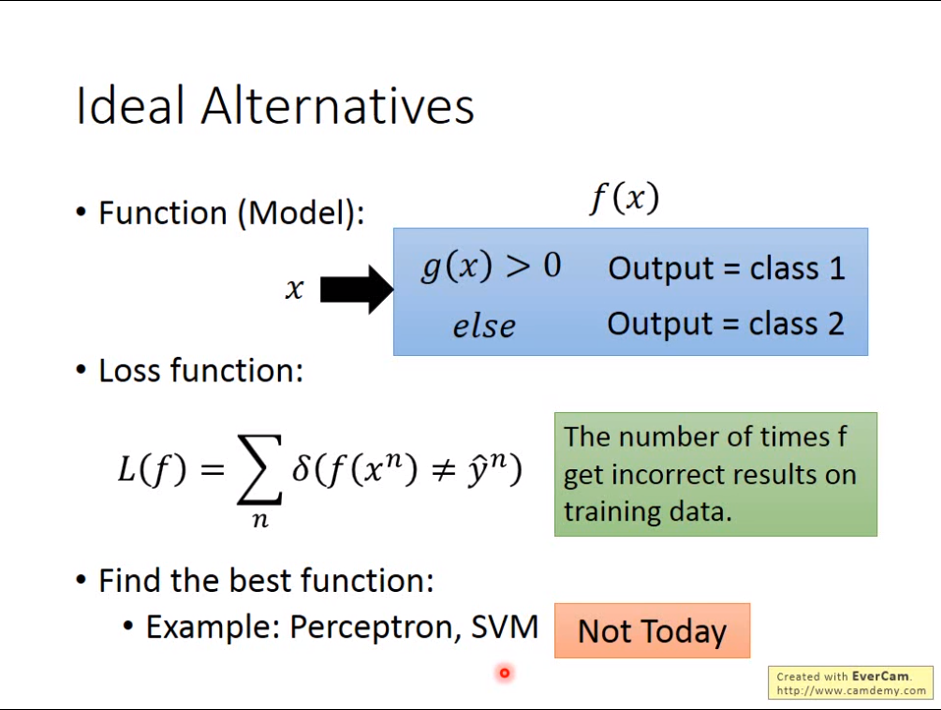
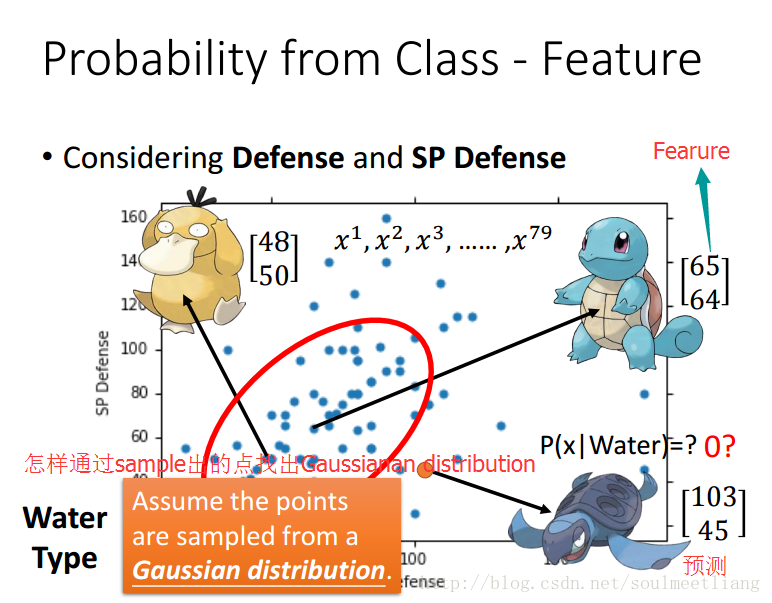
## 分类：概率生成模型(possibility generative model)

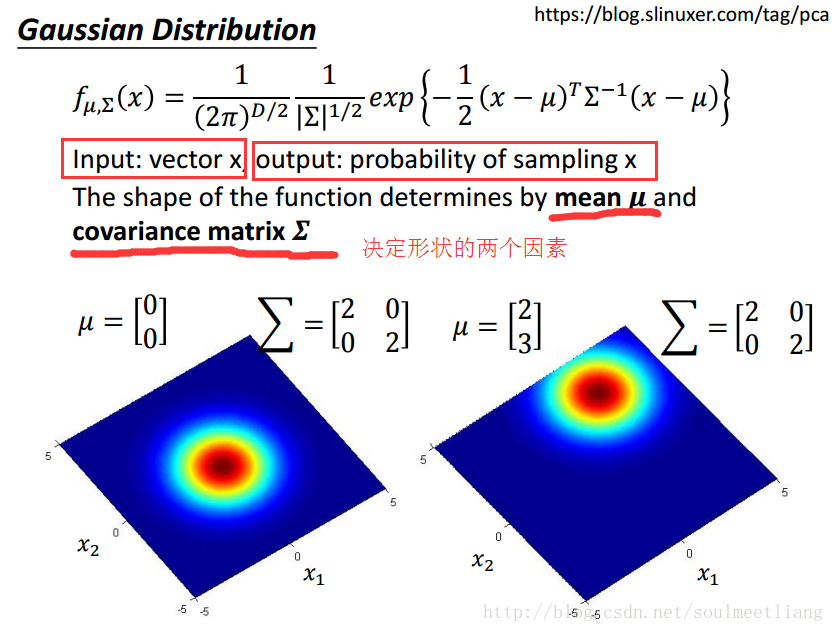
### Ideal Alternatives

Classification不能简单的套用regression，因为regression会受到异常数据的影响，too correct，并且当每个class之间的关系不太明显时，也很难找到合适的function





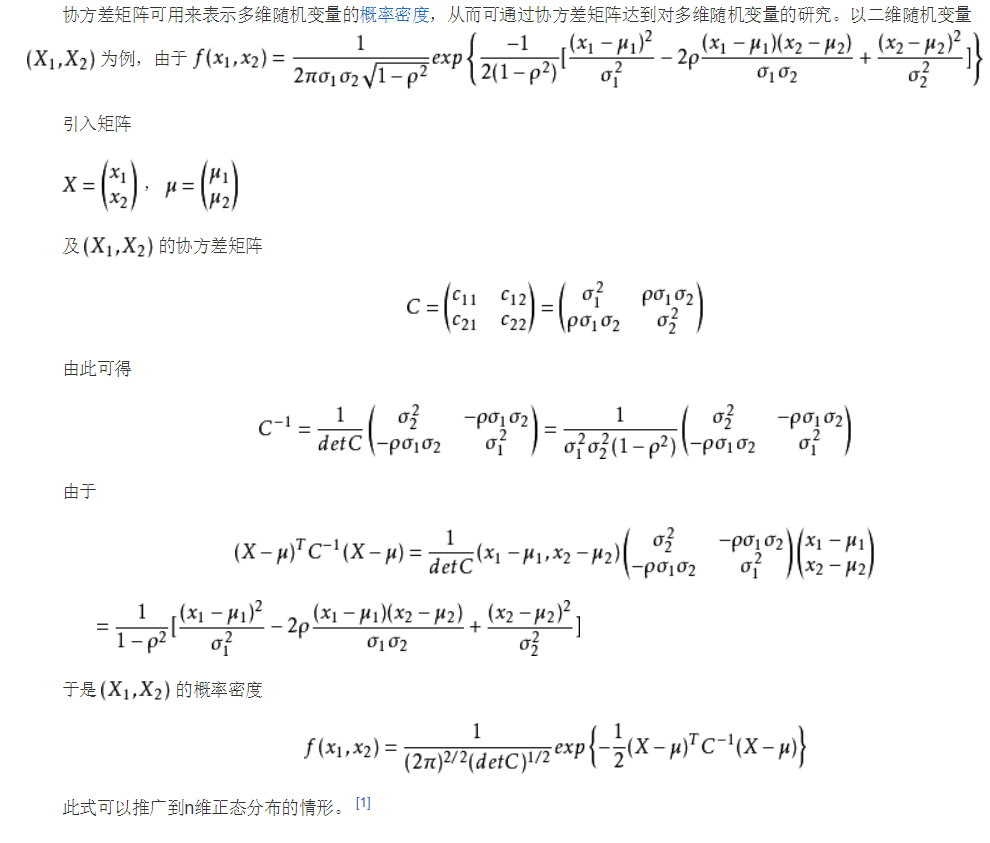
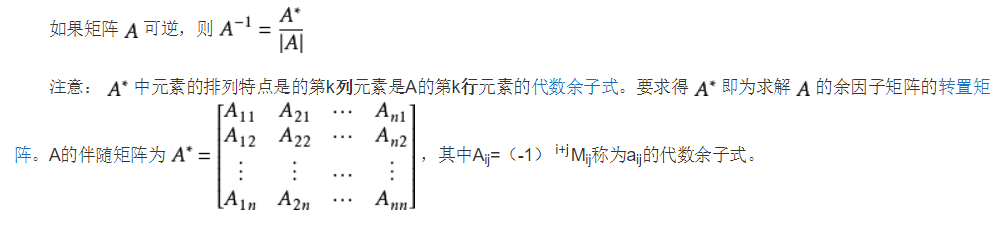
### Gaussian Distribution



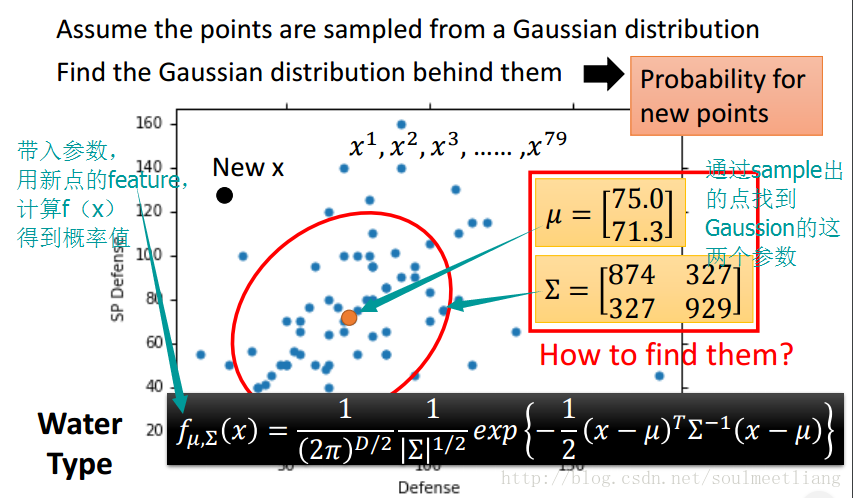
高斯分布（正态分布），这里是通过多维正态分布来推导每一个point应该出现的概率，在training data中训练，得到一个与training data的分布比较接近的μ和∑。

**x、μ和∑都是矢量, ∑是协方差矩阵**

exp，高等数学里以自然常数e为底的指数函数



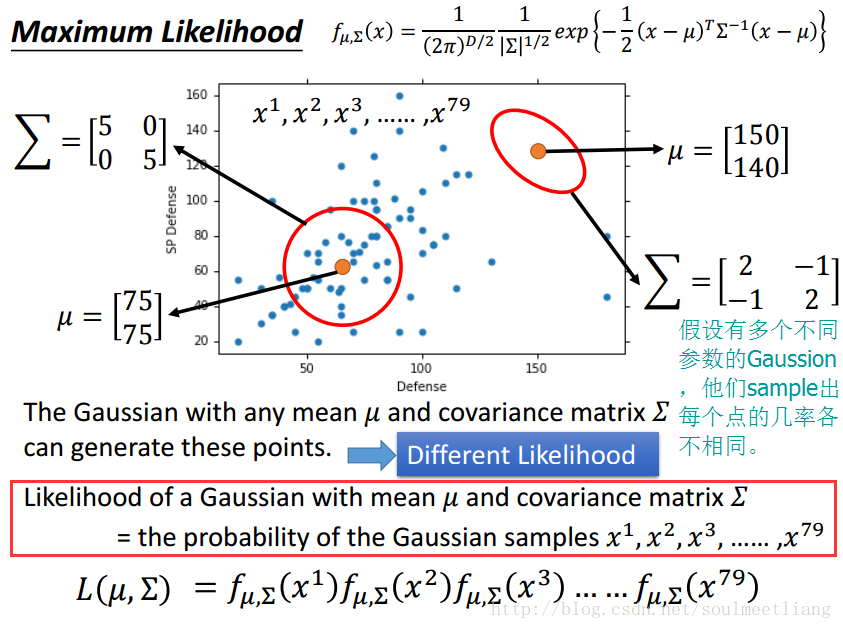
### Probability from each Class

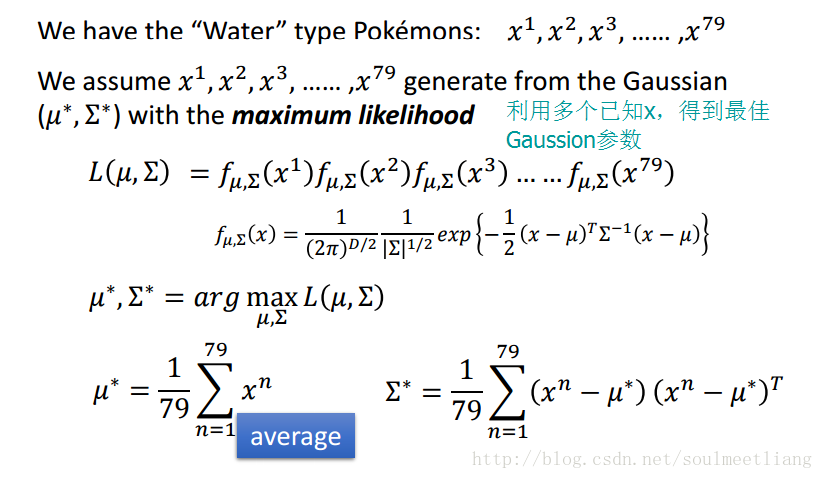


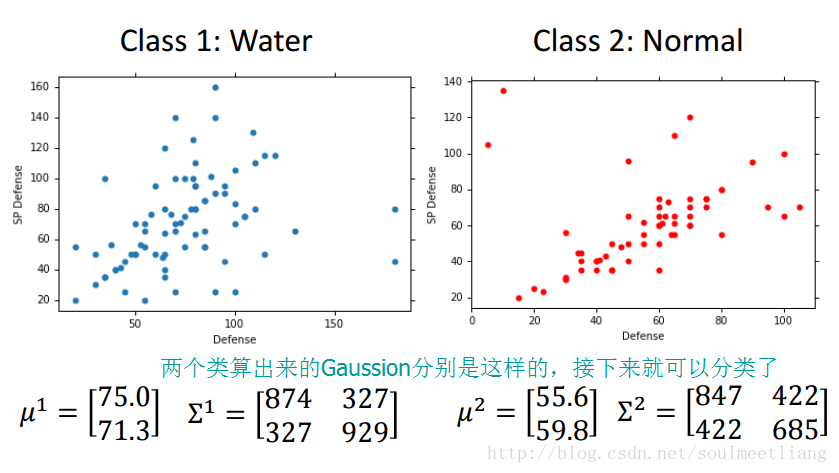
这里的新点指的是其他的training data，不能用training data，否则就是不准确的

Find the Gaussian Distribution

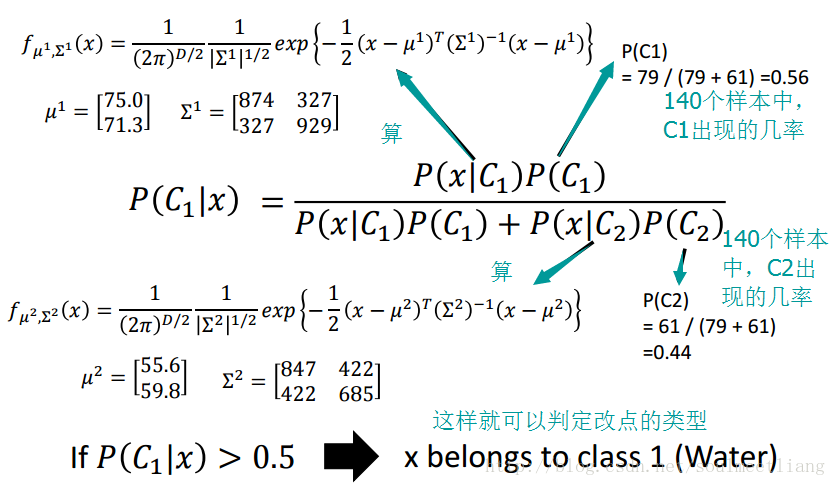
最大似然估计



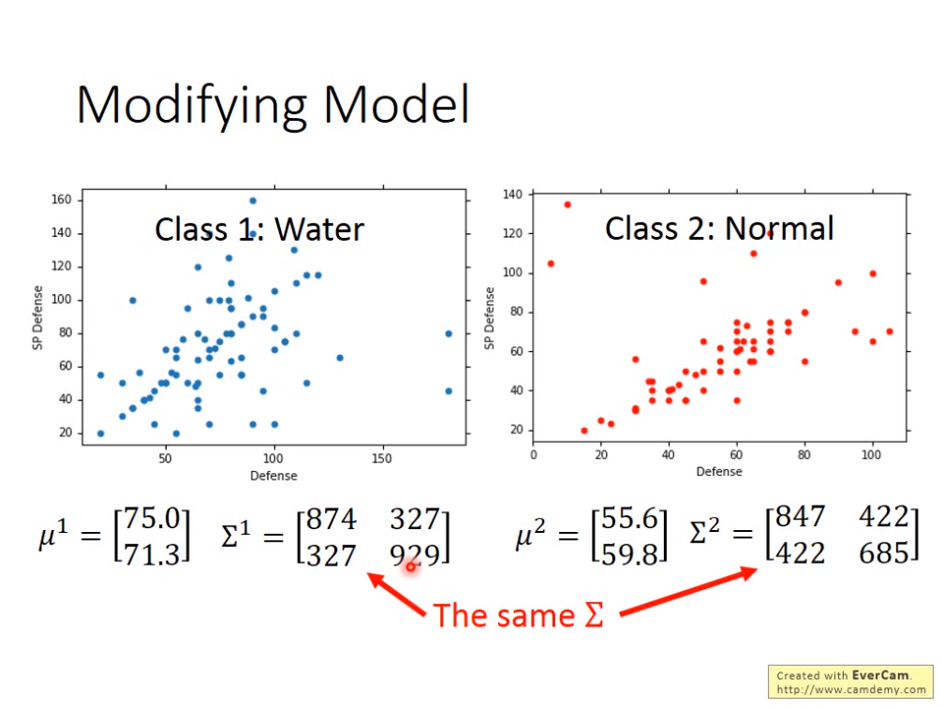




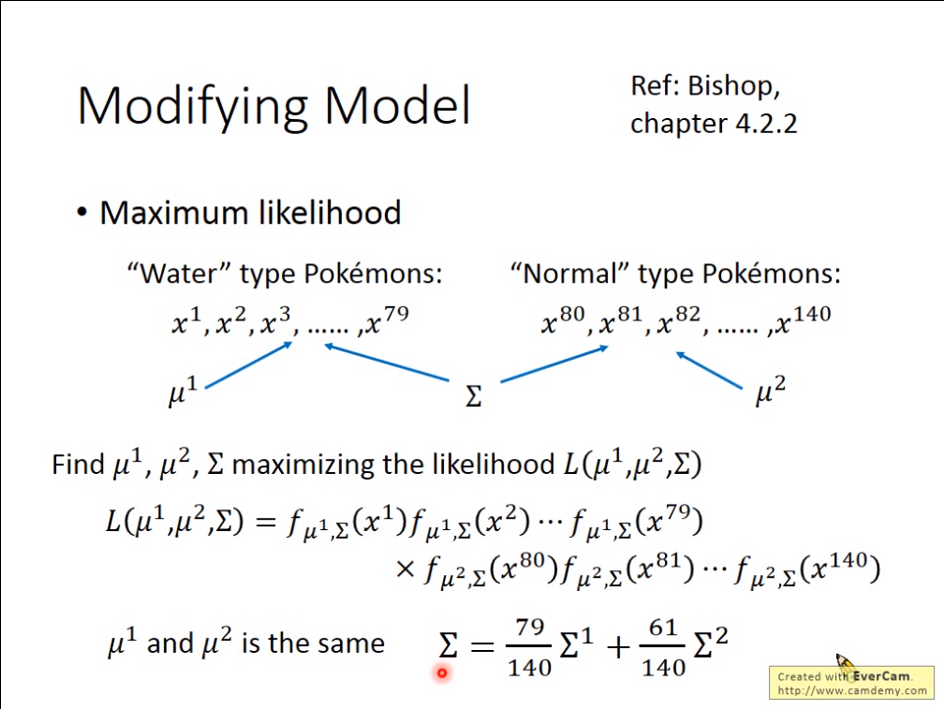
### Do classfication



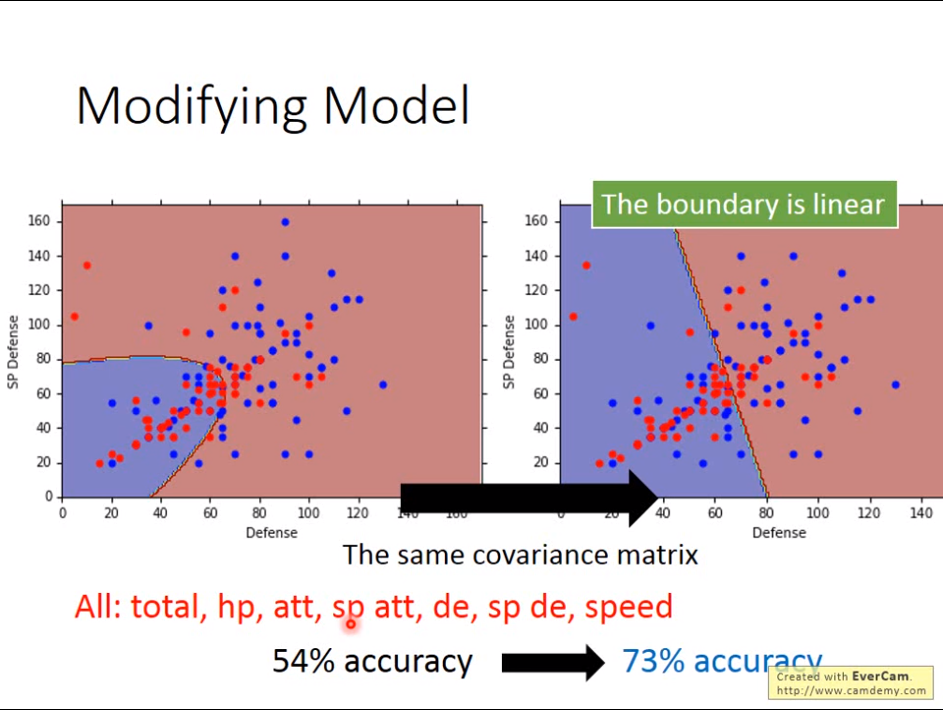
### Modifying Model



因为∑与x平方成正比，当feature size变大时，∑（covariance matrix）的增长速度就会很快，就会出现overfitting的情况，所以现在将两个class∑共用，减少参数。（∑1 + ∑2 -> ∑）

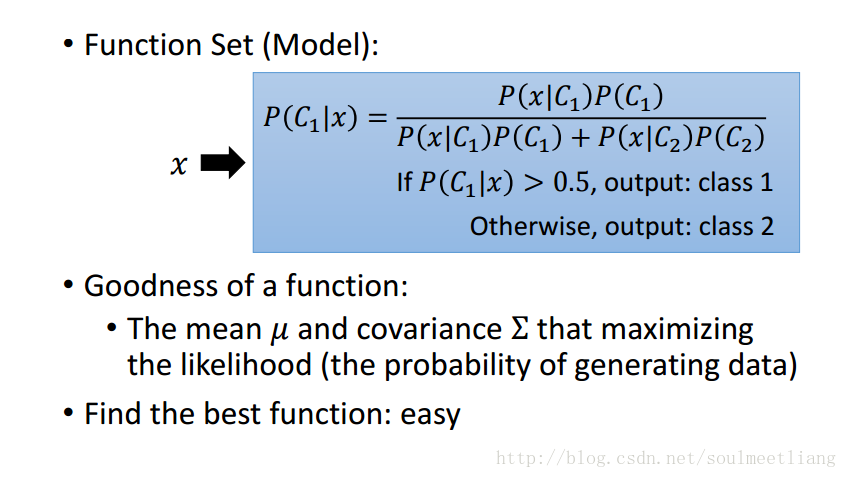


**效果：**

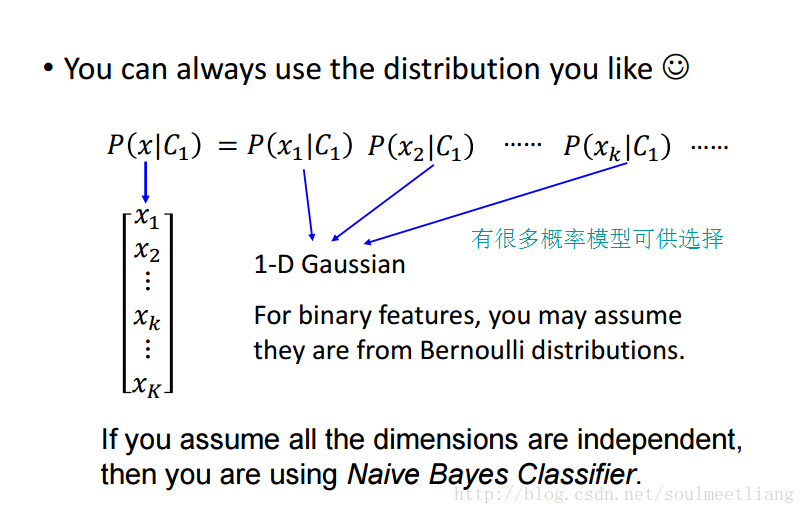


### Summary

#### Three Steps

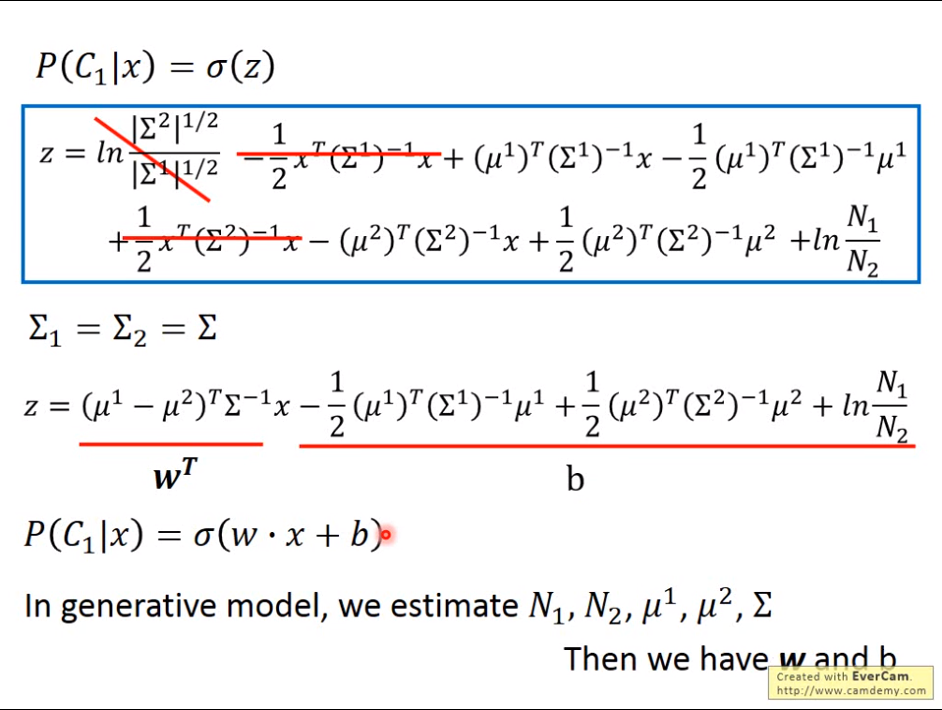


#### Possibility Model（概率模型的选择）



### Warning of Math

具体推导只是数学公式的等价演变，详见视频，此处只说结论



在反复推导后，得到，从而推出无需算出μ和∑，只需一开始计算w和b.

结论：本质上logistic regression也是linear regression问题

Sigmoid函数

https://gss3.bdstatic.com/-Po3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D99/sign=a46bd6f1dd33c895a27e9472d01340df/0df3d7ca7bcb0a4659502a5f6f63f6246b60af62.jpg

https://gss1.bdstatic.com/-vo3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D236/sign=375012cedfca7bcb797bc02c88086b3f/64380cd7912397dde41ab3095182b2b7d0a2875f.jpg

### 总结

1. 朴素贝叶斯

朴素贝叶斯法是基于贝叶斯定理与**特征条件独立假设**的分类方法[1]  。最为广泛的两种分类模型是决策树模型(Decision Tree Model)和朴素贝叶斯模型（Naive Bayesian Model，NBM）。

和决策树模型相比，**朴素贝叶斯分类器(Naive Bayes Classifier,或 NBC)**发源于古典数学理论，有着坚实的数学基础，以及稳定的分类效率。同时，NBC模型**所需估计的参数很少，对缺失数据不太敏感**，算法也比较简单。理论上，NBC模型与其他分类方法相比具有最小的误差率。但是实际上并非总是如此，这是因为NBC模型假设属性之间相互独立，这个假设在实际应用中往往是不成立的，这给NBC模型的正确分类带来了一定影响。

**朴素贝叶斯分类器基于一个简单的假定：给定目标值时属性之间相互条件独立。**

1. 判别模型（**Discriminative Model**）与生成模型（**Generative Model**）

**判别模型（Discriminative Model）**，又可以称为条件模型，或条件概率模型。估计的是条件概率分布(conditional distribution)，p(class|context)。利用正负例和分类标签，主要关心判别模型的边缘分布。其目标函数直接对应于分类准确率。 （**判别模型多数放在分类**）

      主要特点：寻找不同类别之间的最优分类面，反映的是异类数据之间的差异。

      优点：（1）分类边界更灵活，比使用纯概率方法或生产模型得到的更高级；（2）能清晰的分辨出多类或某一类与其他类之间的差异特征；（3）在聚类、视角变化、部分遮挡、尺度改变等方面效果较好；（4）适用于较多类别的识别；（5）判别模型的性能比生成模型要简单，比较容易学习。

      缺点：（1）不能反映训练数据本身的特性，即能力有限，可以告诉你的是1还是2，但没有办法把整个场景描述出来；（2）缺少生成模型的优点，即先验结构的不确定性；（3）黑盒操作，即变量间的关系不清楚，不可视。

      常见的主要有：logistic regression、SVMs、traditional neural networks、Nearest neighbor、Conditional random fields。

      主要应用：Image and document classification、Biosequence analysis、Time series prediction。

**生成模型（Generative Model）**，又叫产生式模型。

估计的是联合概率分布（joint probability distribution），p(class,context)=p(class|context)\*p(context)。用于随机生成的观察值建模，特别是在给定某些隐藏参数情况下。在机器学习中，或用于直接对数据建模（用概率密度函数对观察到的样本数据建模），或作为生成条件概率密度函数的中间步骤。通过使用贝叶斯规则可以从生成模型中得到条件分布。如果观察到的数据是完全由生成模型所生成的，那么就可以拟合生成模型的参数，从而仅可能的增加数据相似度。但数据很少能由生成模型完全得到，所以比较准确的方式是直接对条件密度函数建模，即使用分类或回归分析。与描述模型的不同是，描述模型中所有变量都是直接测量得到。

     主要特点：（1）一般主要是对后验概率建模，从统计的角度表示数据的分布情况，能够反映同类数据本身的相似度；（2）只关注自己的类本身（即点左下角区域内的概率），不关心到底决策边界在哪。

      优点：（1）实际上带的信息要比判别模型丰富；（2）研究单类问题比判别模型灵活性强；（3）模型可以通过增量学习得到；（4）能用于数据不完整（missing data）情况；（5）很容易将先验知识考虑进去。

      缺点：（1）容易会产生错误分类；（2）学习和计算过程比较复杂。

      常见的主要有：Gaussians、Naive Bayes、Mixtures of multinomials、Mixtures of Gaussians、Mixtures of experts、HMMs、Sigmoidal belief networks、Bayesian networks、Markov random fields。

      主要应用：（1）传统基于规则的或布尔逻辑系统正被统计方法所代替；（2）医学诊断。

      注：所列举的生成模型也可以用判决模型的方法来训练，比如GMM或HMM，训练的方法有EBW(Extended Baum Welch)，或最近Fei Sha提出的Large Margin方法。

 过去的报告认为判别模型在**分类问题**上比生成表现更加好（比如Logistic Regression与Naive Bayesian的比较，再比如HMM与Linear Chain CRF的比较）。

当然，生成模型的图模型也有一些难以代替的地方，比如更容易结合无标注数据做semi-or-un-supervised learning。

**所以生成模型和判别模型的主要区别在于：添加了先验概率**

**即：生成模型：p(class, context)=p(class|context)\*p(context)**

**判别模型： p(class|context)**