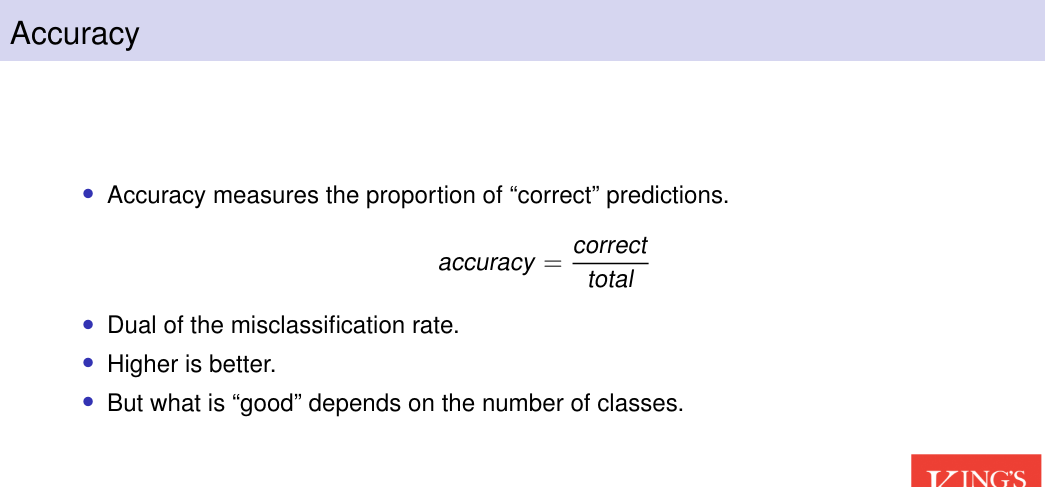
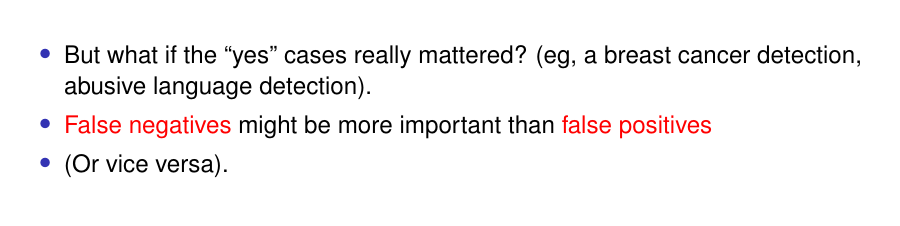
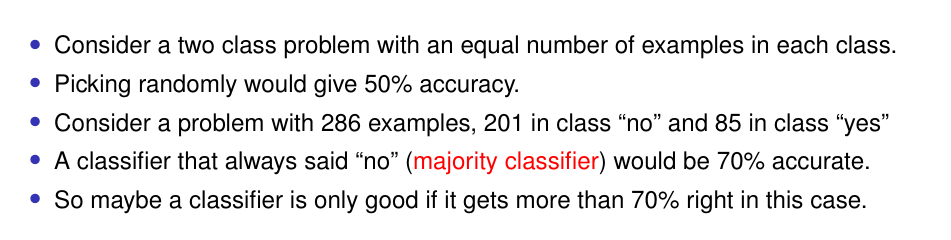
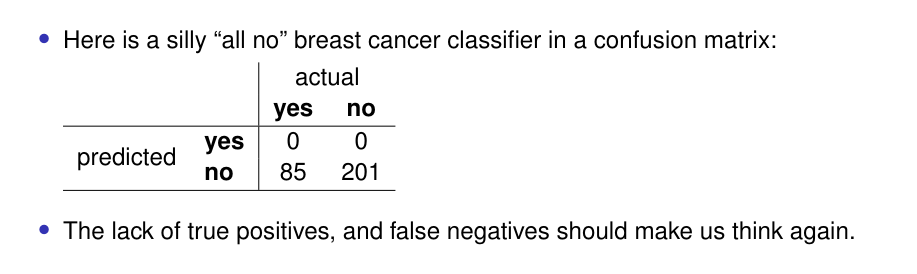
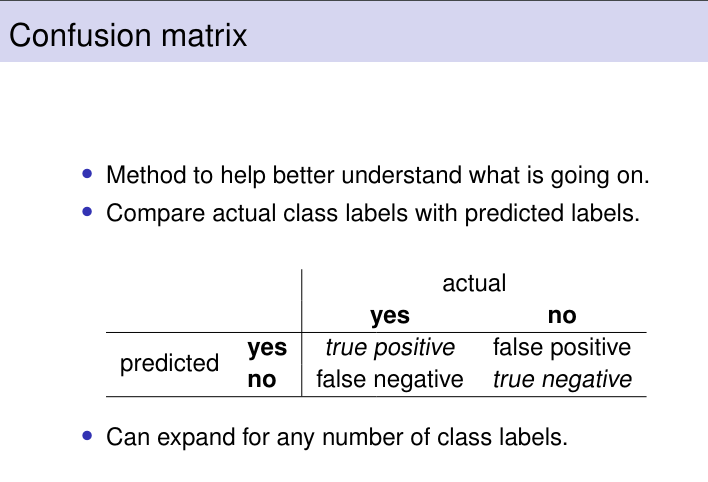
# Week 1

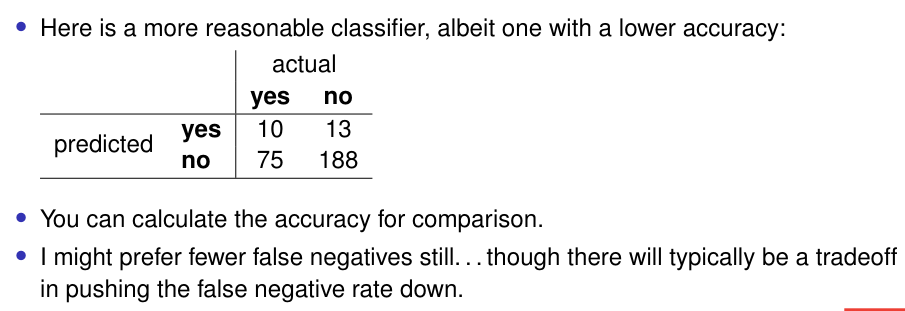
## Accuracy





## Confusion matrix



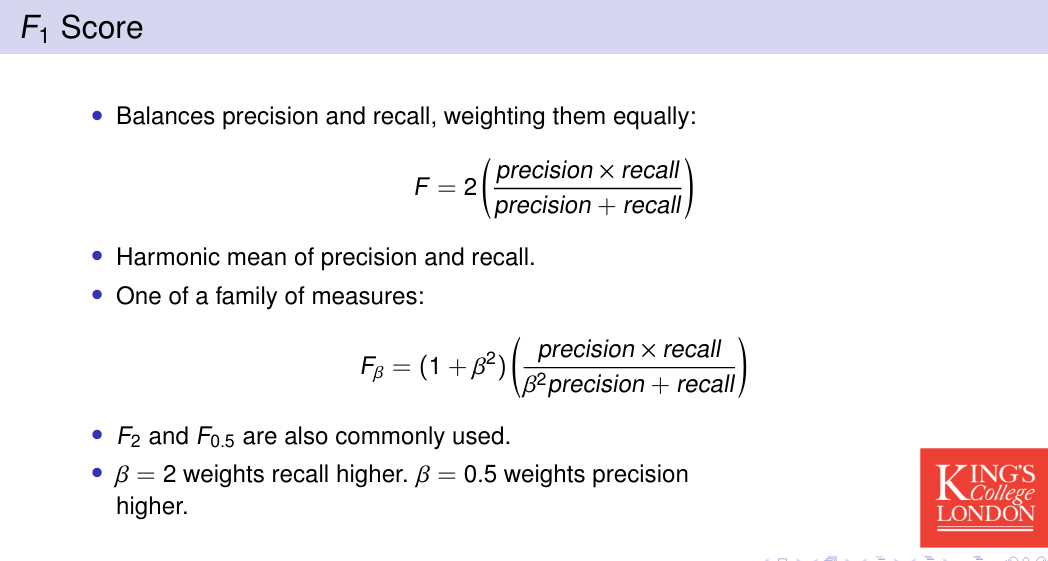


## Precision



## Recall

## F1 Score



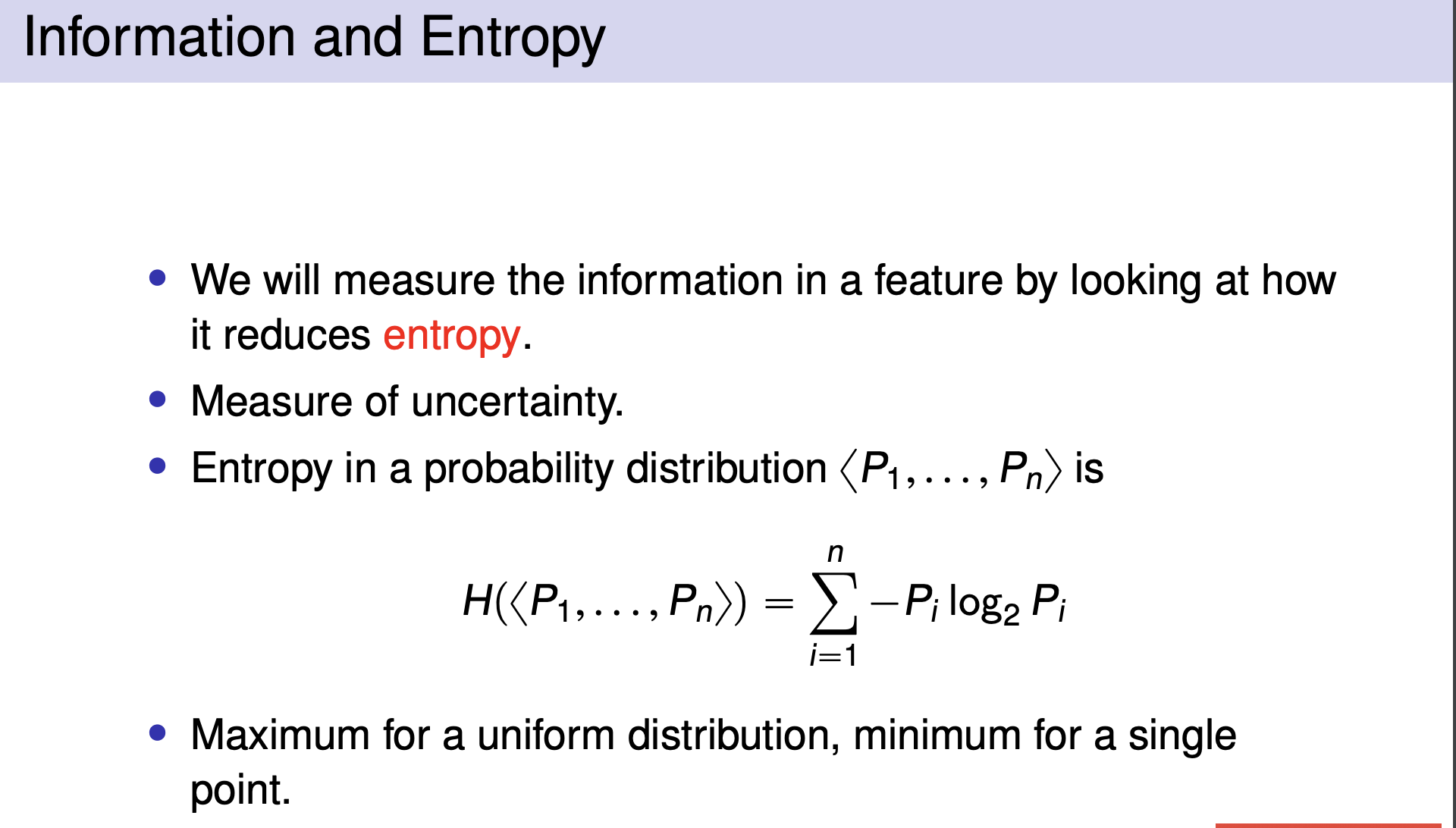
# Week 2

## Calculating log2

## 

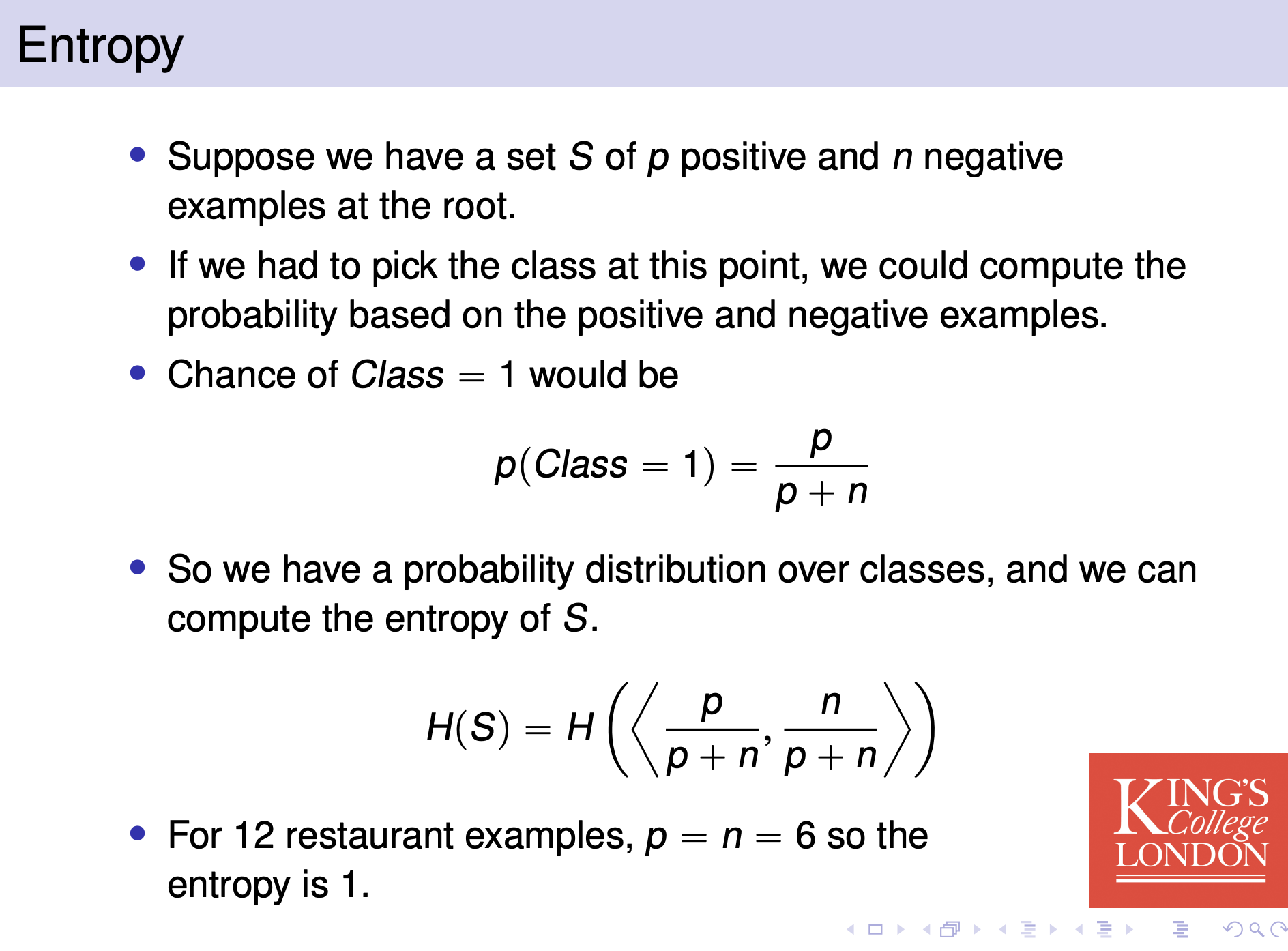
## 

## Entropy

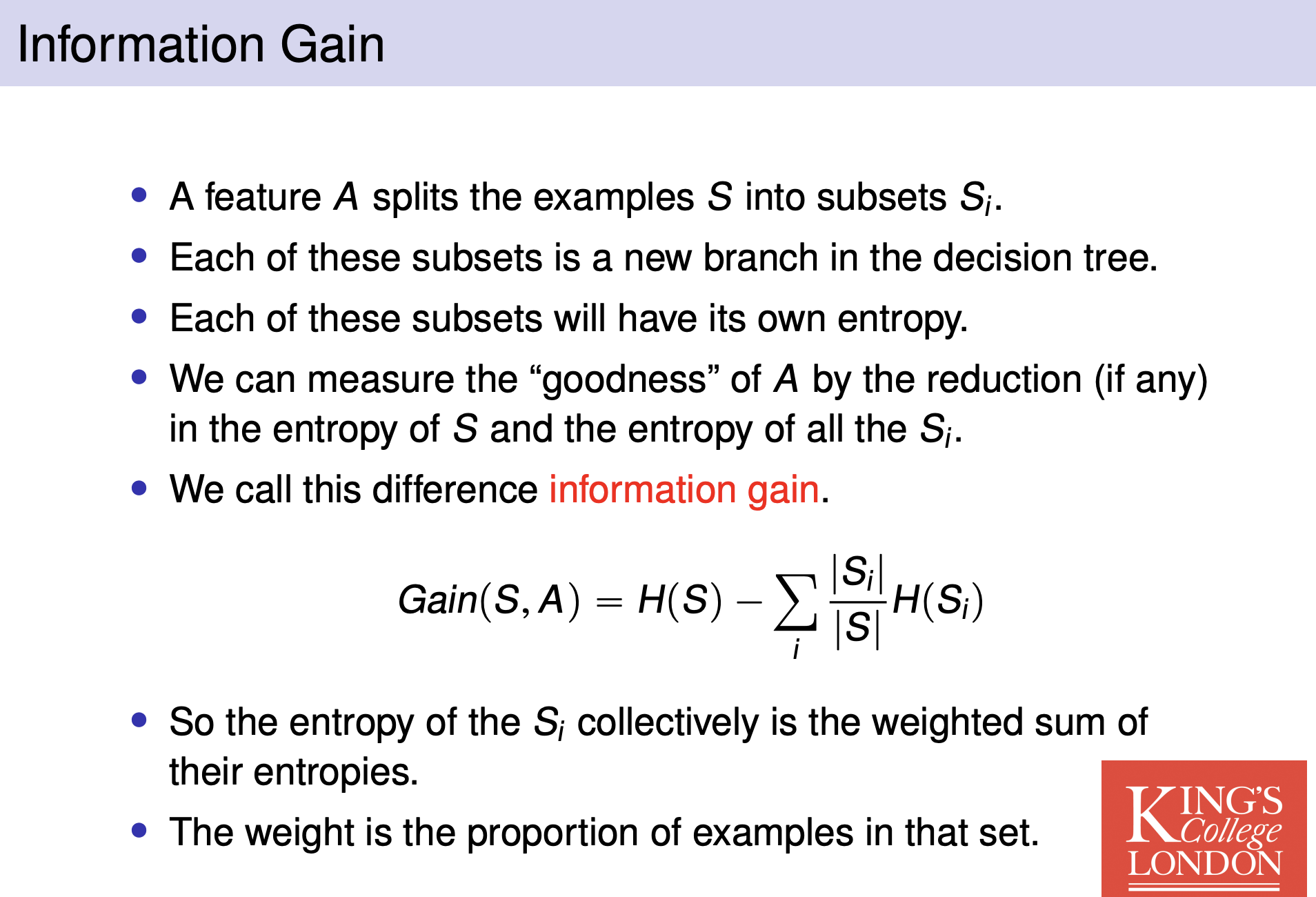


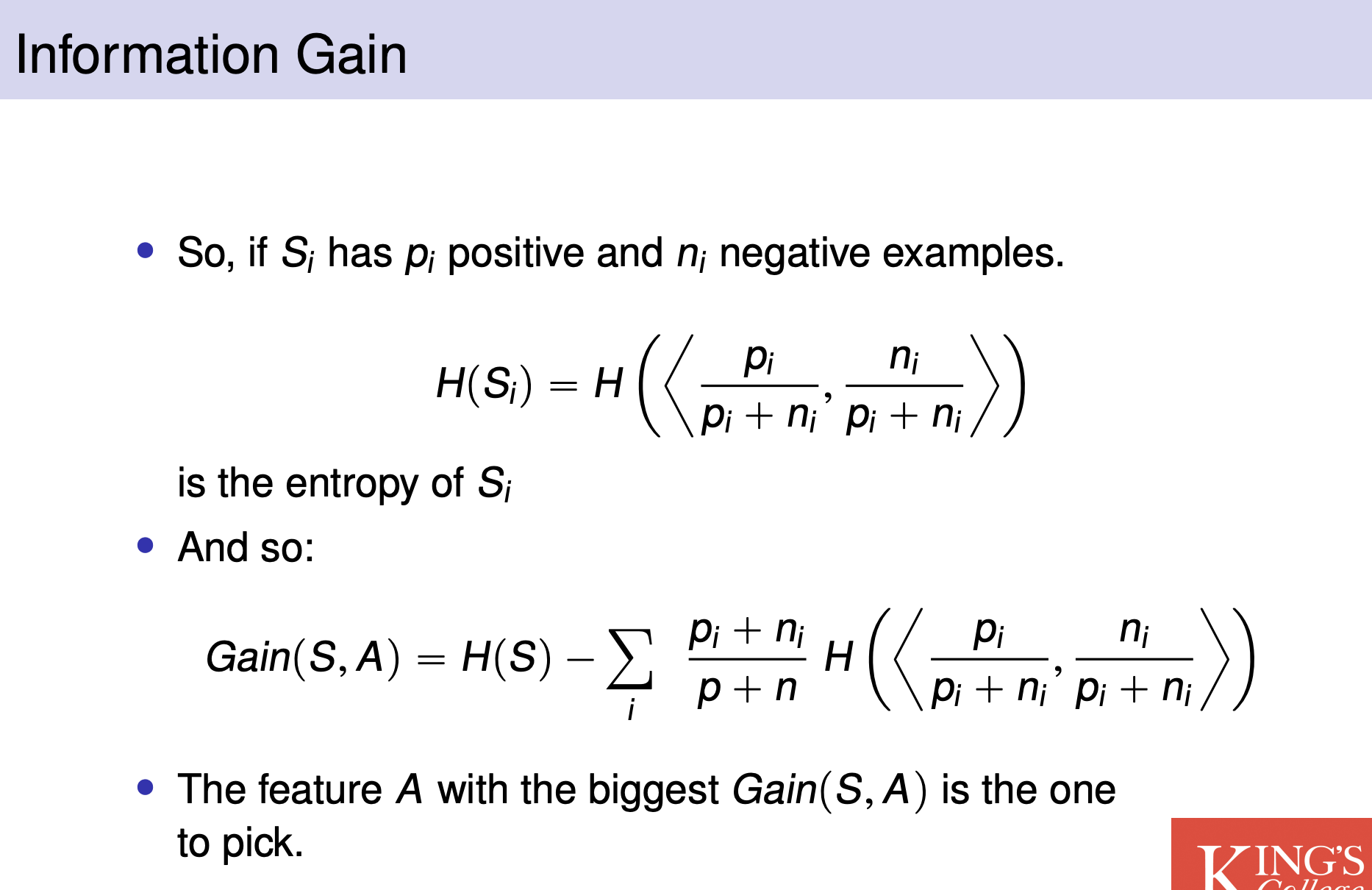
这个计算比较简单，比如说存在一个set，一共有6个variable，两个positive，四个negaitve，计算方法则是 -1 \* (⅓ log2 ⅓ + ⅔ log2 ⅔ ）（因为这里一共有6个variable，其中两个是positive占了三分之一，剩下四个是negative，占了三分之二）最终得到差不多0.918

熵越低越好，最好是0

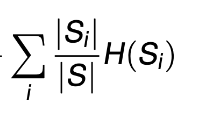


### Information gain

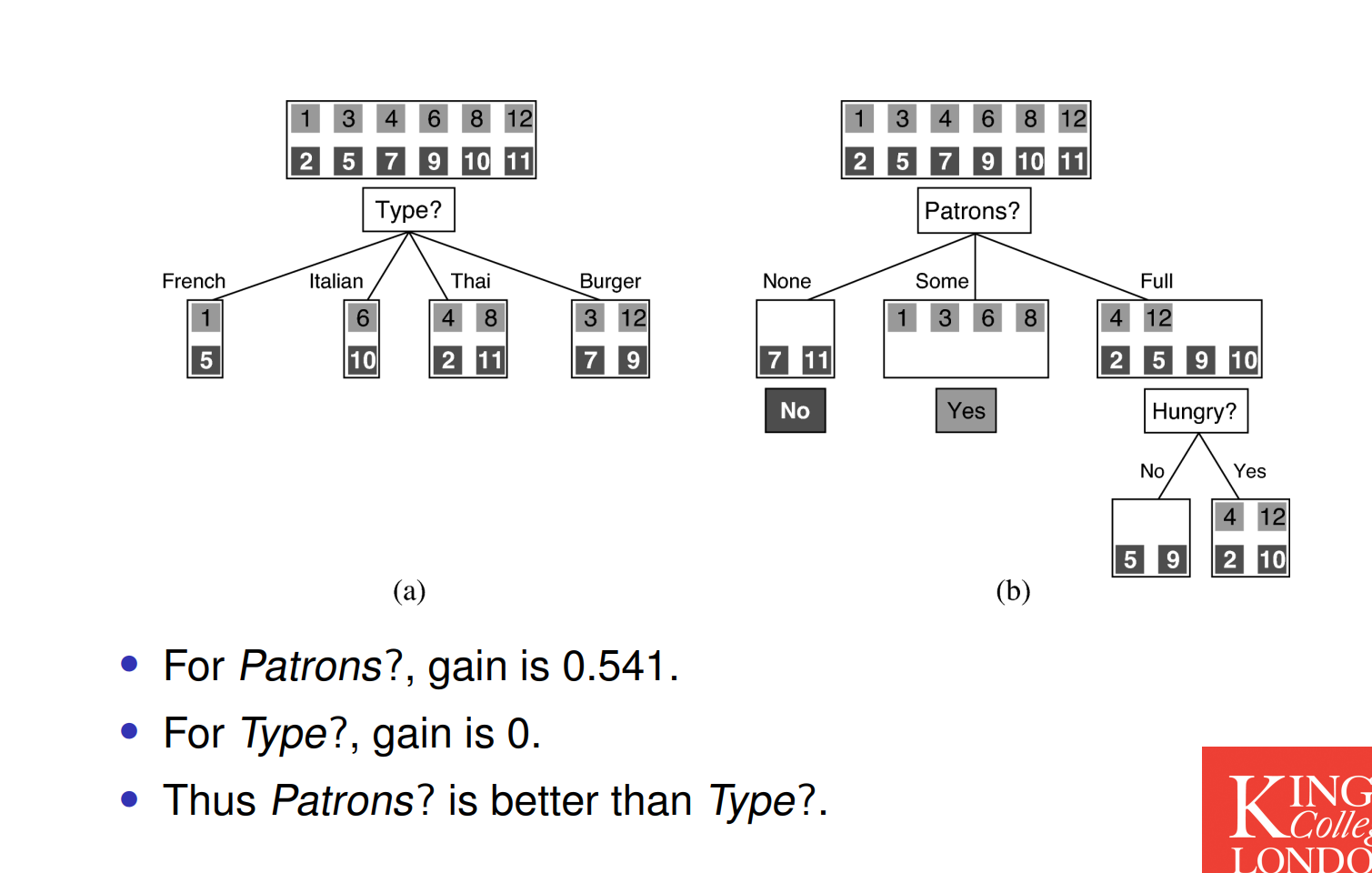




这里ppt做的比较狗屎，其实本质上就是旧的entropy减去新的entropy，上述的内容都可以参考下面的一个例子。

这里的其实指的是，假如这个set一共有12个variable，这12个变量又被分成三个小组，可能第一组两个，第二组四个，第三组六个。每个小组都能计算出一个自己的熵，最终的熵就是把每个小组的熵按比例加起来。比如第一个组占了2/12 = ⅙，因此他们组的熵（假设这里得到0.7）将会乘以⅙，也就是0.7 x ⅙ = 0.116，依此类推。

### Entropy例子

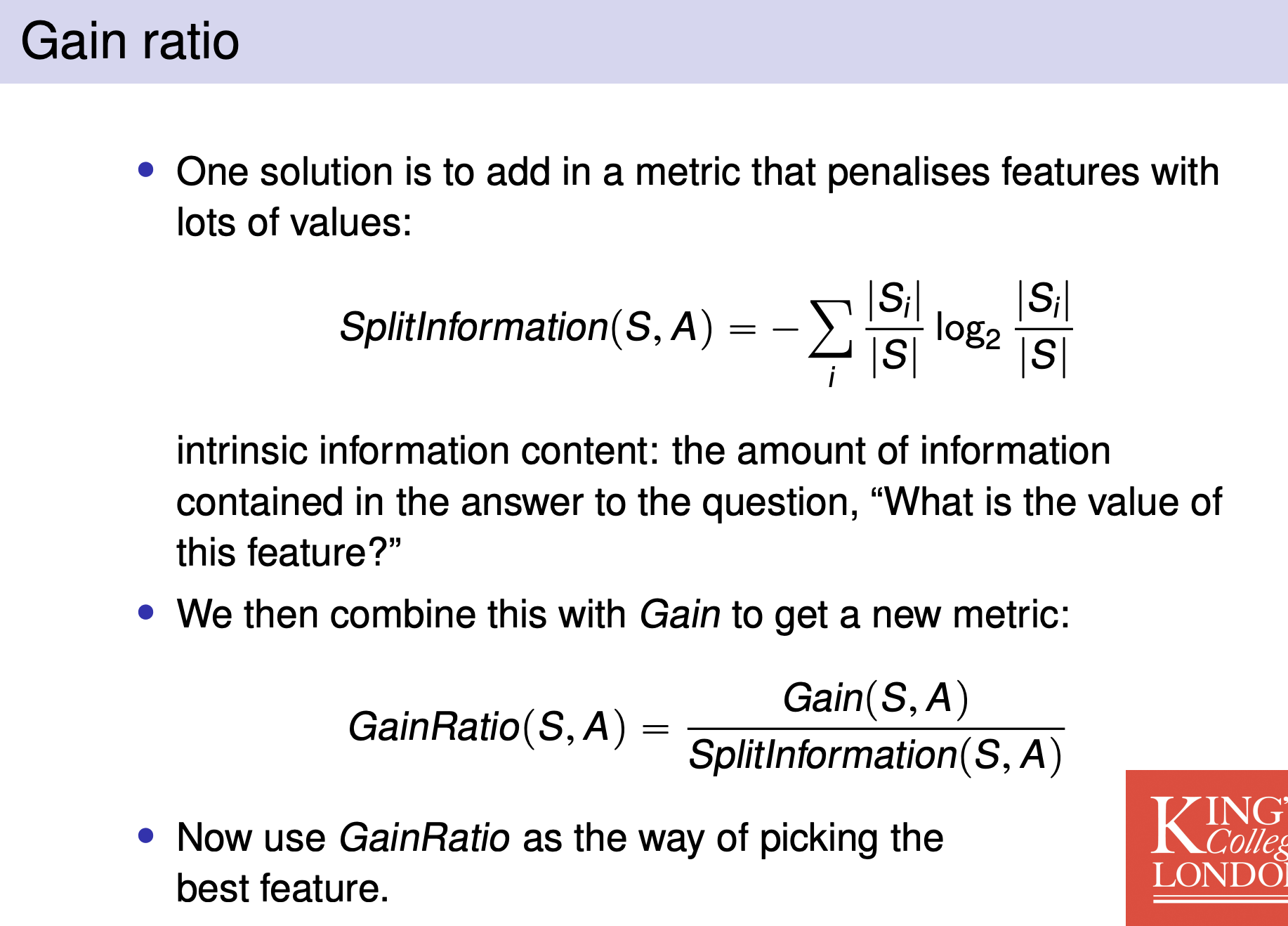


图a因为每一组都有相同数量的positive 和 negative因此其实就是按比例加起来最后得到1,（⅙ \* 1 + ⅙ \* 1 + ¼ \* 1 + ¼ \* 1）

图b则不一样，首先前两组要不就是只有positive，要不就是只有negative，因此他们的熵都是0，第三组的计算方法则是 -1 \* (⅓ log2 ⅓ + ⅔ log2 ⅔ ）（因为这里一共有6个variable，其中两个是positive占了三分之一，剩下四个是negative，占了三分之二）得到差不多0.918，但是这个数字我们还需要乘以他的比例，这里他有6个variables，占了总数的一半，因此就是乘以0.5，得到0.459。 同时因为其他两组都是0，因此这个分类的最终的熵就是0.459，而gain就是1-0.459 = 0.541   
Gain is 1 is because the original set has 6 positive and 6 negative so the entropy for the entire set will be 1, however is set arrange differently then the entropy will be different, and we will use different number for H(s)

补充：上面的positive和negative指的是最终的结果，比如在这个例子中我们想要知道是否选择这家餐厅吃饭，positive就代表最终选择了吃，反正则是不吃

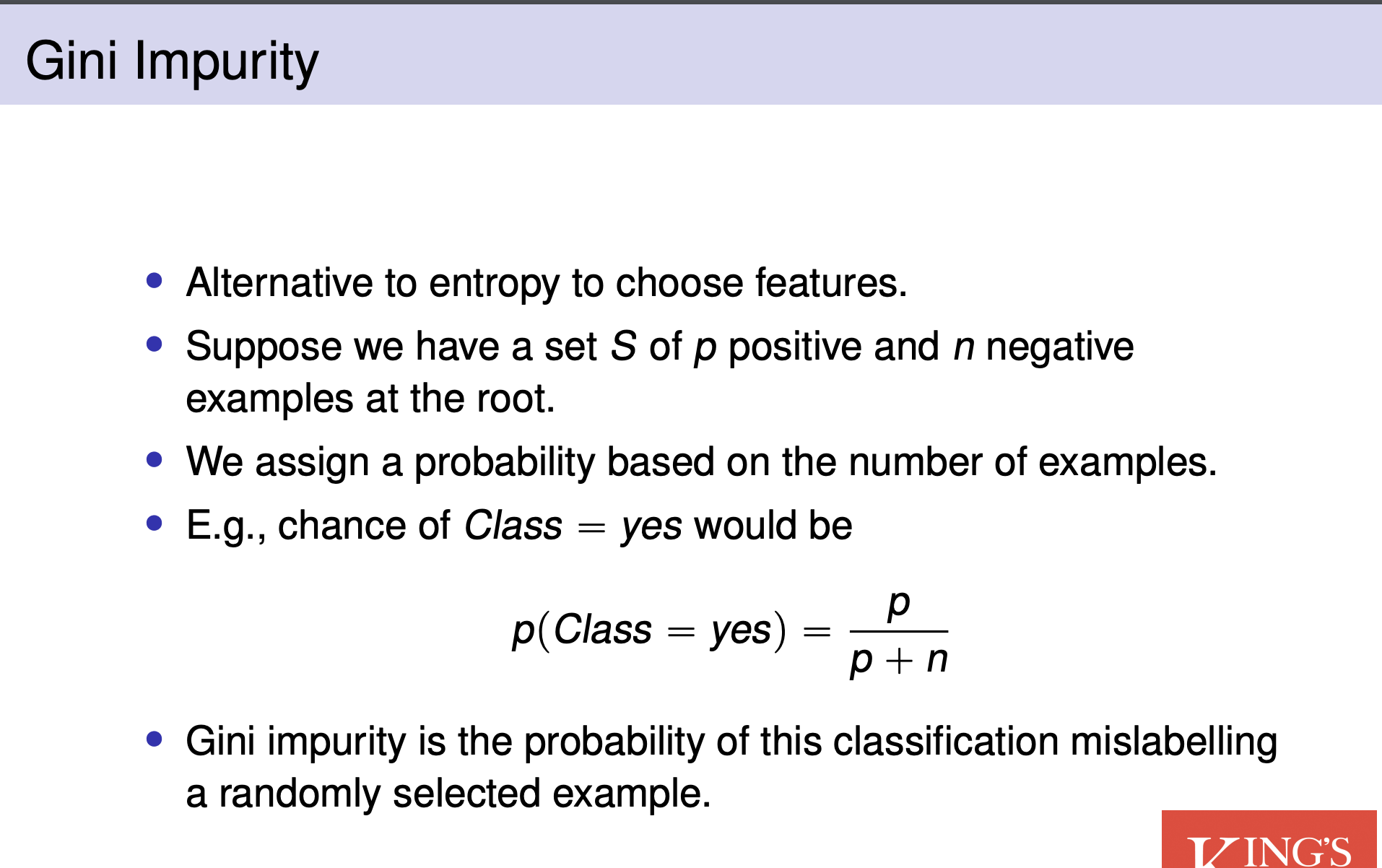
## Gain ratio

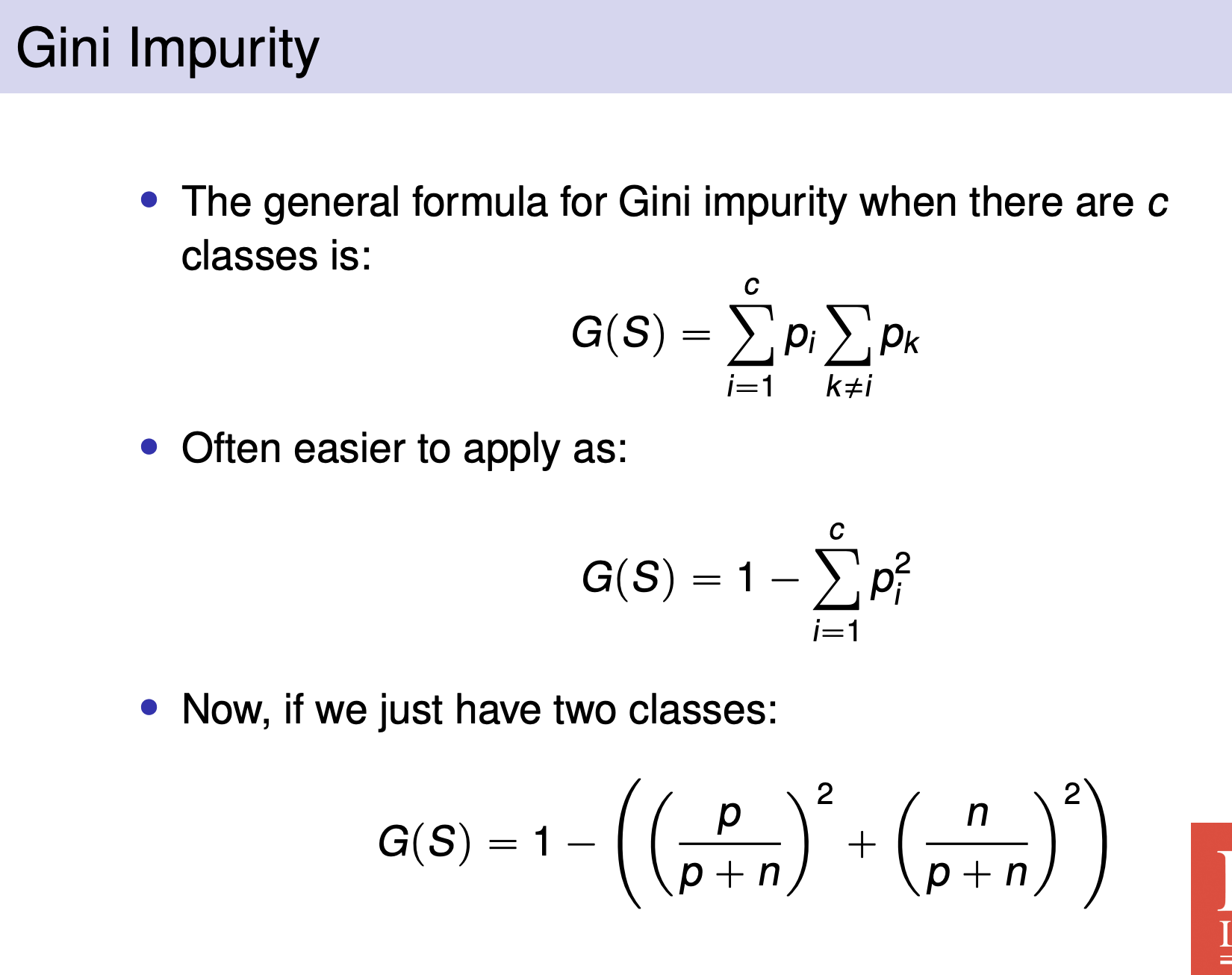


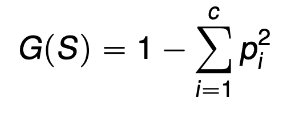
这个例子容易和熵计算混淆，还是之前的上面的example，我们已经计算出gain是0.541， 然后上面根据variable的分布来计算，三个组的概率分别是⅙，⅓和½，因此我们需要对每个概率应用splitinformation的公式，比如说⅙ 就是 -(⅙ log2 ⅙)，也就得到了0.431，同样对剩下两个概率进行同样的计算得到0.528和0.5，最后把三个数字相加就得到了分裂信息的值也就是1.459，而gain ratio的计算要把gain的值也就是0.541除以1.459得到0.371

Gain ratio越大越好

## Gini impurity

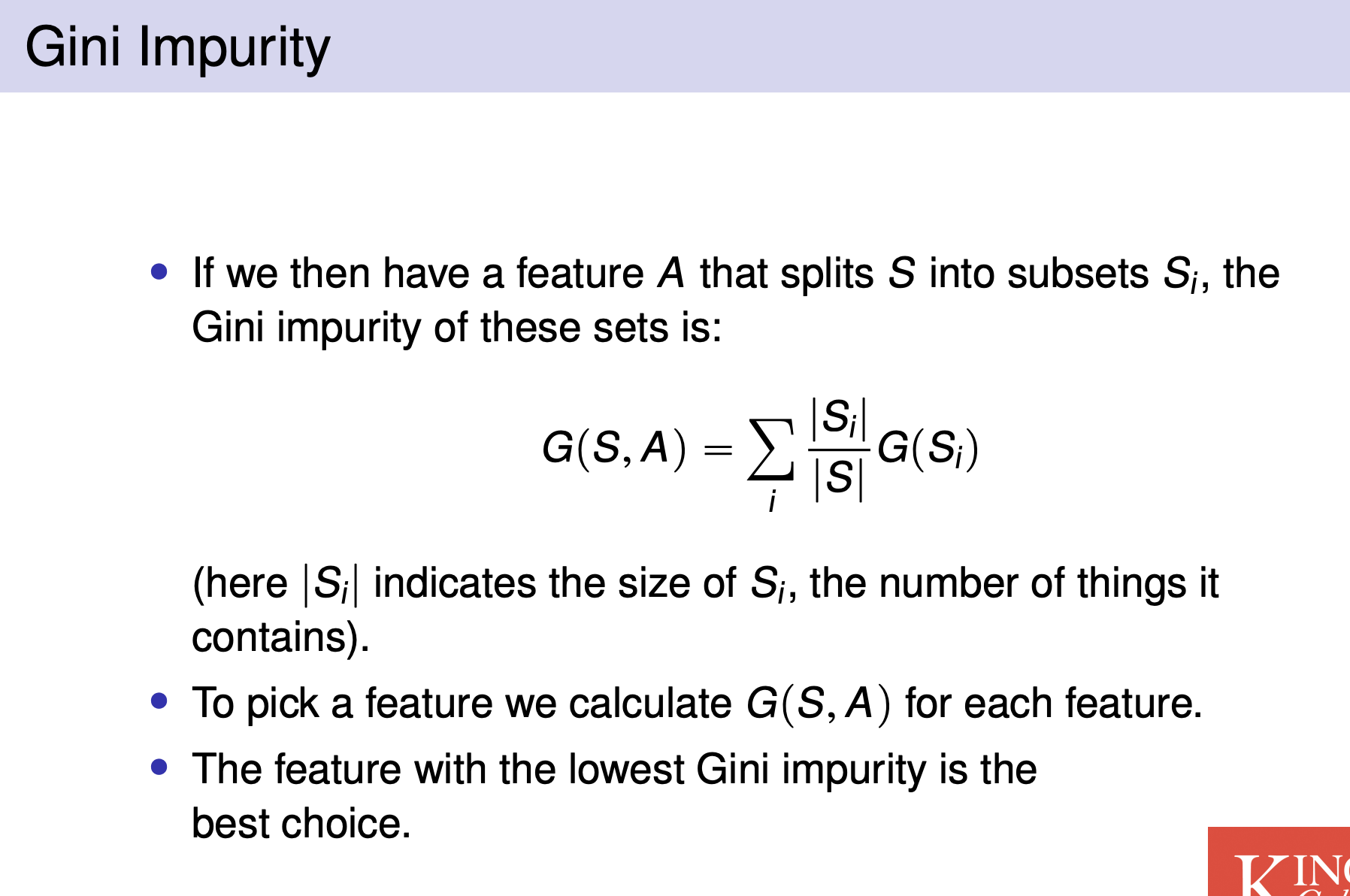




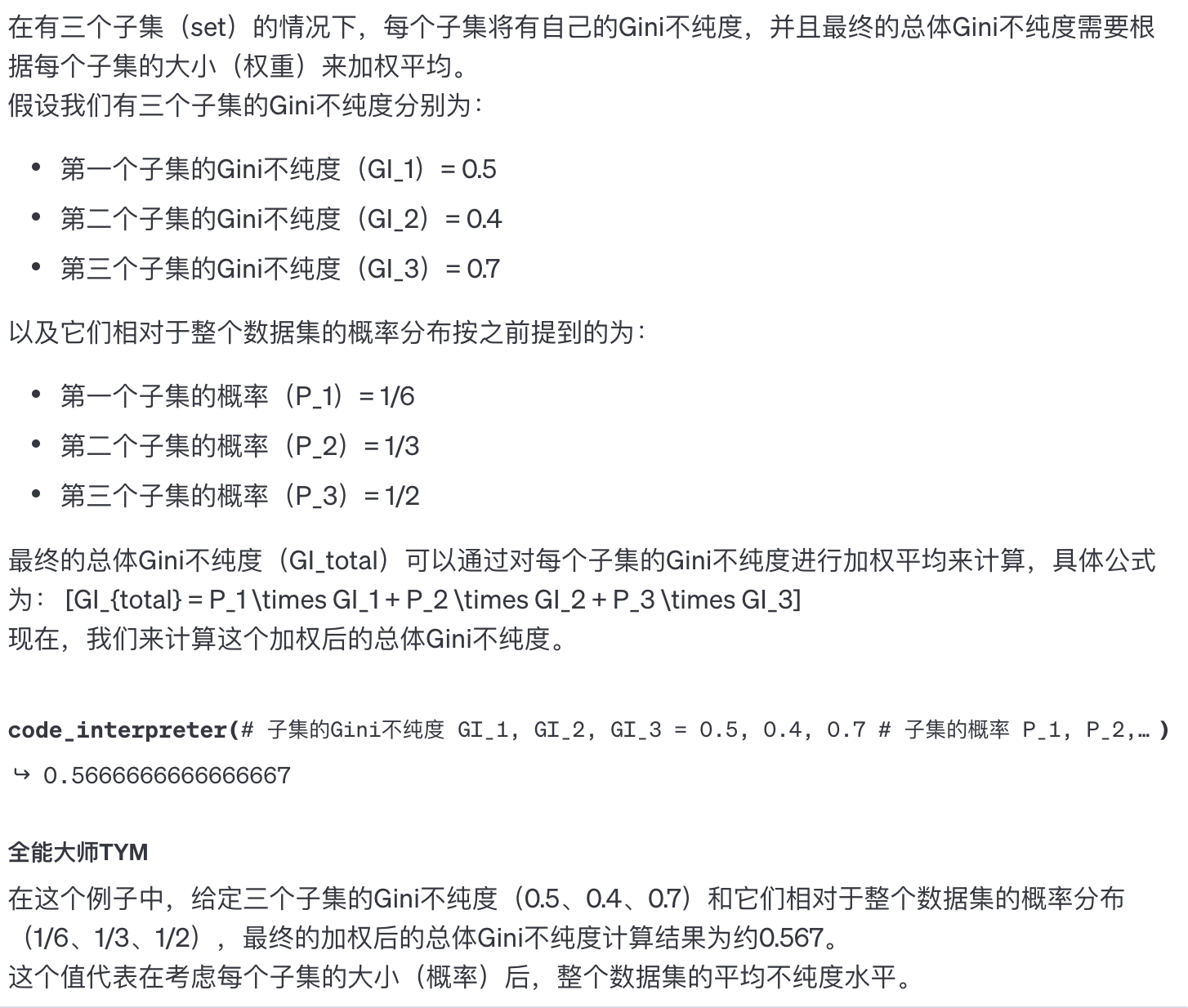
如果pk写作（1-pi），则公式可以写成这样子

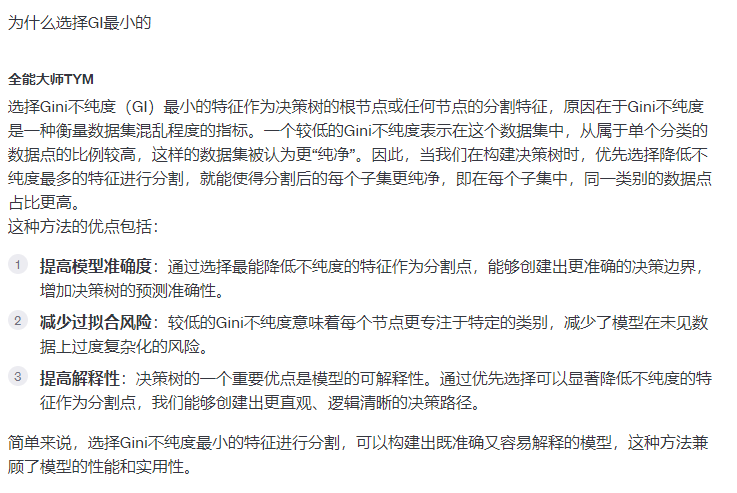
具体计算方法如下，让我们计算一个具体的例子：假设一个数据集中有60%的样本属于类别A（(p\_A = 0.6)），40%的样本属于类别B（(p\_B = 0.4)），计算这个数据集的Gini不纯度。

也就是1-（）=0.48

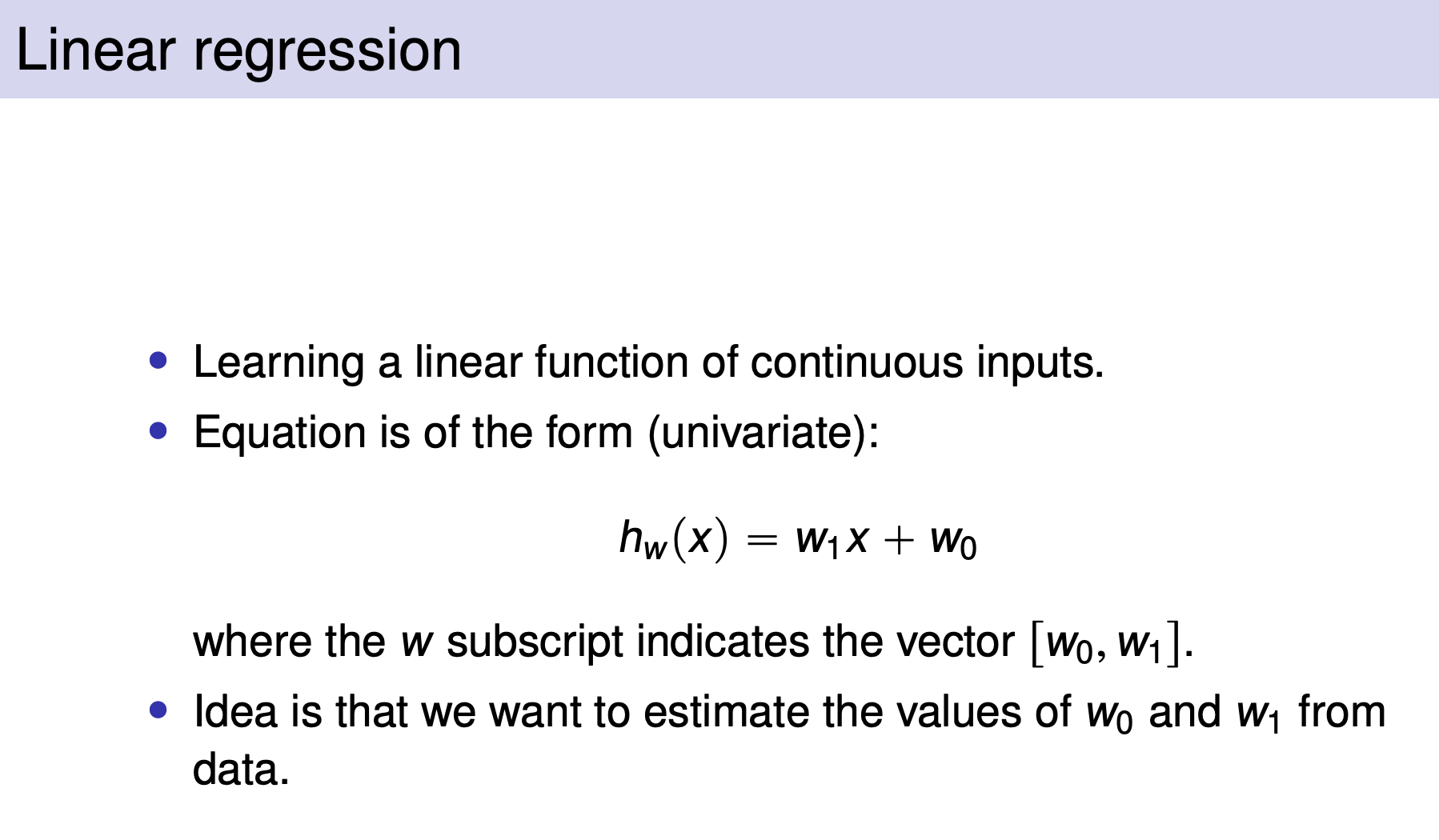


还是用之前的例子：





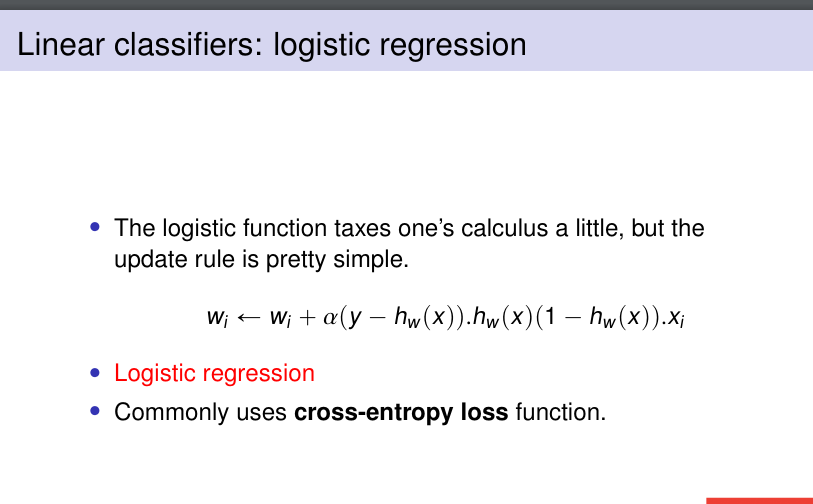
## Linear regression



这个其实就是y=mx+c

## Logistic function

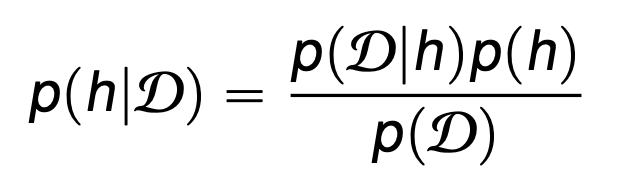
## 

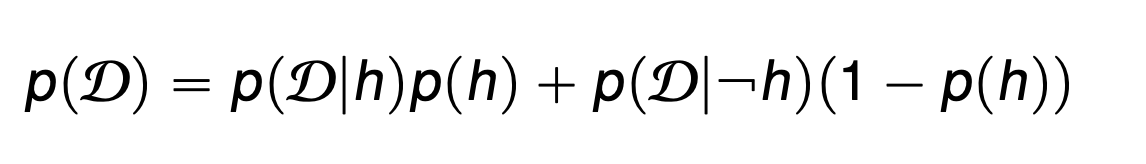
<https://www.youtube.com/watch?v=9zw76PT3tzs&ab_channel=ritvikmath>

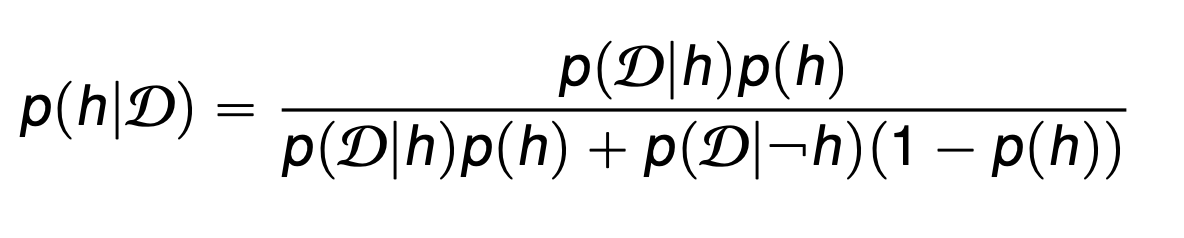
## 

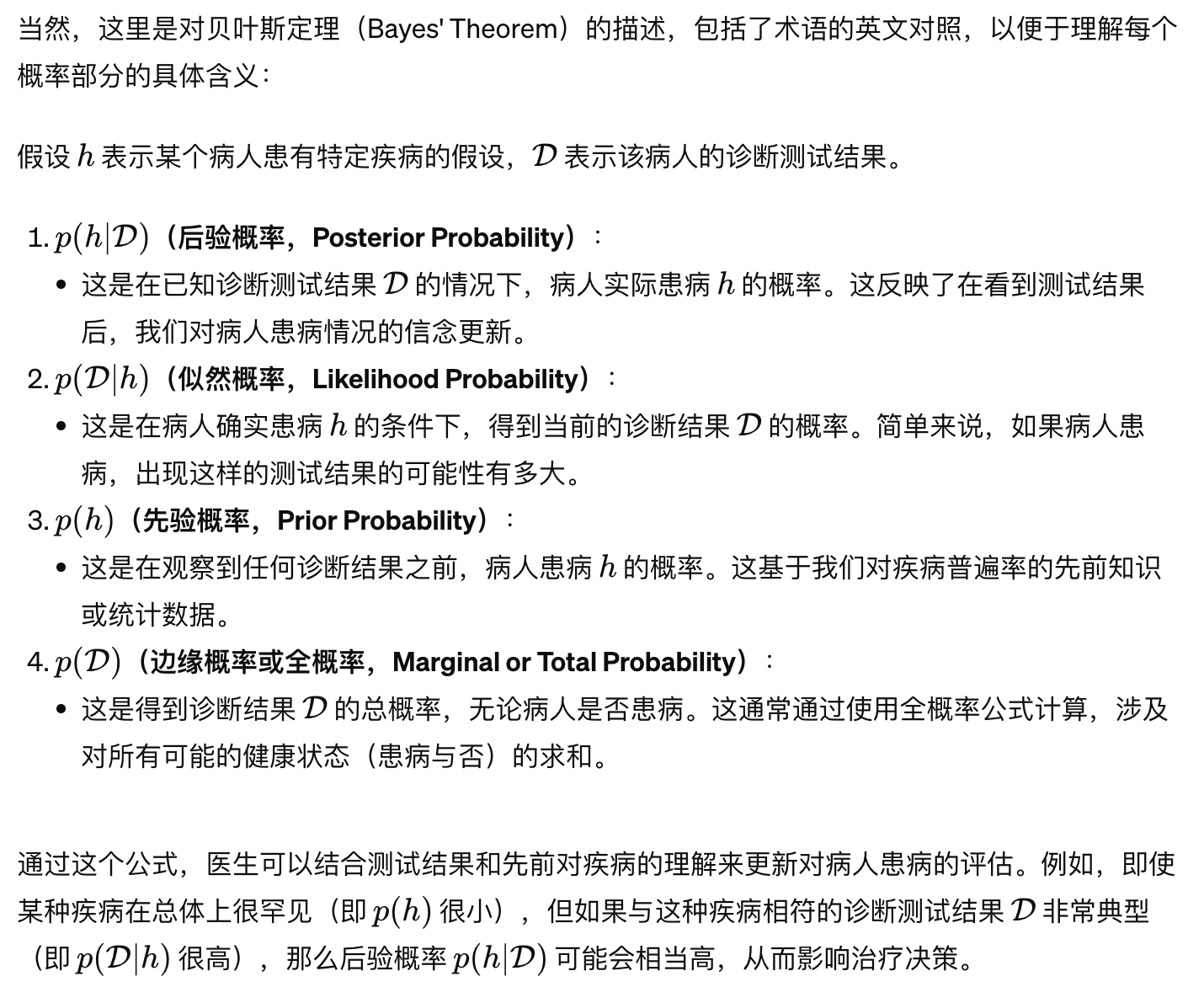
# Week 3

## Bayes

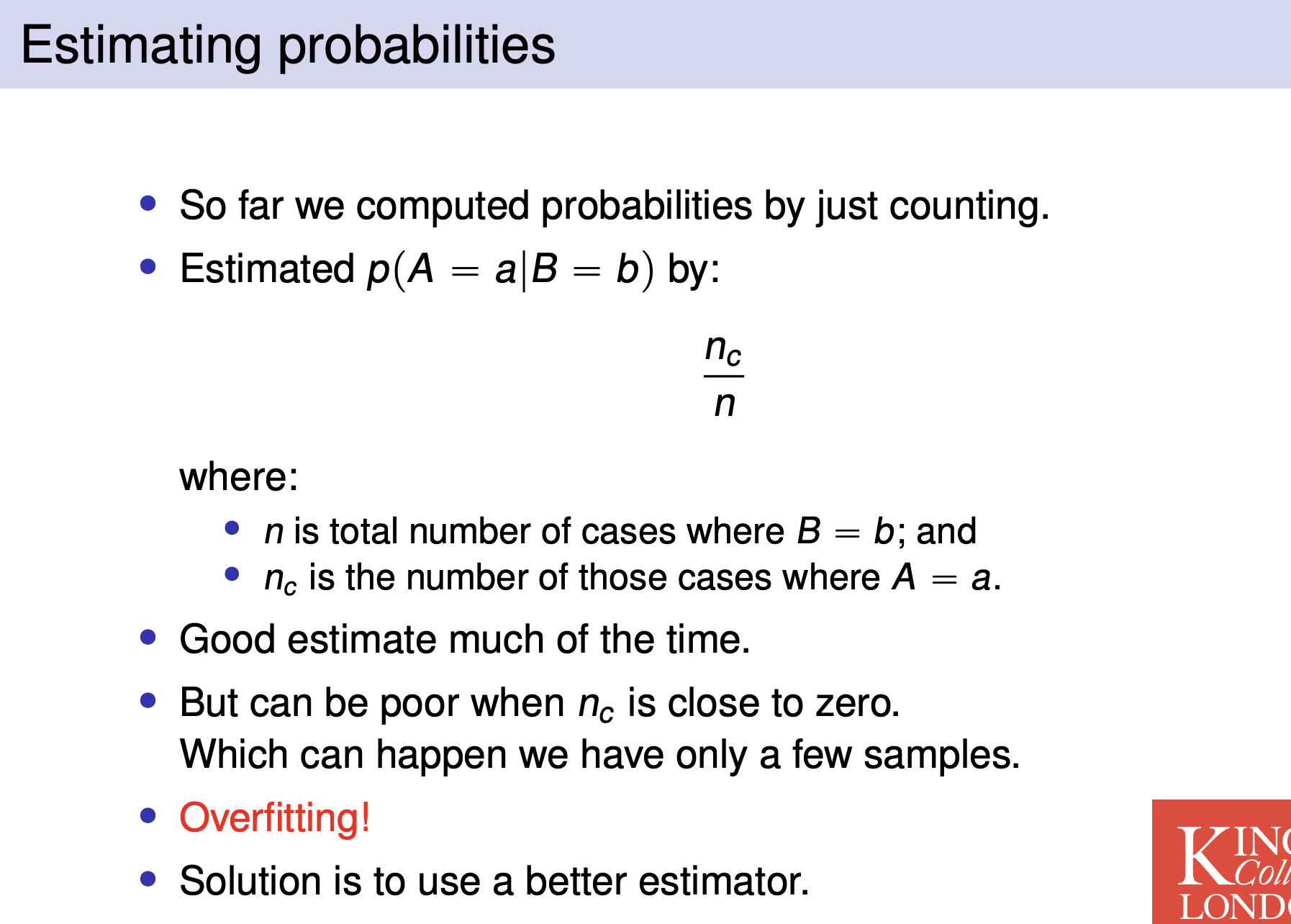


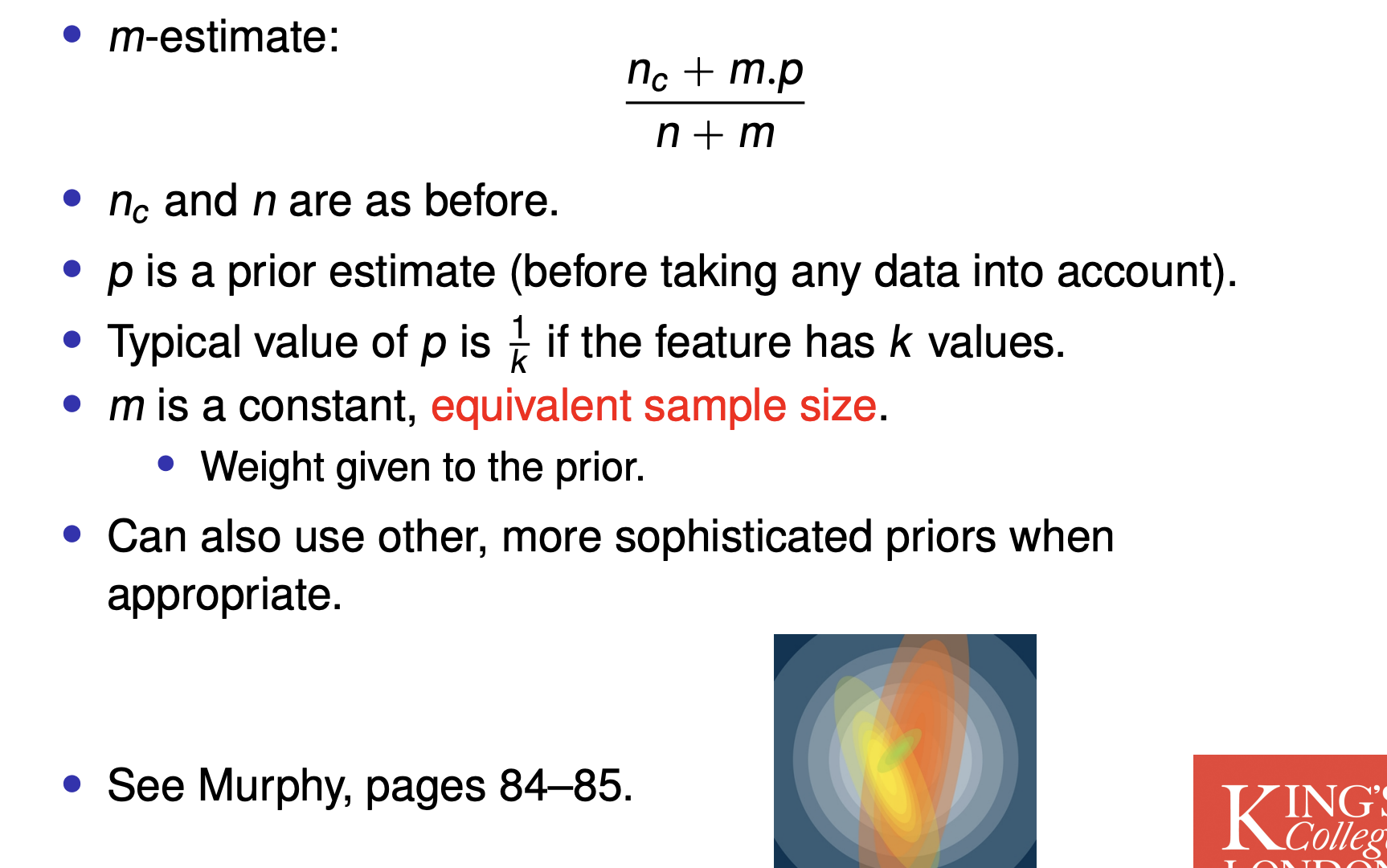




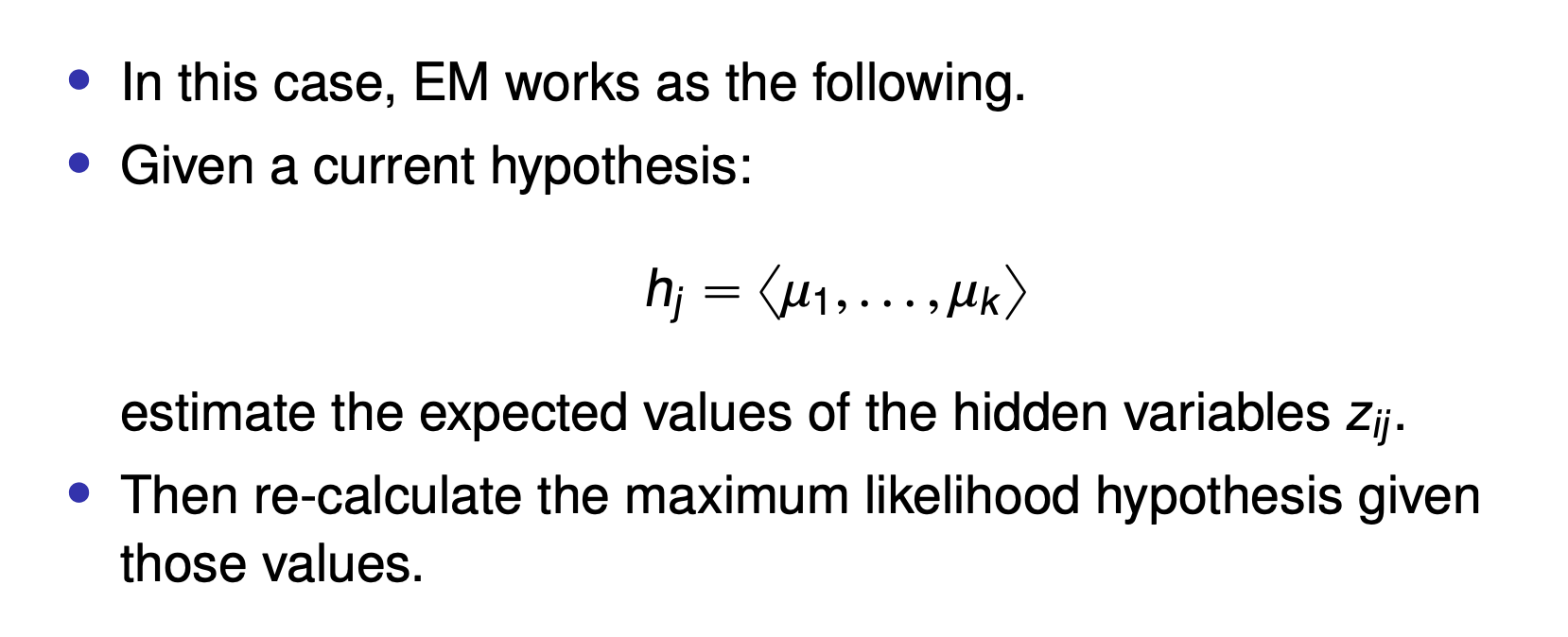


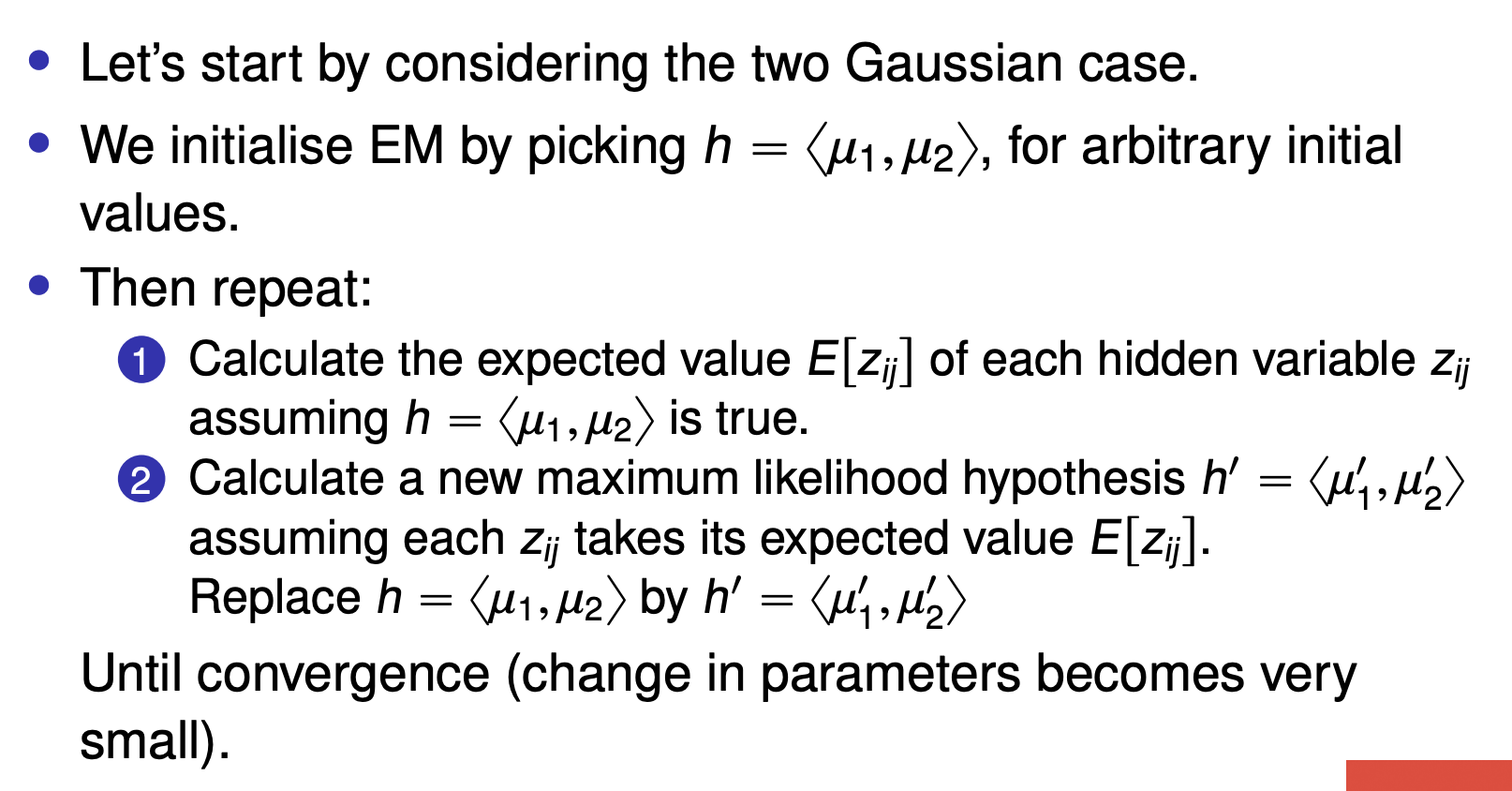
## M estimate

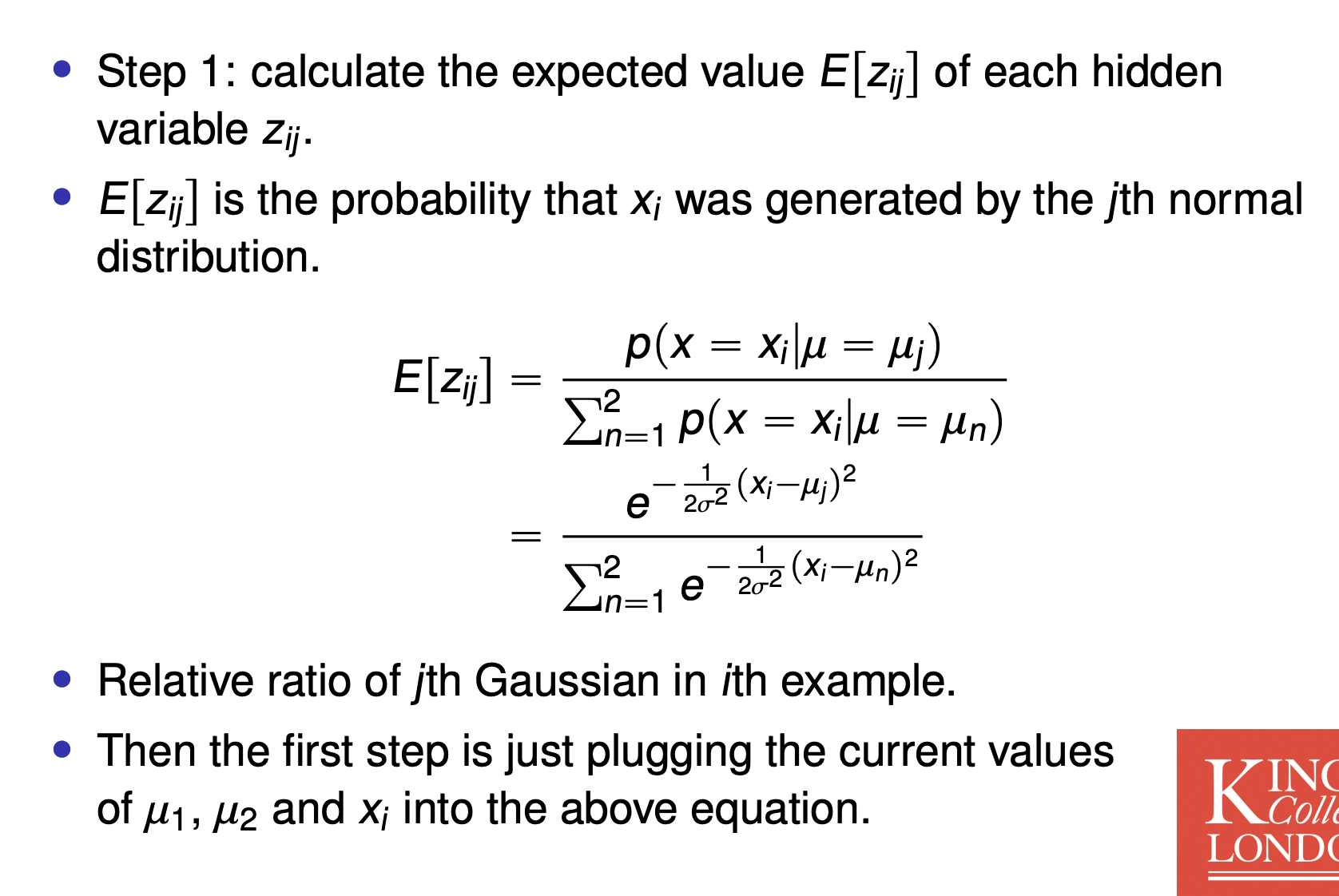




## Expectation Maximisation







## K means distance calculation

manhattan



Euclidean

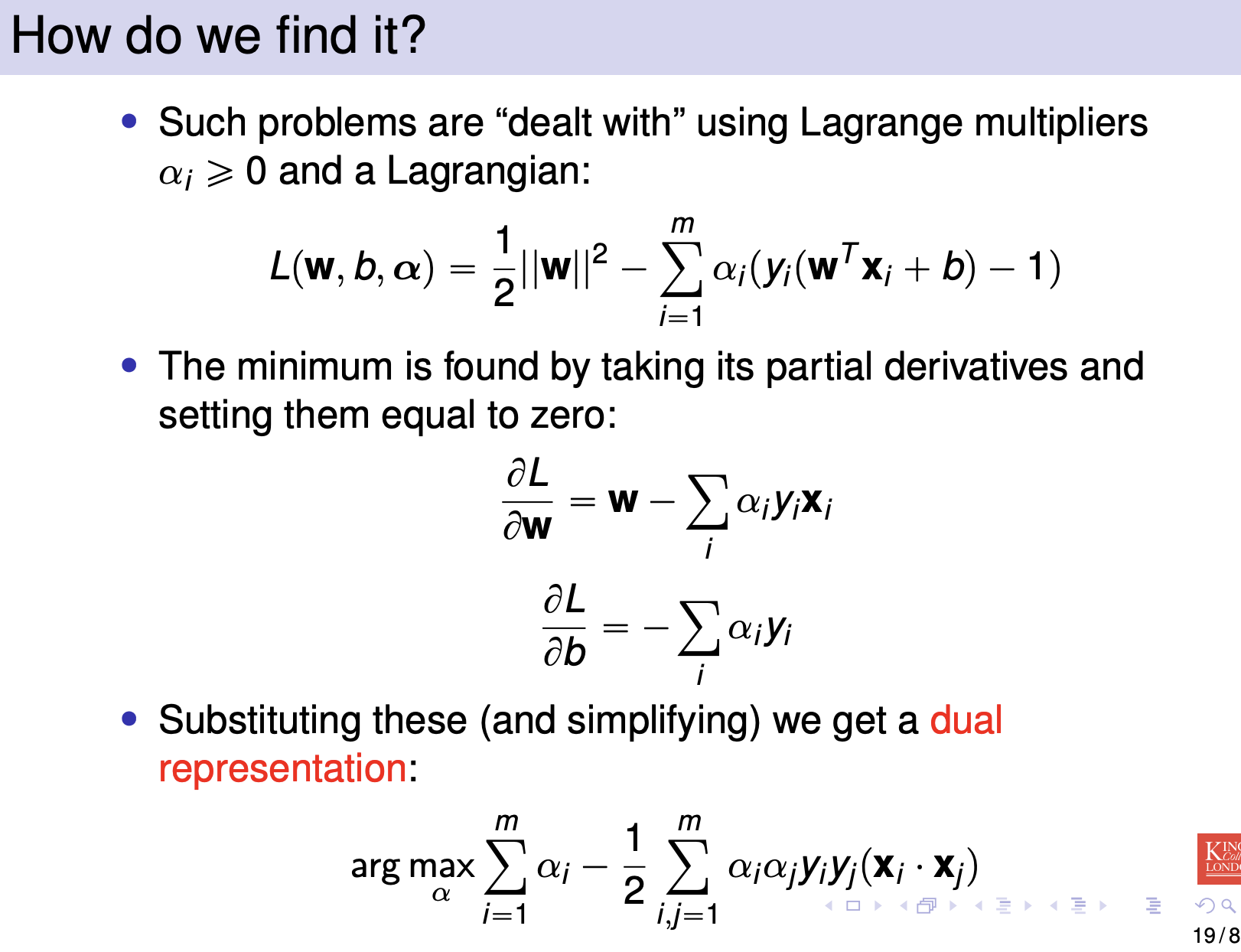
d = √[(x2 – x1)2 + (y2 – y1)2]

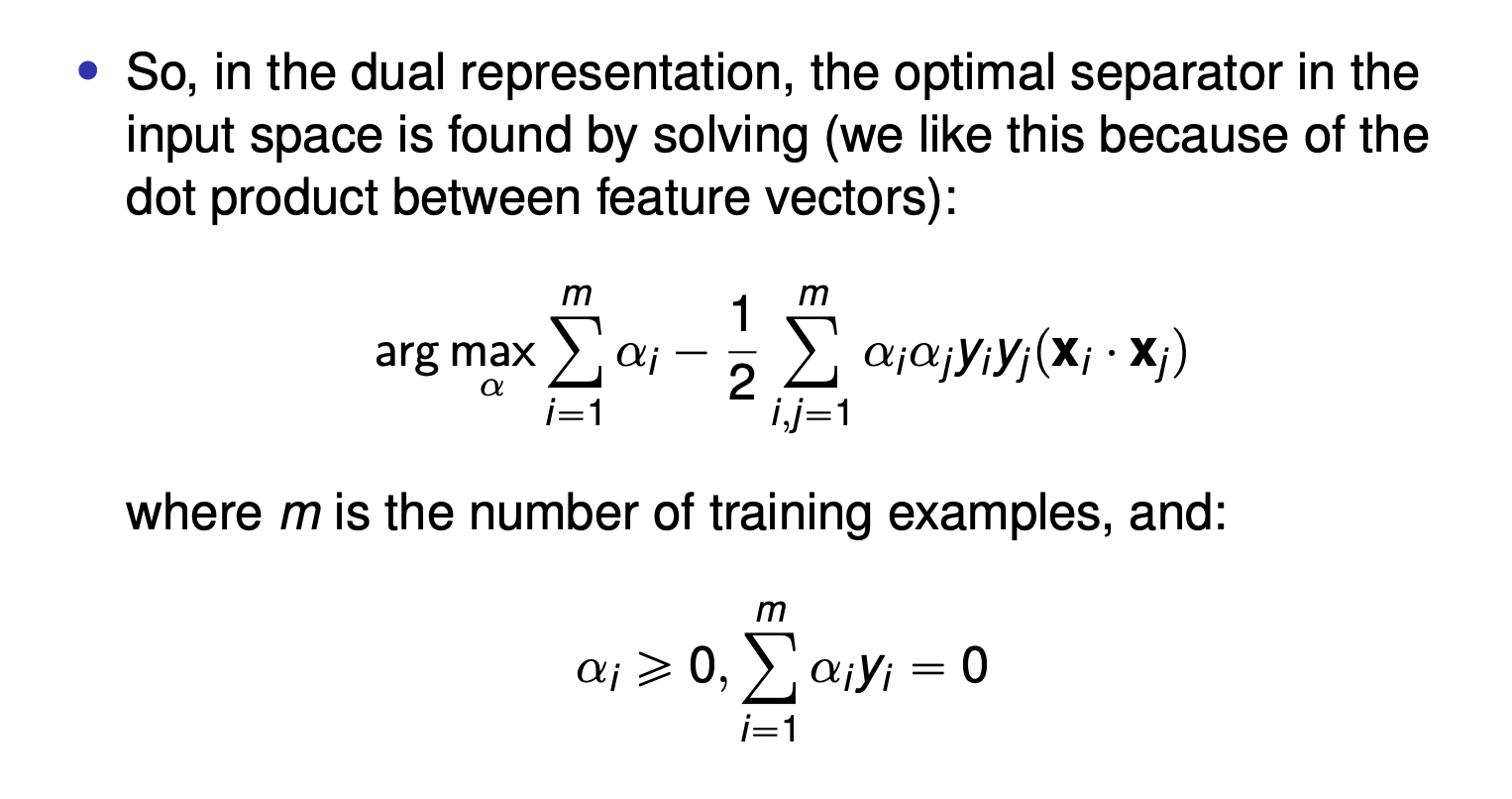
# 

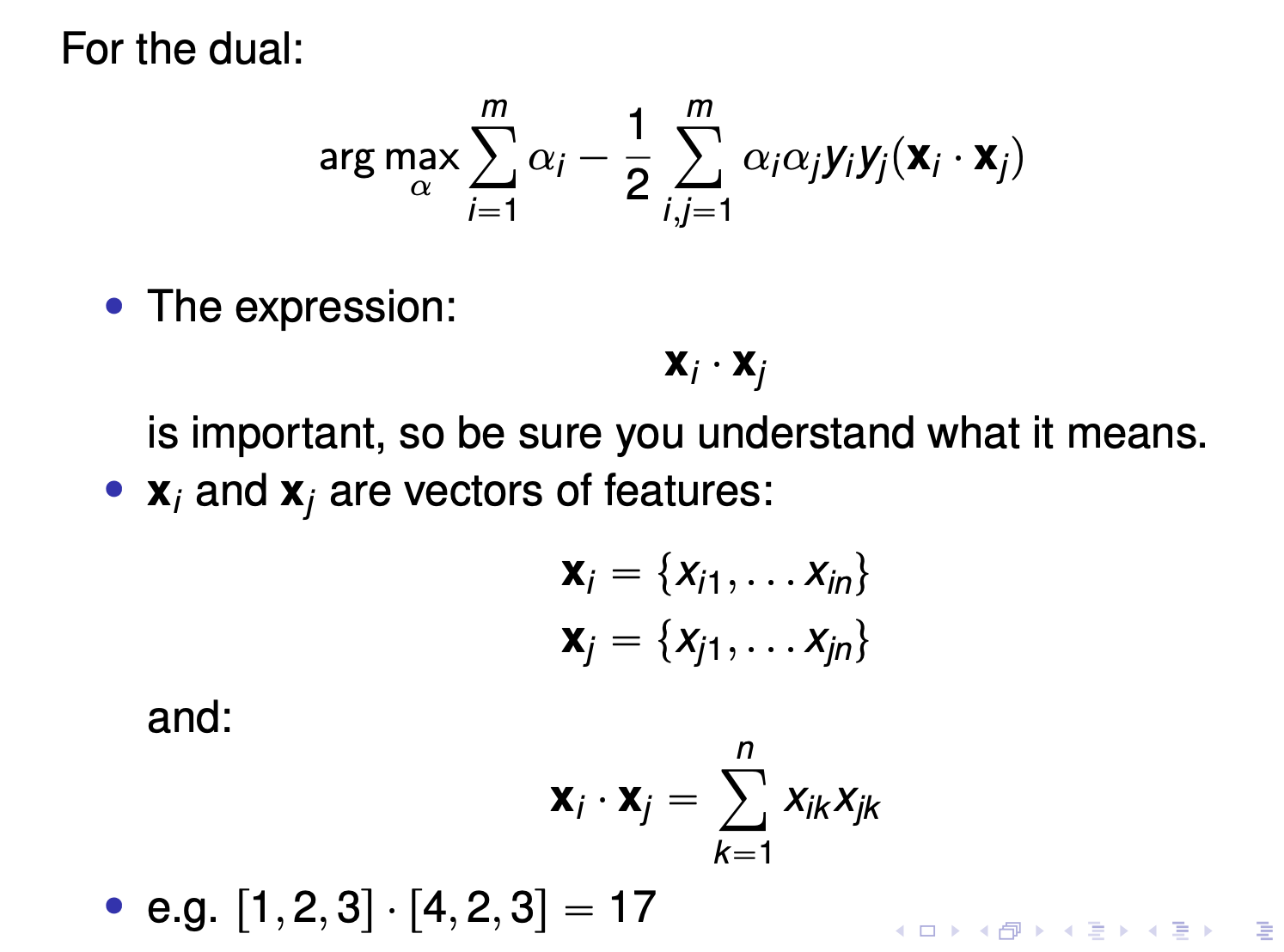
# Week 4

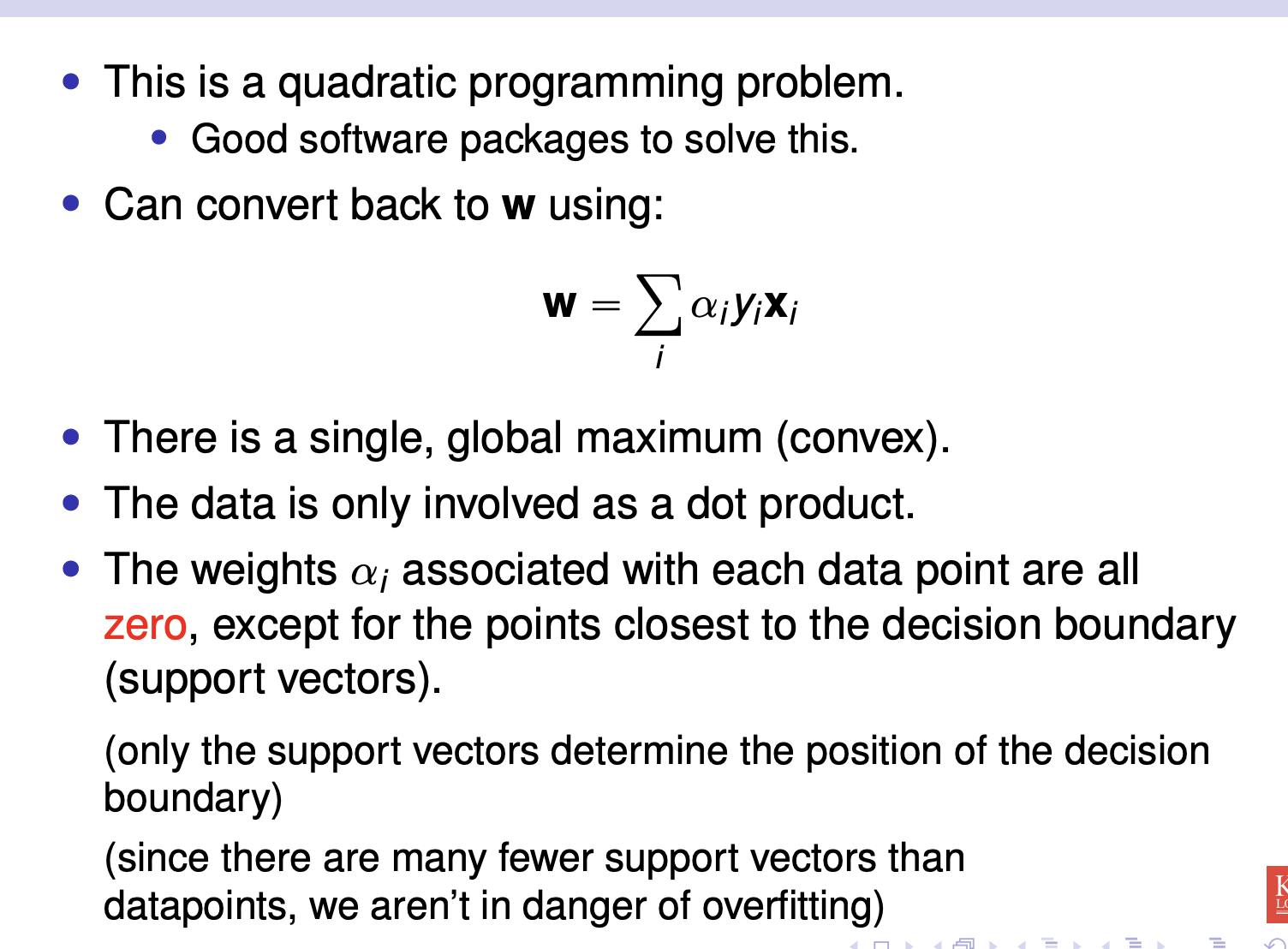
# Week 5

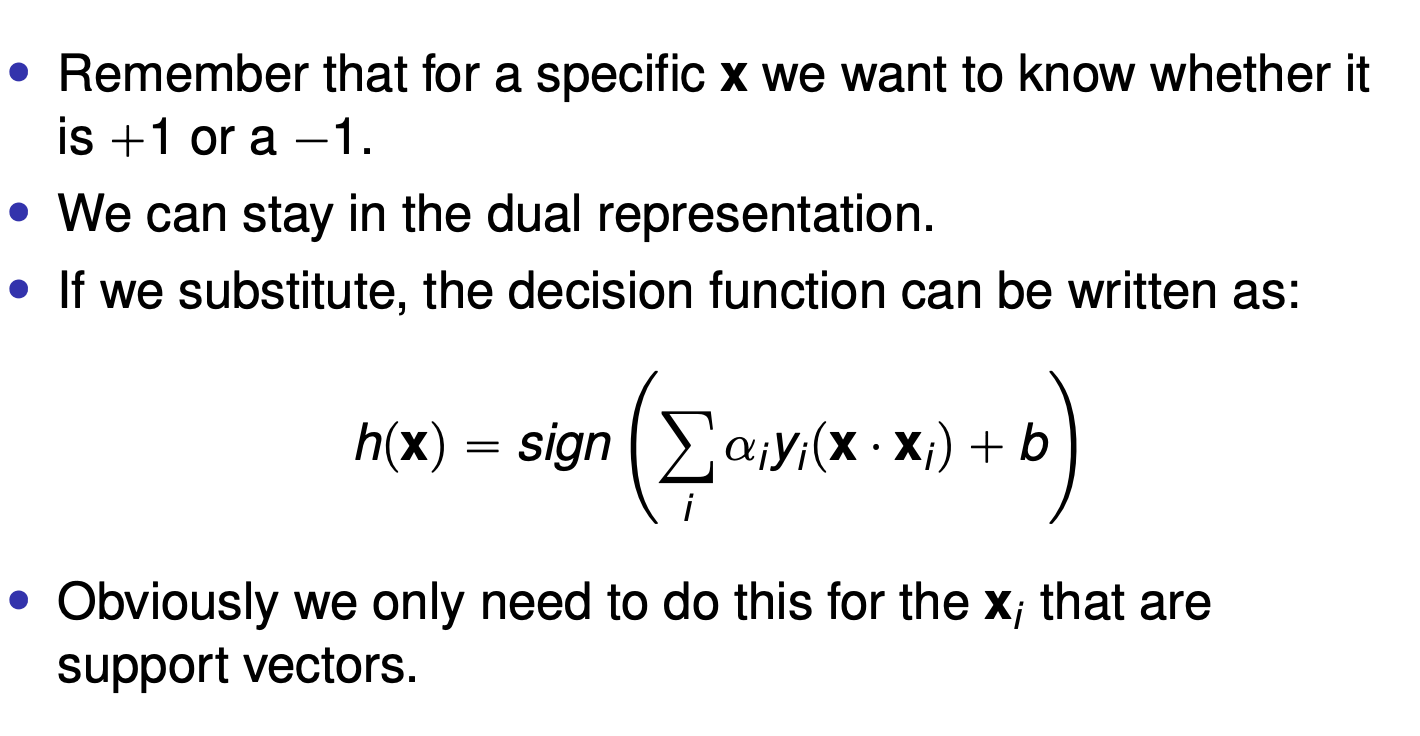
## Support vector machine



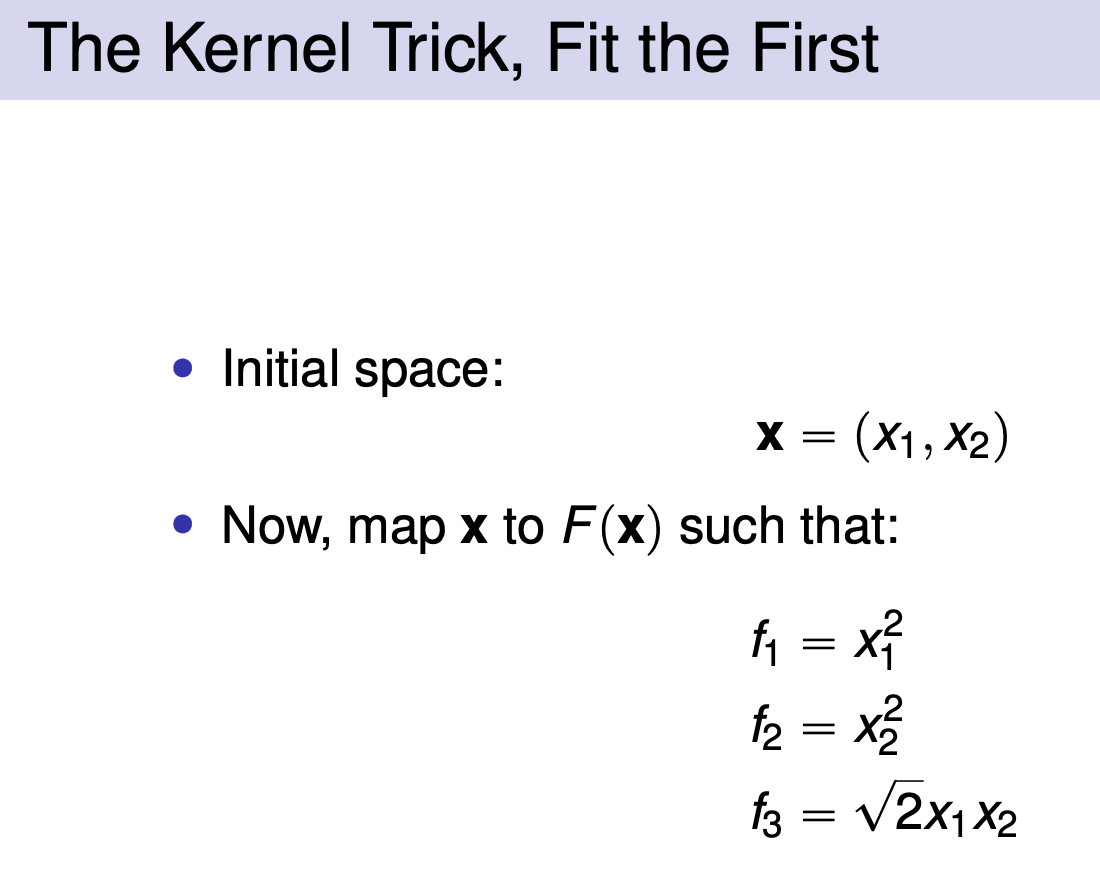


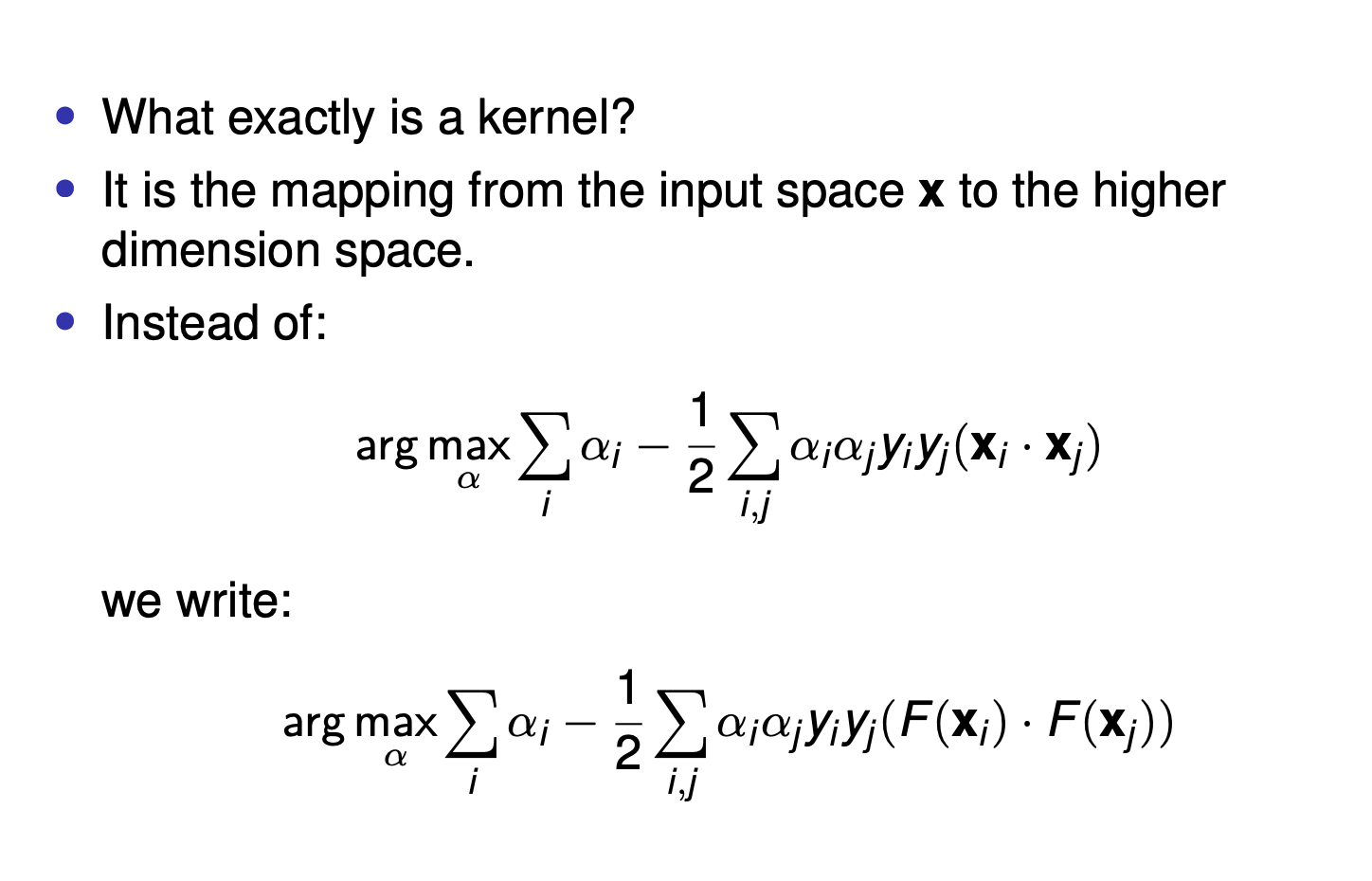


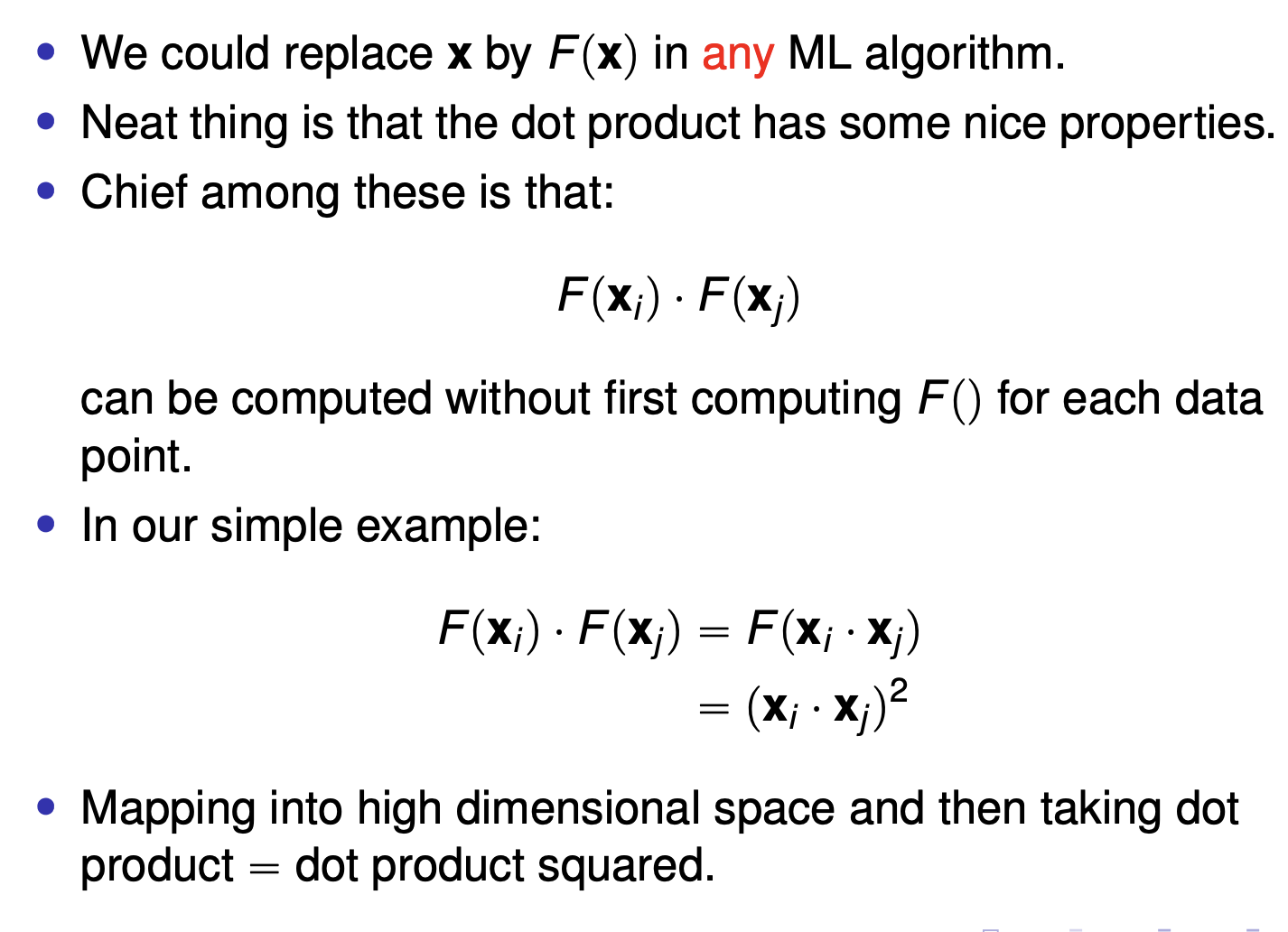


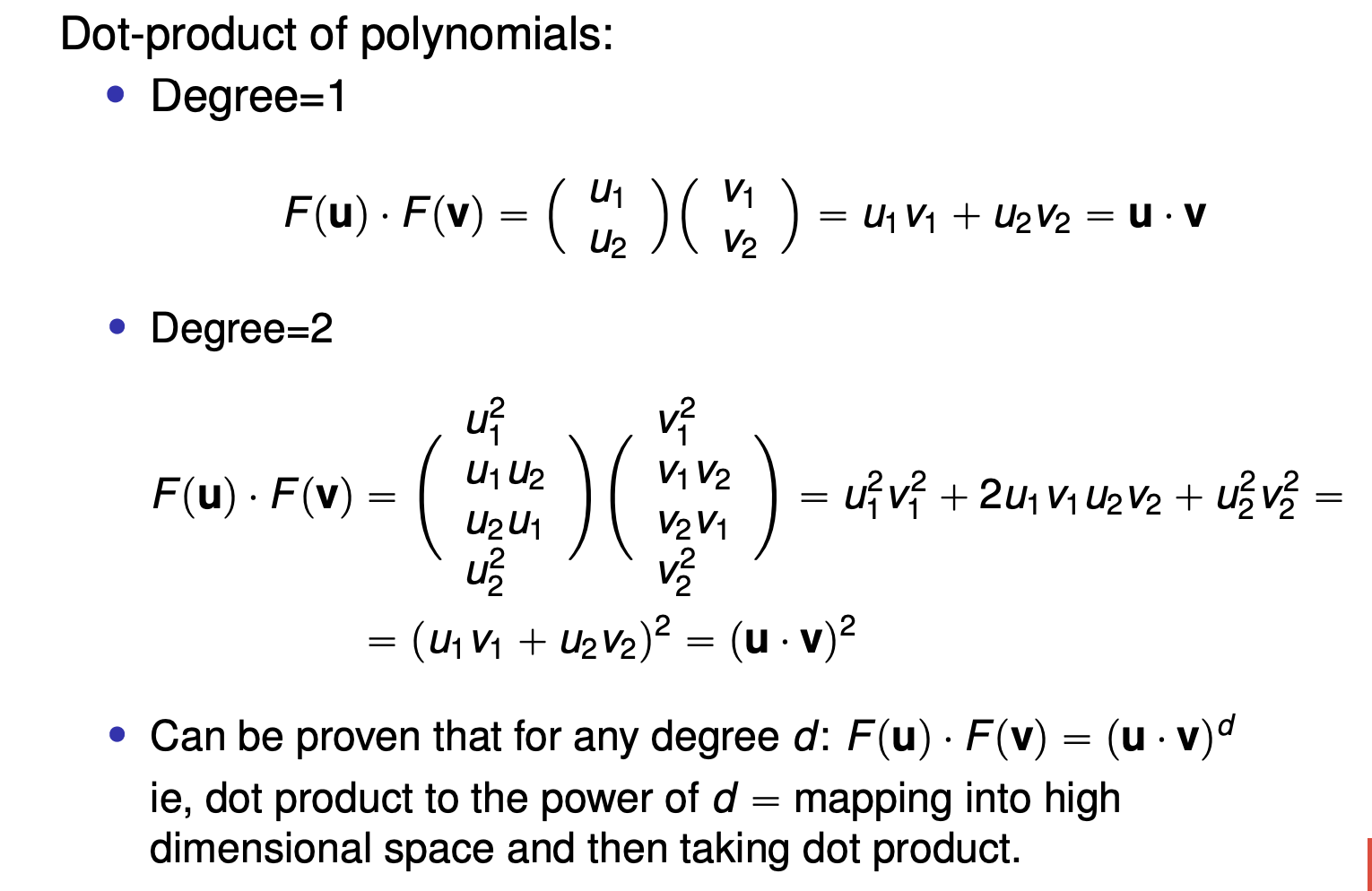


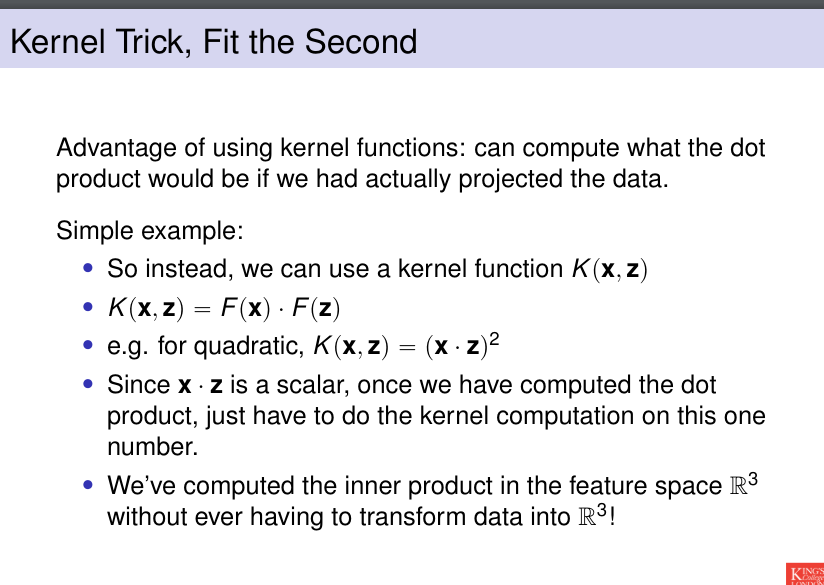
## Kernel Trick

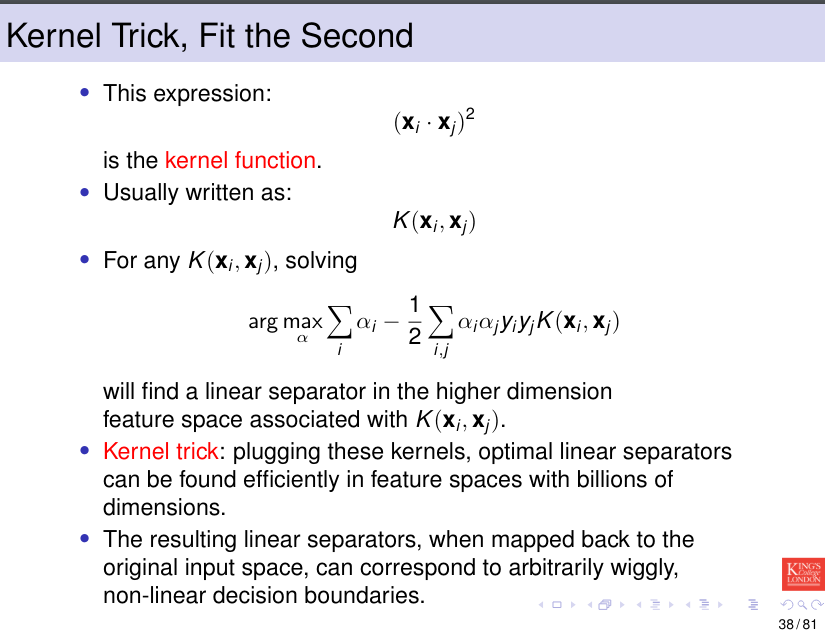


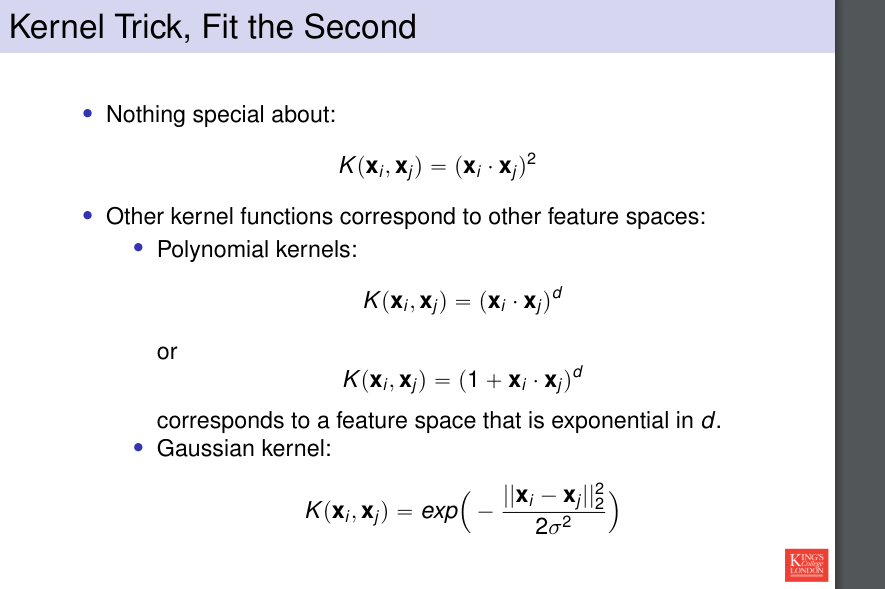


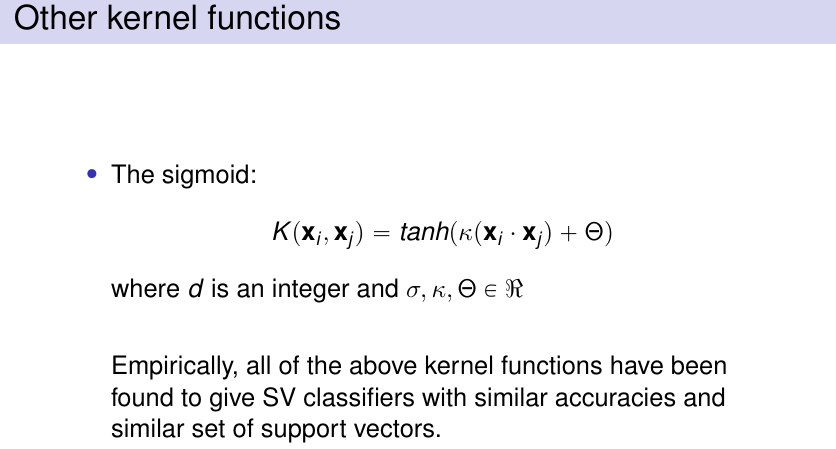


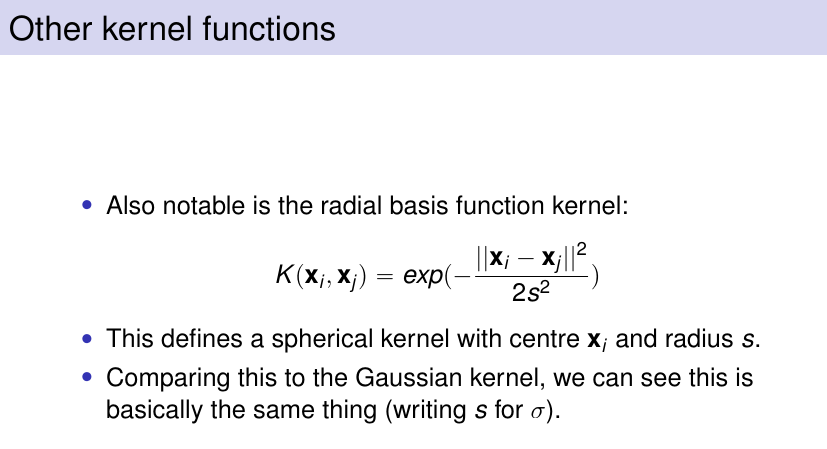




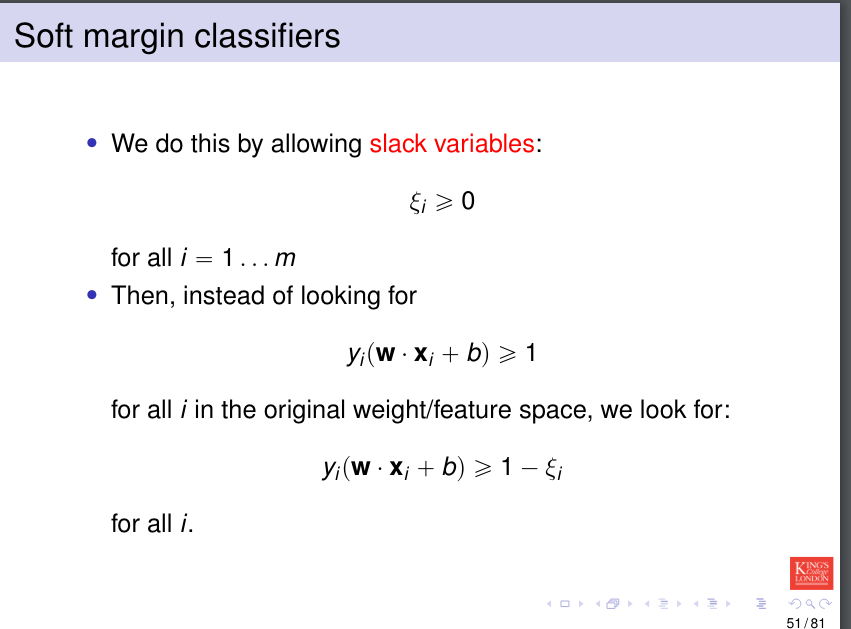


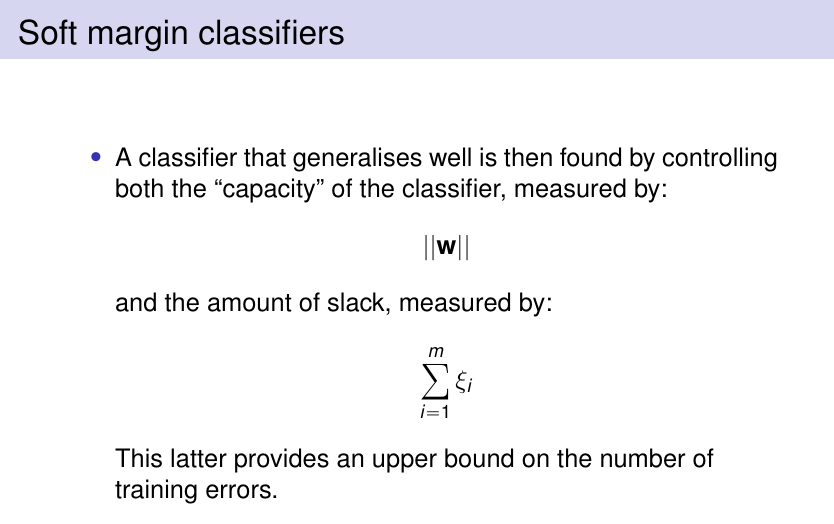


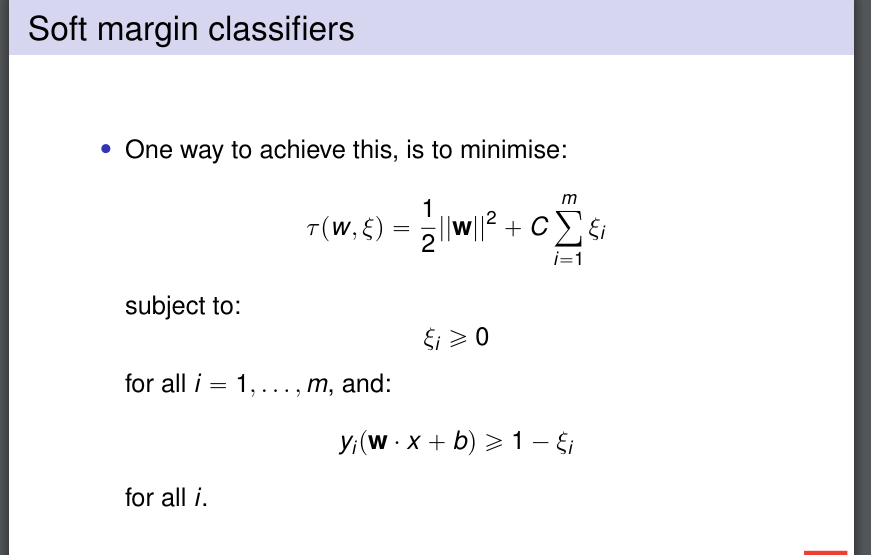


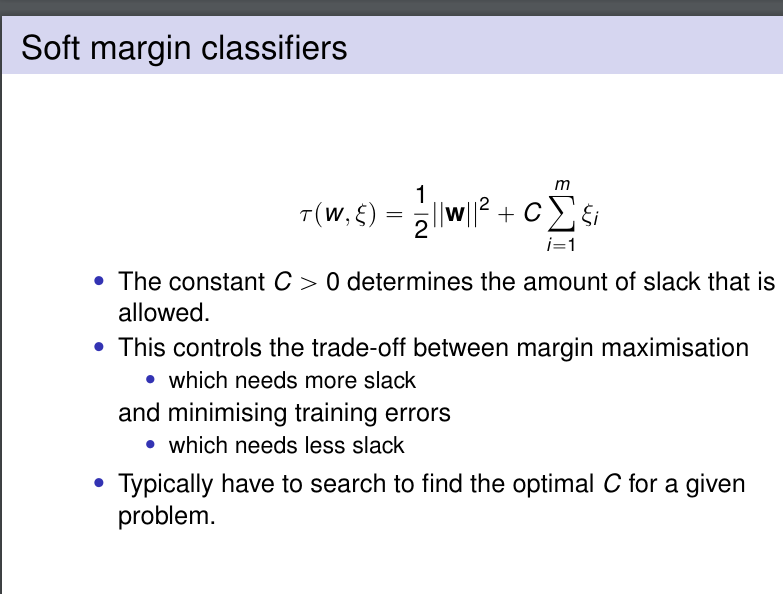


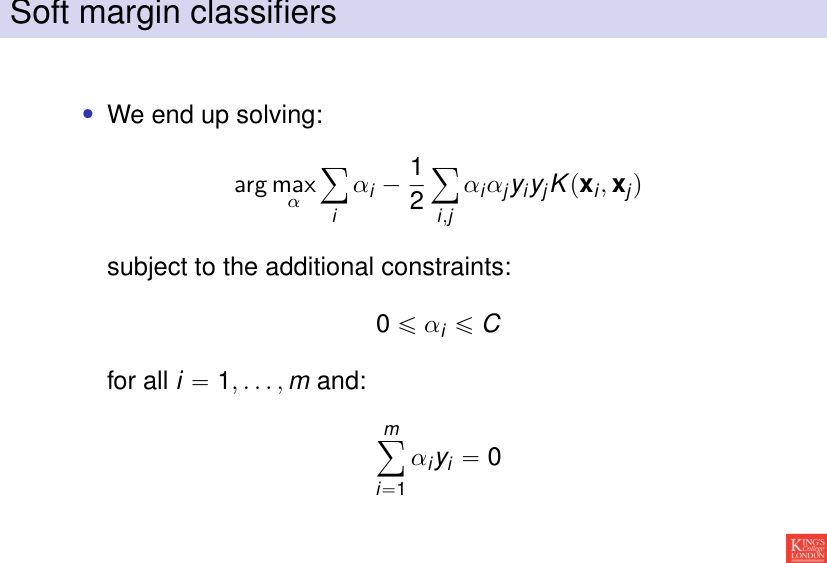
## Soft margin classifier



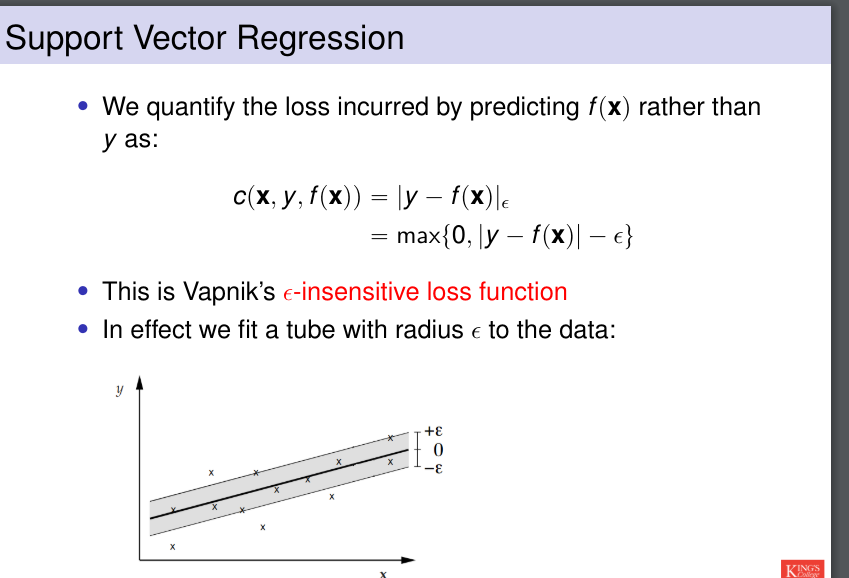


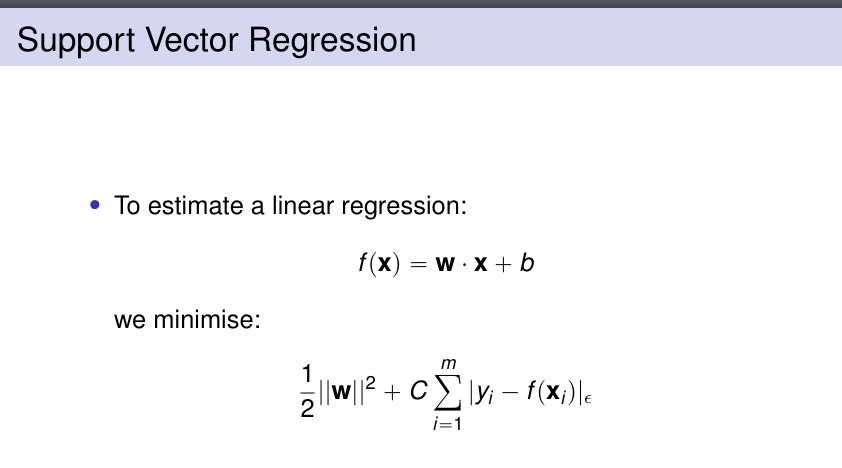


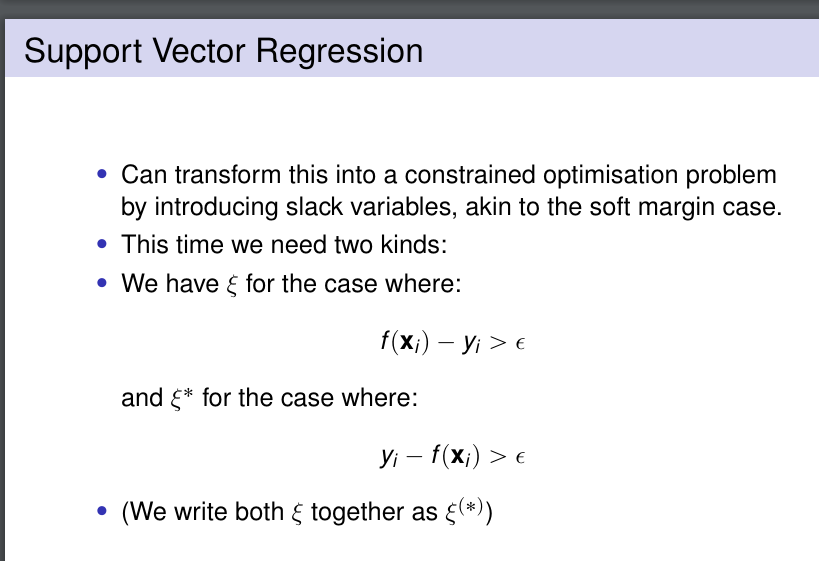


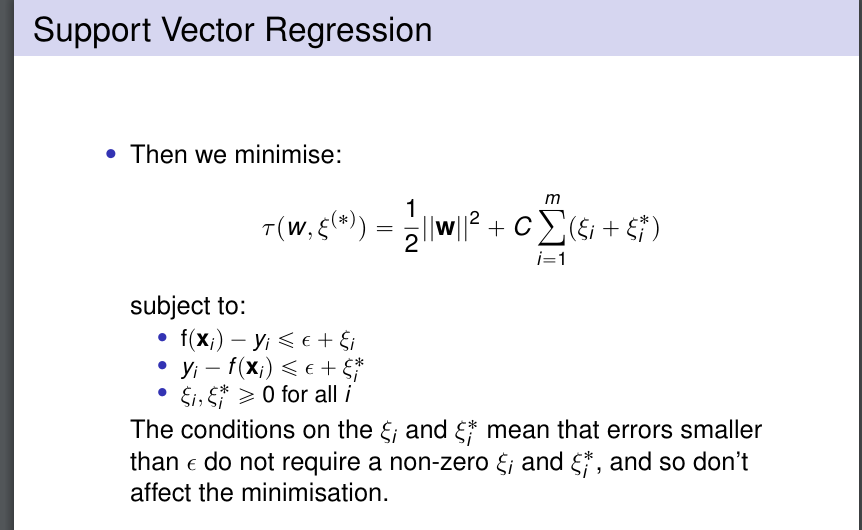


## Support vector regression

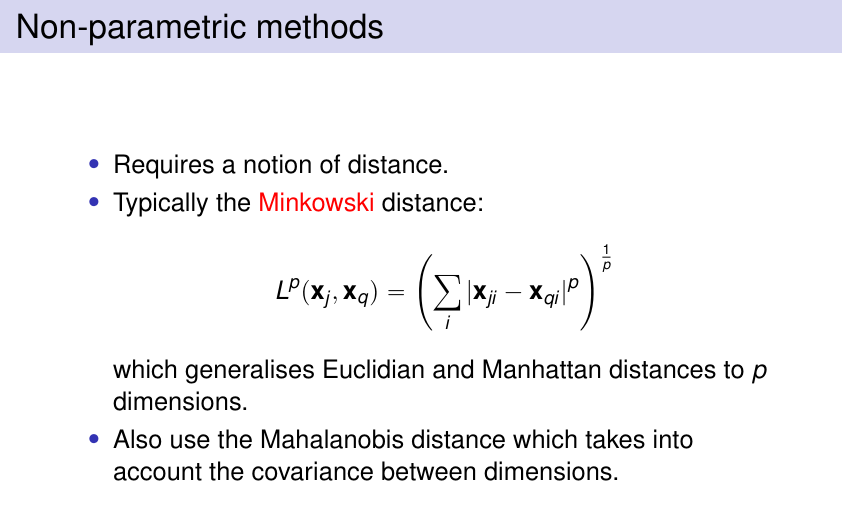




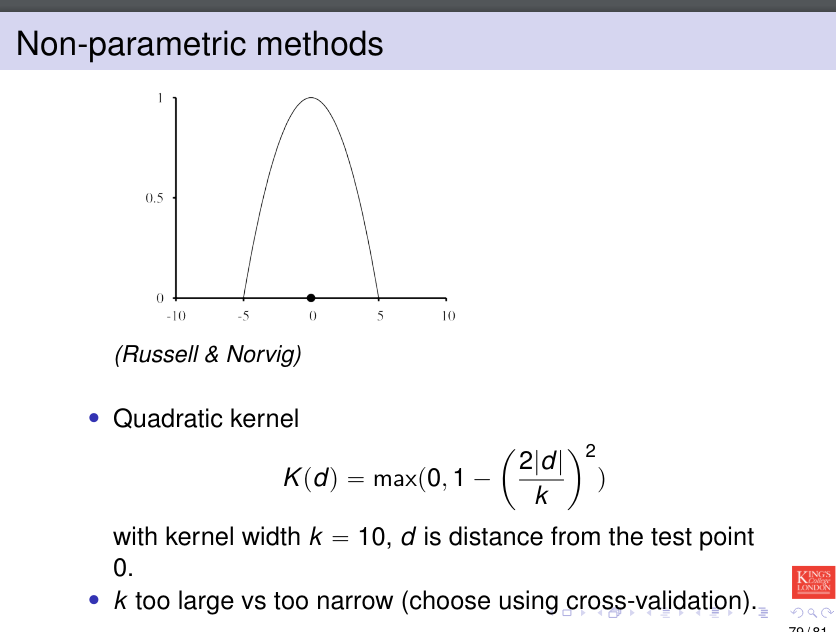




## Non parametric method



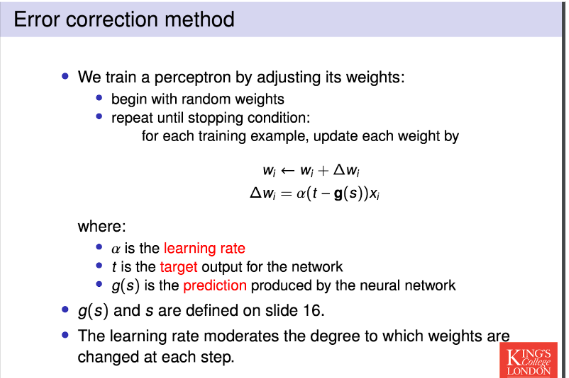
就是范数

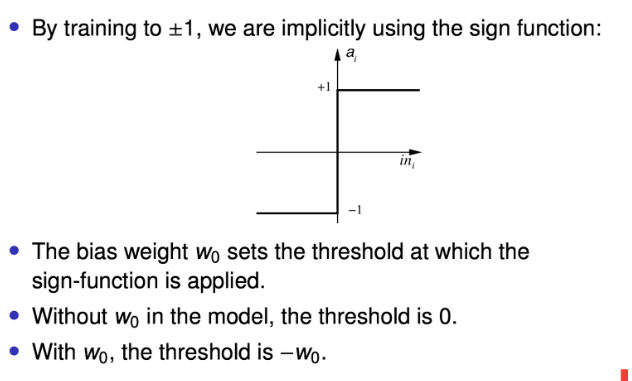




# Week 7

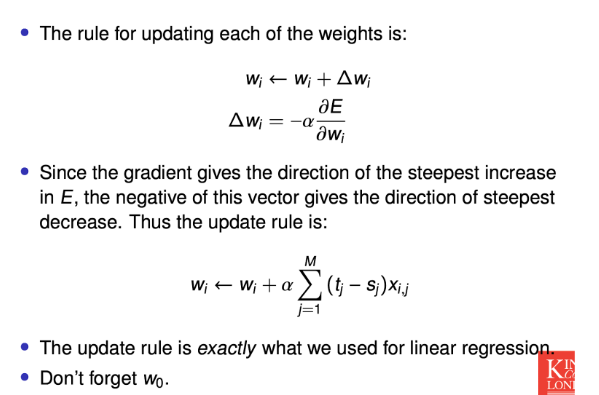
## Error correction method

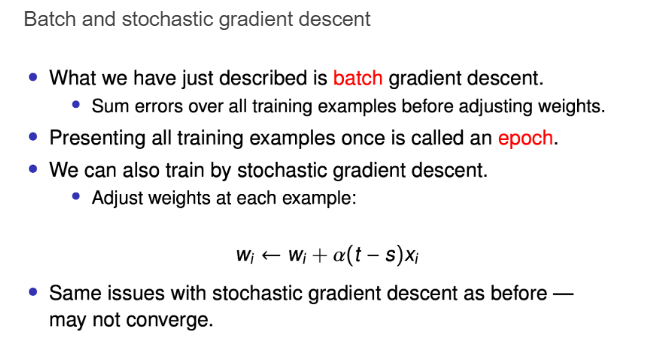






## Delta rule

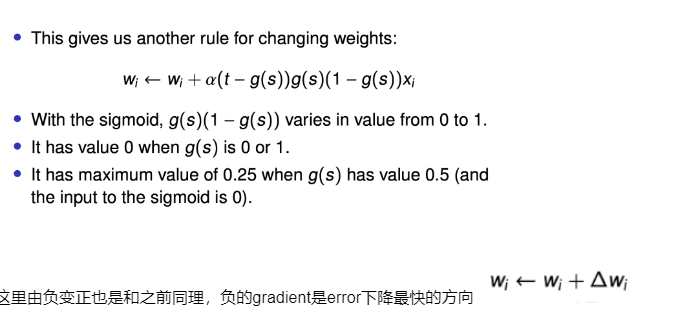




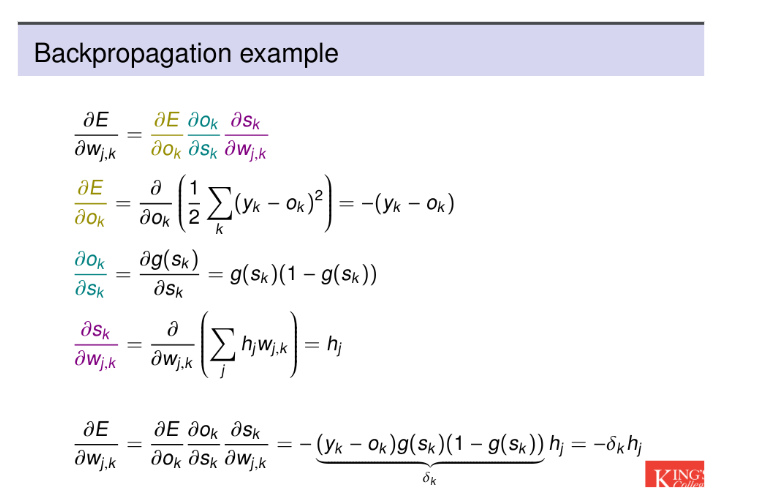


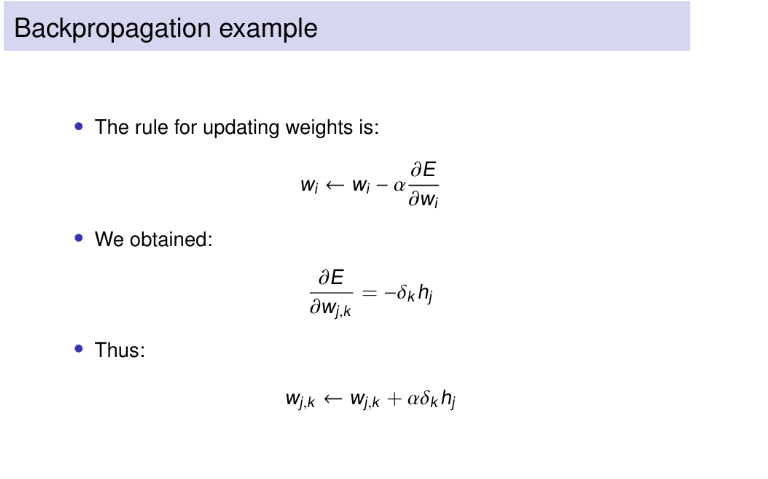
## Sigmoid

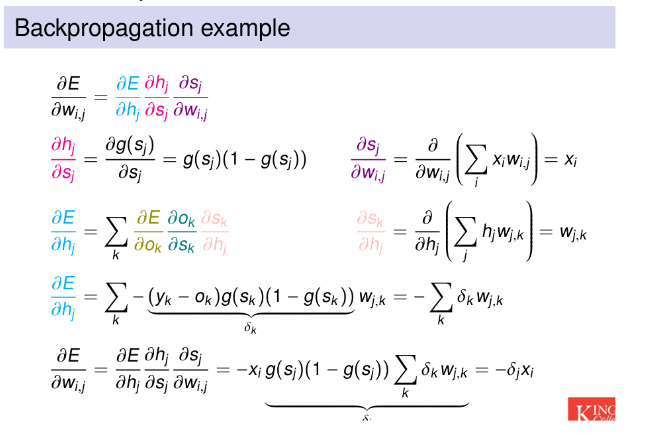
## Generalized delta rule

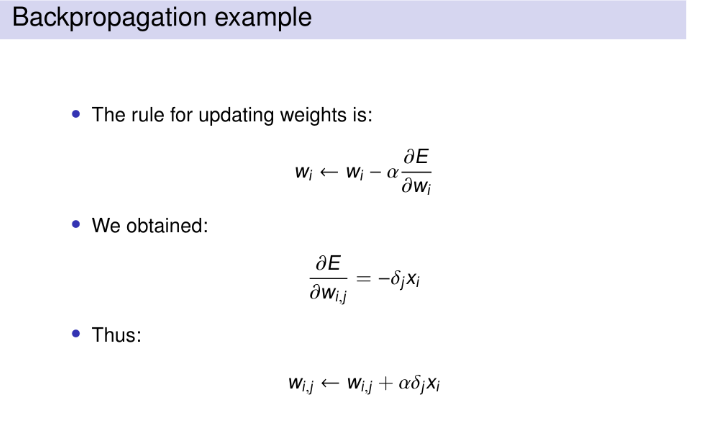


## Backpropagation









# Week 8

## Softmax

