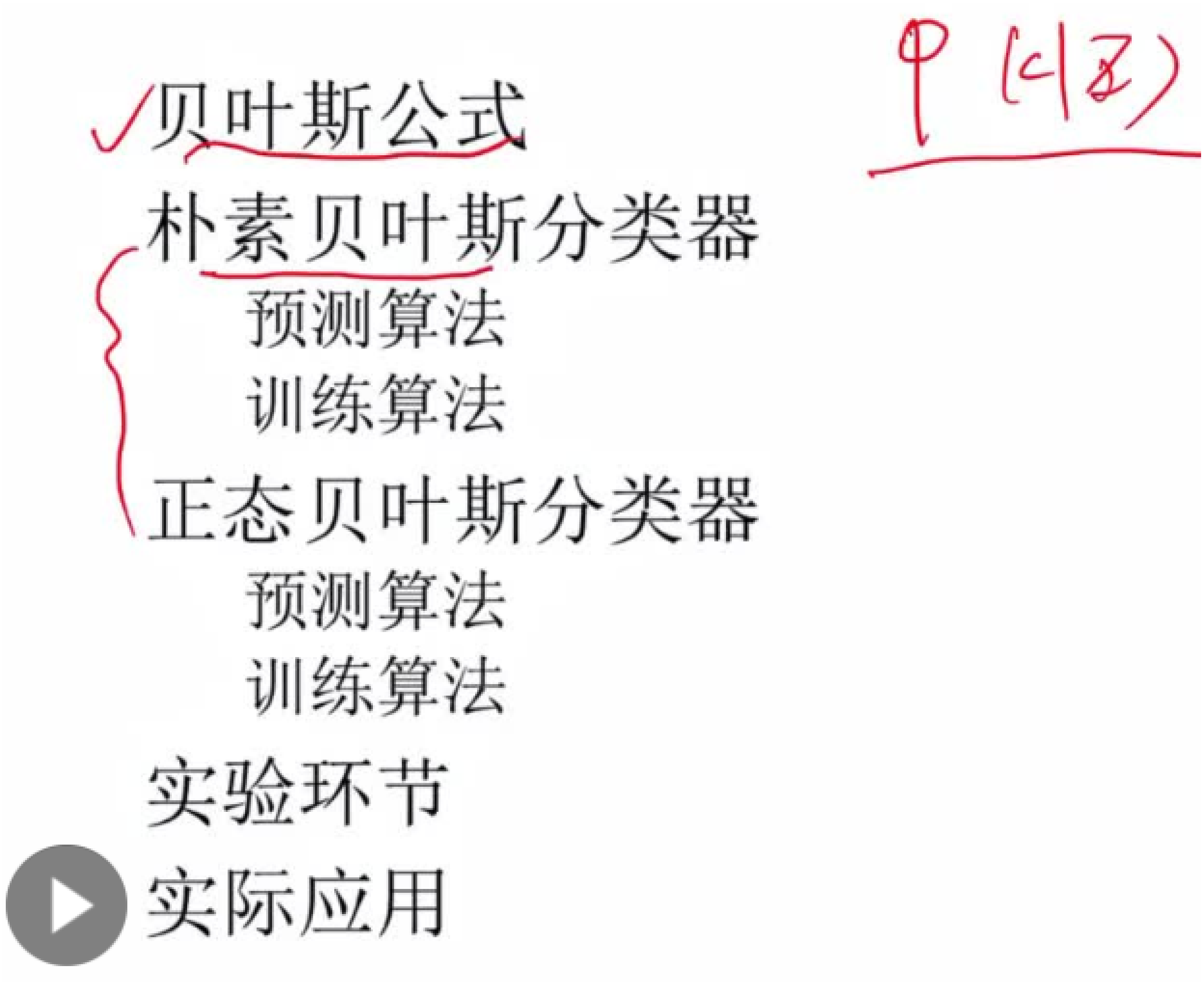
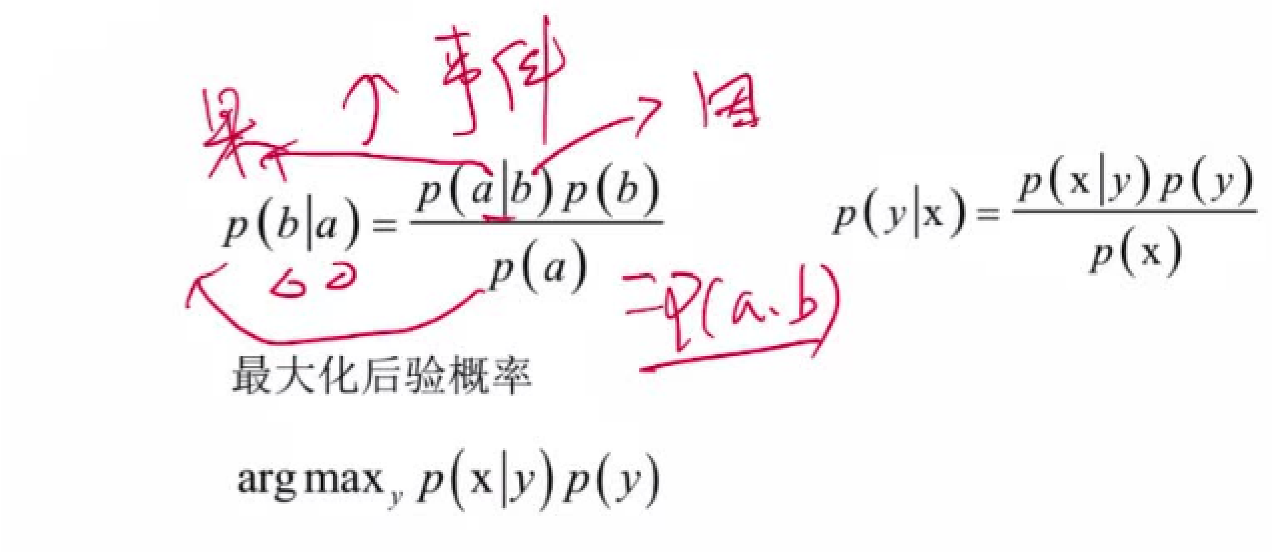
5 贝叶斯分类器知识要点

5.1 本集简介



5.2 贝叶斯公式

随机事件是在随机试验中，可能出现也可能不出现，而在大量重复试验中具有某种规律性的事件叫做随机事件(简称事件)。随机事件通常用大写英文字母A、B、C等表示。



随机事件推广到随机变量。

贝叶斯分类器。分类器。

P(z|y)

P(y|z)

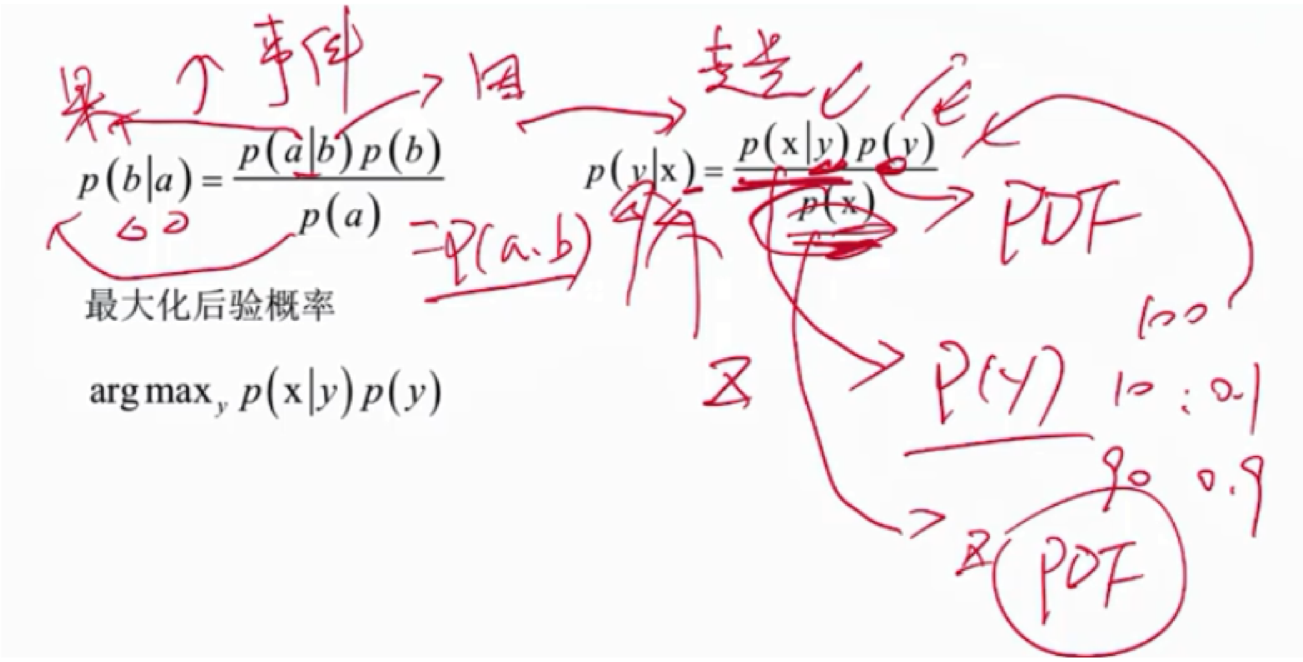
$$\frac{p(b|a) = p(a|b)p(b)}{p(a)}$$

推广到随机变量

$$\frac{p(y|x) = p(x|y)p(y)}{p(x)}$$

最大化后验概率

$$arg\ max\ y\ p(x|y)p(y)$$



重点理解最大化后验概率

概率密度函数，PDF

什么是先验概率？

条件概率，全概率公式，这些都是要记住的。

例如：男性的脚比女性的脚大。

连续型特征　$X: X_1\ X_2 \cdots X_n$　X_i

$$f(x_i = x | y = c) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

$$\arg\max_c\ p(y=c) \prod_{i=1}^{n} f(x_i | y=c)$$

知识补充

条件概率
贝叶斯公式

$$p(b|a) = \frac{p(a,b)}{p(a)}$$

$$p(a|b) = \frac{p(a,b)}{p(b)}$$

$$p(a|b) = \frac{p(a)p(b|a)}{p(b)}$$

5.3 朴素贝叶斯分类器

重点记住这两个公式

朴素贝叶斯分类器

$$p(y=c_i|\mathrm{x})=\frac{p(y=c_i)p(\mathrm{x}|y=c_i)}{p(\mathrm{x})}$$

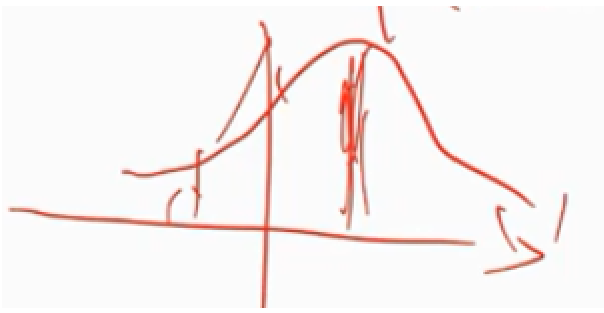
$$p(y=c_i|\mathrm{x})=\frac{p(y=c_i)\prod_{j=1}^{n}p(x_j|y=c_i)}{Z}$$

$$\arg\max_{c} P(x|c)P(c)$$

分为两个部分，一个是离散型随机变量？

怎么根据样本来估计它的PDF？

现实世界中的事件，一般假设服从正态。



对最大后验概率的理解

盲点

独立同分布的**中心极限定理**。

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty}\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \emptyset(x)$$

离散型的随机变量

离散型特征

$$p\left(x_j = v \middle| y = c\right) = \frac{\sum N_{x_{ij}=v \wedge y_i = c}}{N_{y=c}}$$

$$\arg\max_{y}\, p(y=c)\prod_{j=1}^{n} p\left(x_j = v \middle| y = c\right)$$

$$p\left(x_j = v \middle| y = c\right) = \frac{\sum N_{x_{ij}=v \wedge y_i = c} + 1}{N_{y=c} + k}$$

$x_1, x_2 \ldots x$

离散型特征

类条件概率

$$p\left(x_i = v \mid y = c\right) = \frac{N_{x_i = v, y = c}}{N_{y=c}}$$

类概率

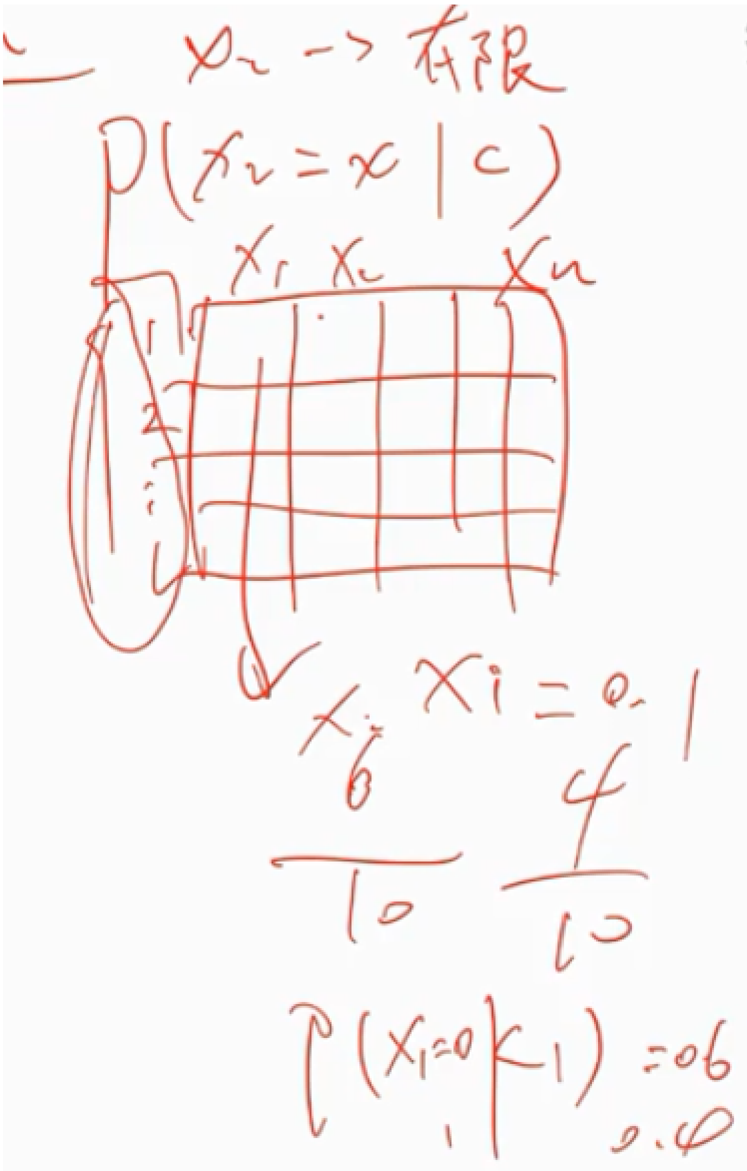
$$p\left(y = c\right) = \frac{N_{y=c}}{N}$$

预测函数

$$\arg\max_y p\left(y = c\right) \prod_{i=1}^{n} p\left(x_i = v \mid y = c\right)$$

拉普拉斯平滑

$$p\left(x_i = v \mid y = c\right) = \frac{N_{x_i = v \wedge y = c} + 1}{N_{y=c} + k}$$



分母，分子

$$p\left(x_j = v \middle| y = c\right) = \frac{\sum N_{x_{ij}=v \wedge y_i = c}}{N_{y=c}}$$

有一个问题是，分母可能为零。所以要在分母上加一个k， 分子加上1

$$p\left(x_j = v \middle| y = c\right) = \frac{\sum N_{x_{ij}=v \wedge y_i = c} + 1}{N_{y=c} + k}$$

连续型随机变量

连续型特征

$$f(x_i = x | y = c) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

$$\arg\max_c \; p(y=c) \prod_{i=1}^{n} f(x_i | y = c)$$

如果是连续的，假设服从正态分布

计算MLE

最大似然估计

$c: l$

(μ, σ^2)

$\downarrow$ MLE

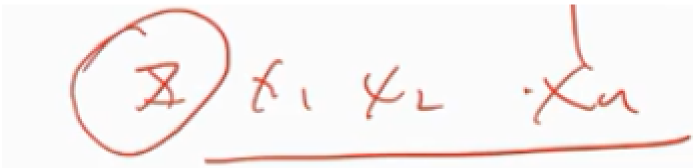
$$\mu = \frac{1}{l} \sum_{i=1}^{l} x_i$$

$$\sigma^2 = \frac{1}{l} \sum_{i=1}^{l} (x_i - \mu)^2$$

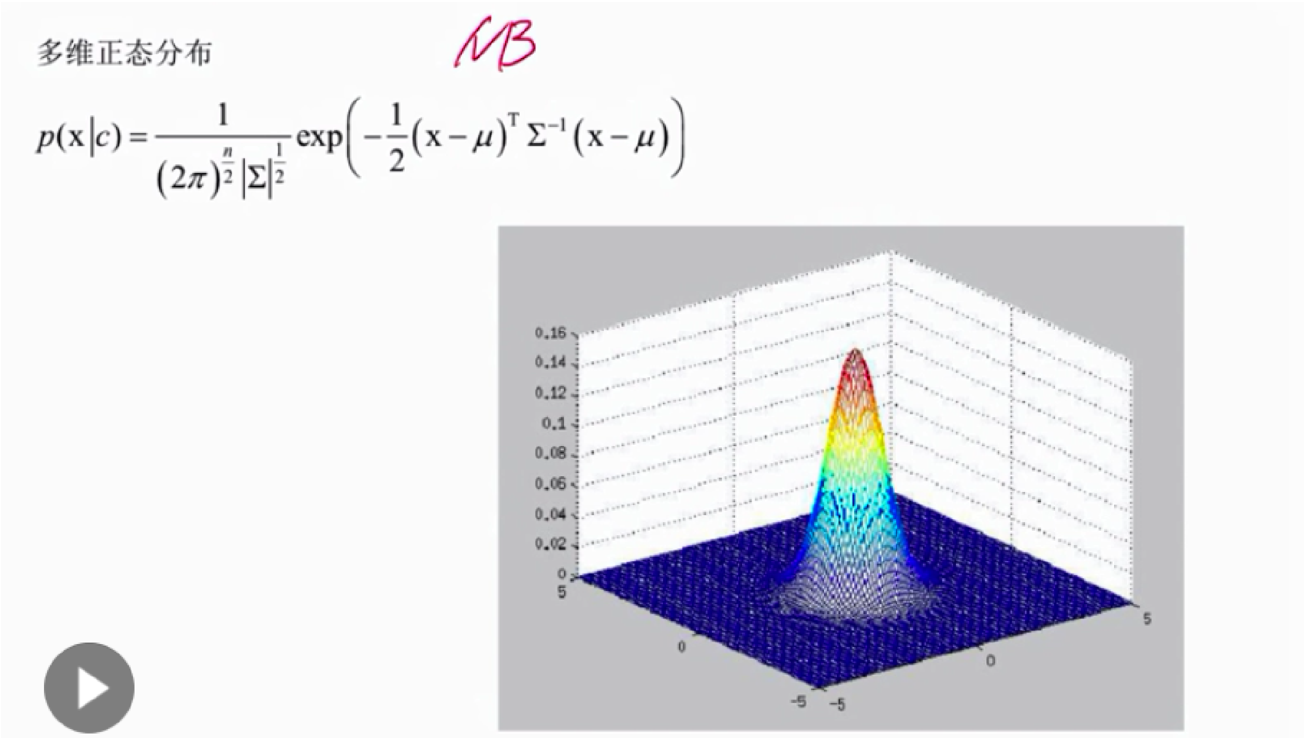
mu和sigma 算出来

总结朴素贝叶斯分类器

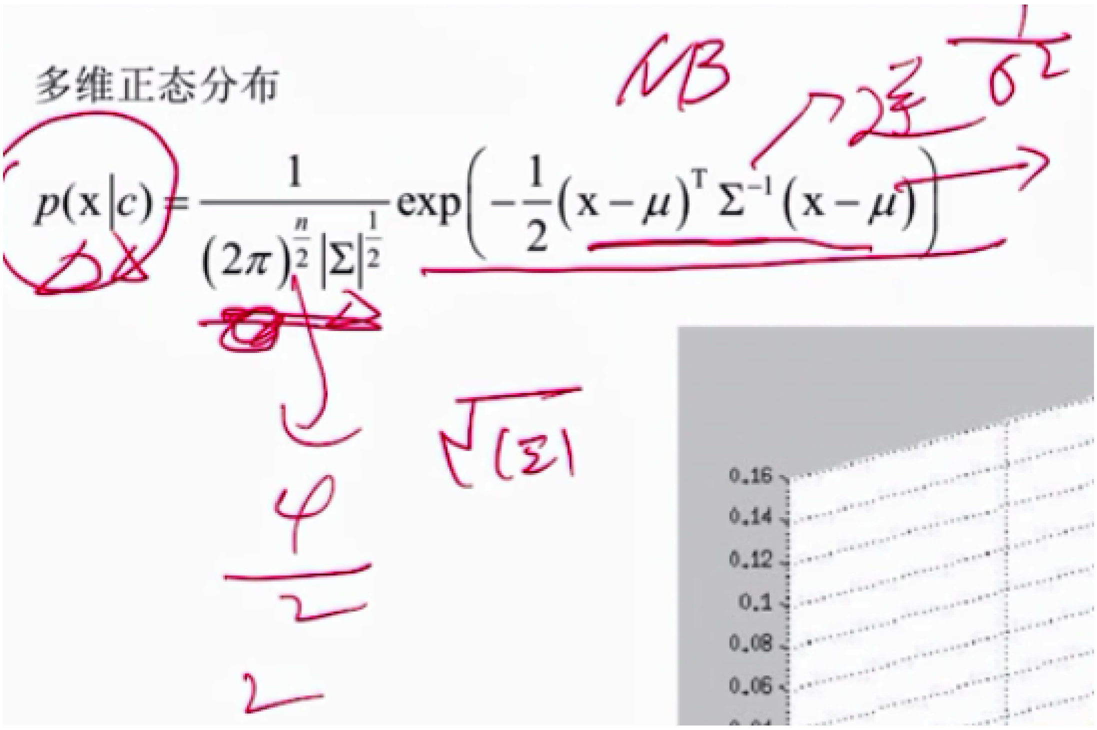
假设分量x1,x2,xn相互独立。



5.4 正态贝叶斯分类器



多维正态分布



p(x)

最大似然估计

正态贝叶斯分类器

正态贝叶斯分类器

$$p(\mathrm{x}|c)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\mathrm{x}-\mu)^T\Sigma^{-1}(\mathrm{x}-\mu)\right)$$

$$\Sigma=\mathrm{UWU}^T$$

$$(\Sigma)^{-1}=\left(\mathrm{UWU}^{-1}\right)^{-1}=\mathrm{UW}^{-1}\mathrm{U}^{-1}=\mathrm{UW}^{-1}\mathrm{U}^T$$

$$\arg\max_c\left(p(c|\mathrm{x})\right)=\arg\max_c\left(\frac{p(c)\,p(\mathrm{x}|c)}{p(\mathrm{x})}\right)$$

$$\arg\max_c\left(p(\mathrm{x}|c)\right)$$

$$\ln\left(p(\mathrm{x}|c)\right)=\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\right)-\frac{1}{2}\left((\mathrm{x}-\mu)^T\Sigma^{-1}(\mathrm{x}-\mu)\right)$$

$$\ln\left(p(\mathrm{x}|c)\right)=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln\left(|\Sigma|\right)-\frac{1}{2}\left((\mathrm{x}-\mu)^T\Sigma^{-1}(\mathrm{x}-\mu)\right)\qquad \ln\left(|\Sigma|\right)+\left((\mathrm{x}-\mu)^T\Sigma^{-1}(\mathrm{x}-\mu)\right)$$

$$
p(\mathrm{x}|c)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\mathrm{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathrm{x}-\mu)\right)
$$

$$
\Sigma=\mathrm{UWU}^{\mathrm{T}}
$$

$$
(\Sigma)^{-1}=\left(\mathrm{UWU}^{-1}\right)^{-1}=\mathrm{UW}^{-1}\mathrm{U}^{-1}=\mathrm{UW}^{-1}\mathrm{U}^{\mathrm{T}}
$$

协方差矩阵

实对称矩阵

带到贝叶斯分类器

$$
\arg\max_c\left(p(c|\mathrm{x})\right)=\arg\max_c\left(\frac{p(c)\,p(\mathrm{x}|c)}{p(\mathrm{x})}\right)
$$

$$
\arg\max_c\left(p(\mathrm{x}|c)\right)
$$

奇异值分解

$$
\ln\left(p(\mathrm{x}|c)\right)=\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\right)-\frac{1}{2}\left((\mathrm{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathrm{x}-\mu)\right)
$$

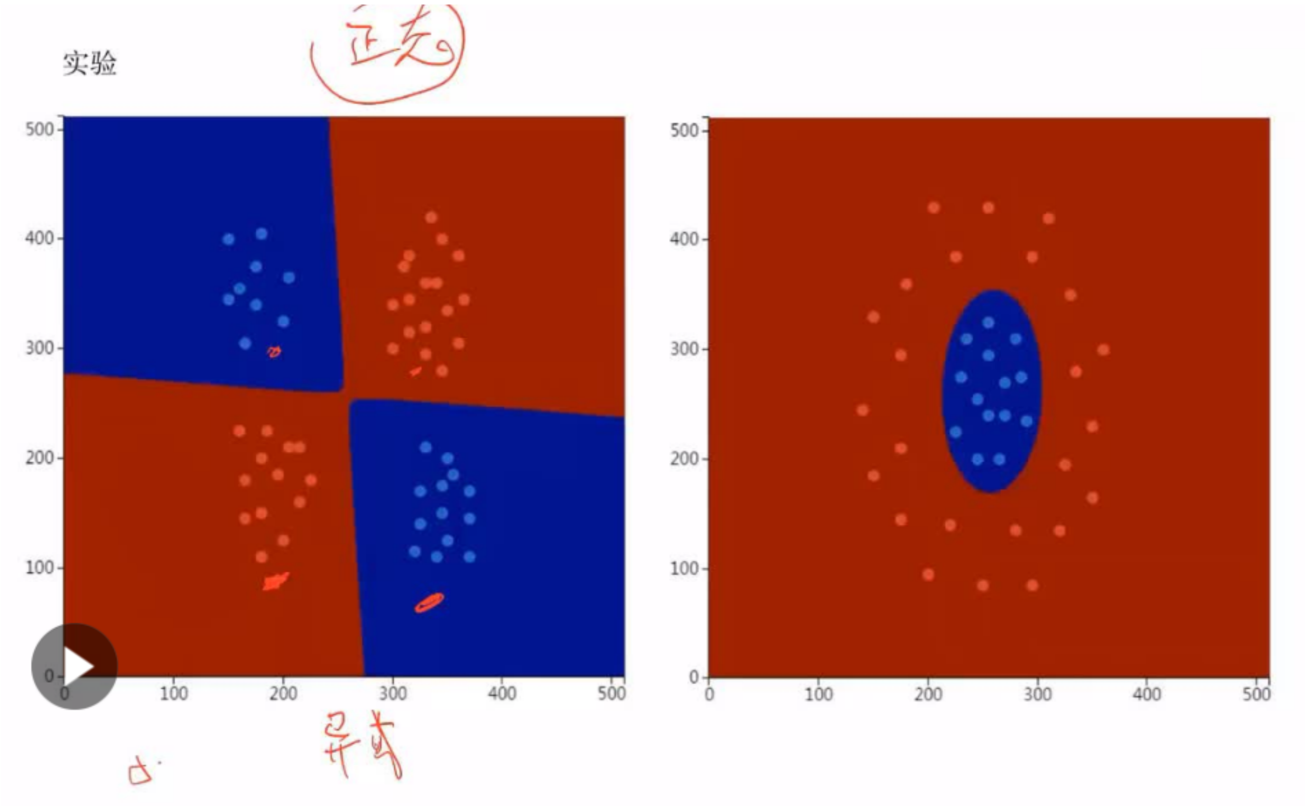
$$
\ln\left(p(\mathrm{x}|c)\right)=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln(|\Sigma|)-\frac{1}{2}\left((\mathrm{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathrm{x}-\mu)\right) \qquad \ln(|\Sigma|)+\left((\mathrm{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathrm{x}-\mu)\right)
$$

正态贝叶斯分类器，

$$
\begin{bmatrix}\mu\\ \Sigma\end{bmatrix}
$$

$$
\ln(|\Sigma|)+\left((\mathrm{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathrm{x}-\mu)\right)
$$

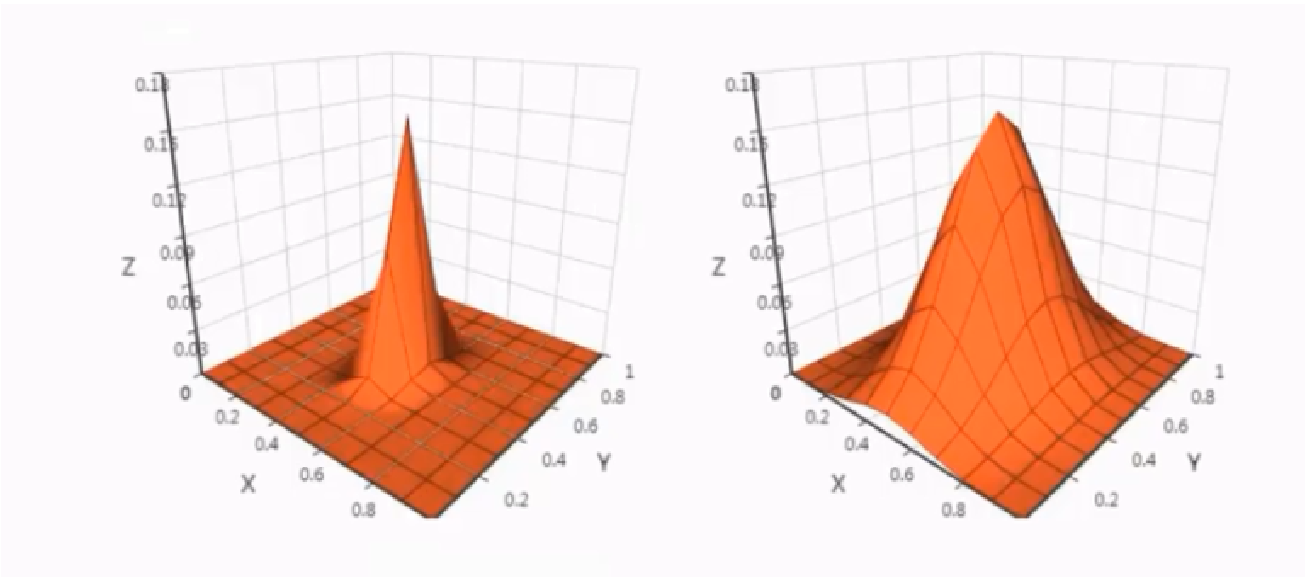
5.5 实验



实验环节

贝叶斯分类器有非线性的模型。

点击打开训练集



5.5 实际应用

应用
垃圾邮件分类问题
Mehran Sahami, Susan T Dumais, David Heckerman, Eric Horvitz. A Bayesian Approach to Filtering Junk E-Mail. 1998.
文本分类问题
Yiming Yang, Xin Liu. A re-examination of text categorization methods.international acm sigir conference on research and development in information retrieval, 1999.
智能视频监控中的背景建模算法
Liyuan Li, Weimin Huang, Irene Y H Gu, Qi Tian. Foreground object detection from videos containing complex background. acm multimedia, 2003.
人脸识别中的联合贝叶斯模型
Dong Chen, Xudong Cao, Liwei Wang, Fang Wen, Jian Sun. Bayesian face revisited: a joint formulation. european conference on computer vision, 2012.

- 垃圾邮件分类
- 文本分类问题
  - TF-IDF 比如分
- 智能视频监控中的背景建模算法
  - 判断图片是全景还是背景，是个二分类问题
- 人脸识别中的联合贝叶斯模型

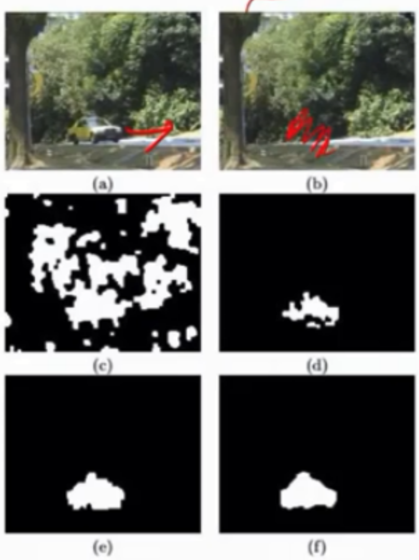
判断是全景还是

$$P(C|\mathbf{v}_t, s) = \frac{P(\mathbf{v}_t|C, s)P(C|s)}{P(\mathbf{v}_t|s)}, \quad C = b \text{ or } f$$

$$P(b|\mathbf{v}_t, s) > P(f|\mathbf{v}_t, s)$$

$$P(\mathbf{v}_t|s) = P(\mathbf{v}_t|b, s) \cdot P(b|s) + P(\mathbf{v}_t|f, s) \cdot P(f|s)$$

$$2P(\mathbf{v}_t|b, s) \cdot P(b|s) > P(\mathbf{v}_t|s)$$



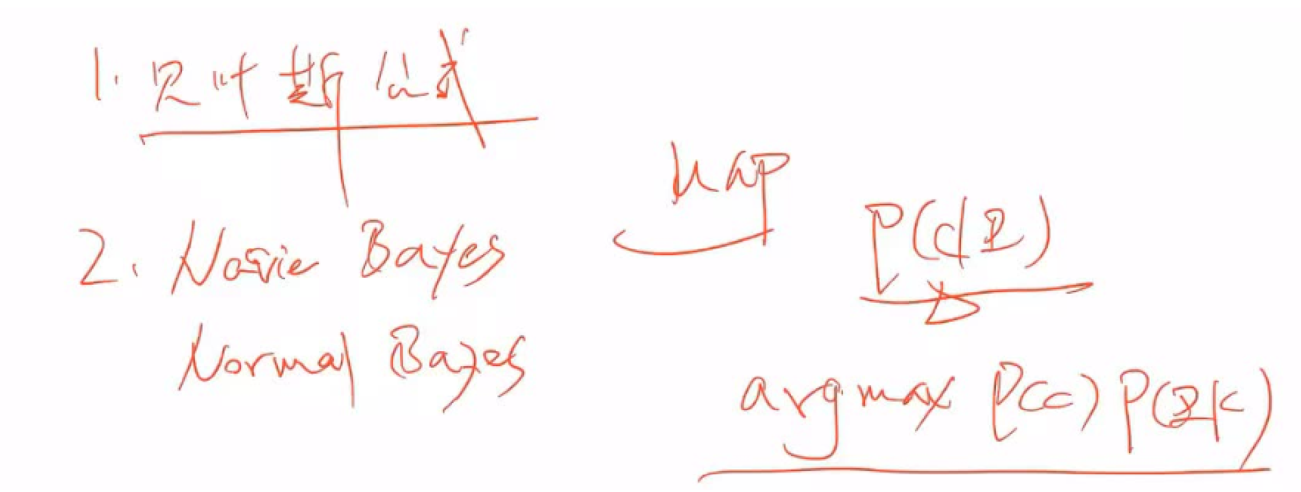
$$P(C|\mathbf{v}_t, s) = \frac{P(\mathbf{v}_t|C, s)P(C|s)}{P(\mathbf{v}_t|s)}, \quad C = b \text{ or } f$$

$$P(b|\mathbf{v}_t, s) > P(f|\mathbf{v}_t, s) \rightarrow b$$

$$P(\mathbf{v}_t|s) = P(\mathbf{v}_t|b, s) \cdot P(b|s) + P(\mathbf{v}_t|f, s) \cdot P(f|s)$$

$$2P(\mathbf{v}_t|b, s) \cdot P(b|s) > P(\mathbf{v}_t|s)$$

5.6 本集总结



1. 贝叶斯公式
2. Bayes，正态贝叶斯
3. 实验的环节

自我总结：这节课的知识挺难的，涉及到的知识面也是比较广。

我的why

为什么朴素贝叶斯是“朴素”的？

Answer：“朴素”是因为它假设了数据集中的**所有特征是同等重要的并且是条件独立的。**然而，这是一个很强的假设，在实际情况中，这个假设通常很难严格成立。

Q5：解释朴素贝叶斯算法中的先验概率、似然和边缘似然概念？

Answer：*先验概率**代表了数据集中因变量的比例，它是指你不依靠额外的信息能做出最有可能的类别猜测。比如，在一个垃圾邮件分类的数据集中，因变量是二进制的（0和1），1（垃圾邮件）的比例是70%而0（非垃圾邮件）的比例是30%，据此，我们可以估计一个新的邮件有70%的可能是垃圾邮件。

似然是指存在其他变量的情况下，把一个给定的观测分类为1的概率。比如，“FREE”这个词出现在垃圾邮件中的概率就是似然。**边缘似然是“FREE”出现在任何邮件中的概率。**