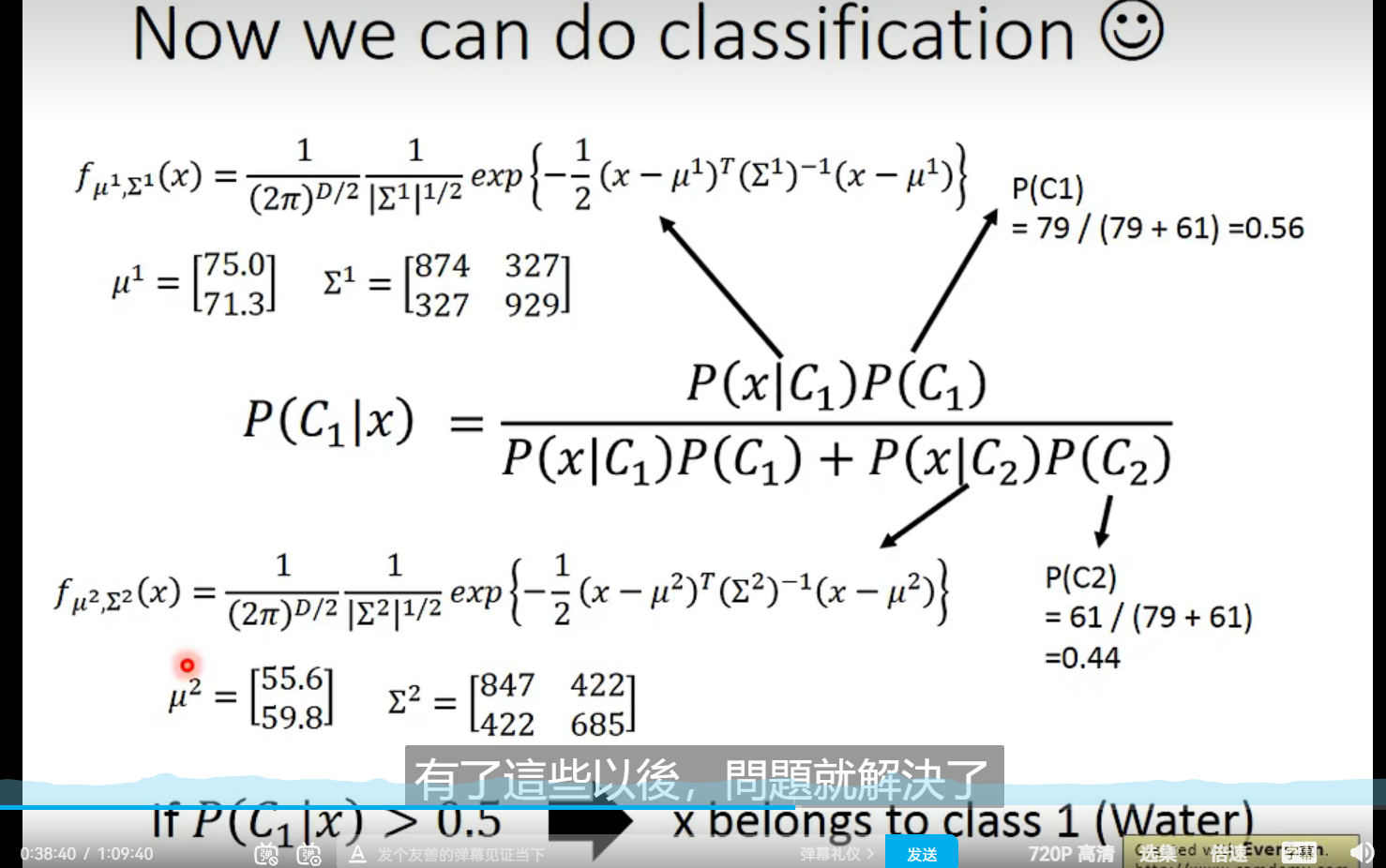
现在的问题就是如何通过training data 中找出这个未知参数

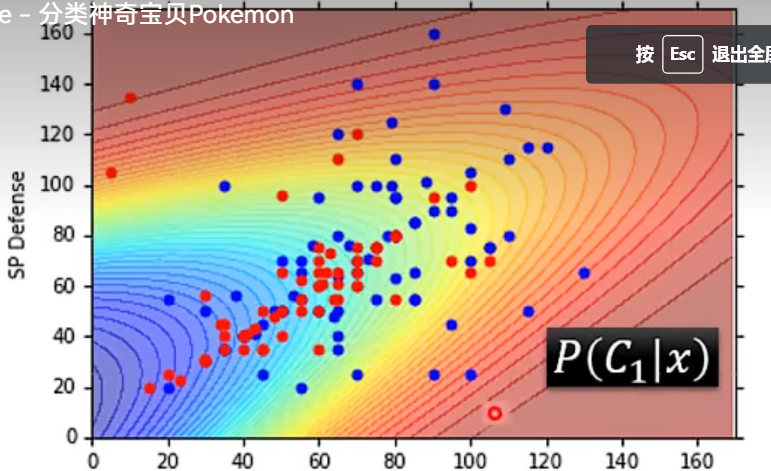




这些样本都是独立的

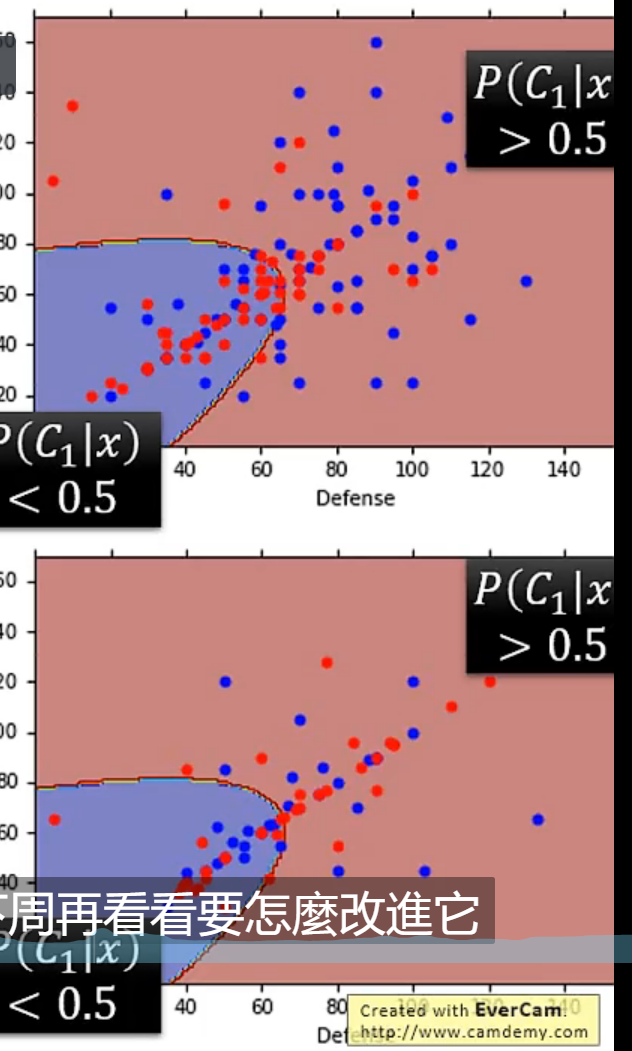
穷举所有的找出最大的，记下此时所对应的参数，这就是我们所需要找的参数，代入值就行了





红色区域概率大于0.5 ->属于水系

蓝色区域概率小于0.5->属于一般系

这还是仅仅是二维的，还可以上升到多维