Bayes' theorem

Likelihood

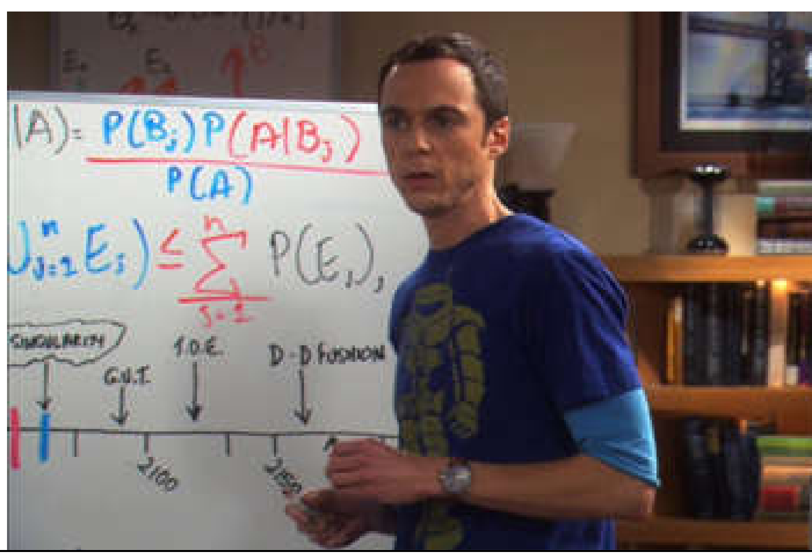
Prior probability

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$

Posterior probability

Evidence

A Puzzle：
校园遇到一个shy的男同学， 他是Math Phd还是Business School的？



A Puzzle：
校园遇到一个shy的男同学，他是Math Phd还是Business School的？

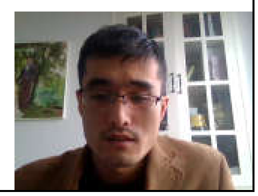
Q:

P (math | shy) 和 P (business | shy)

A:

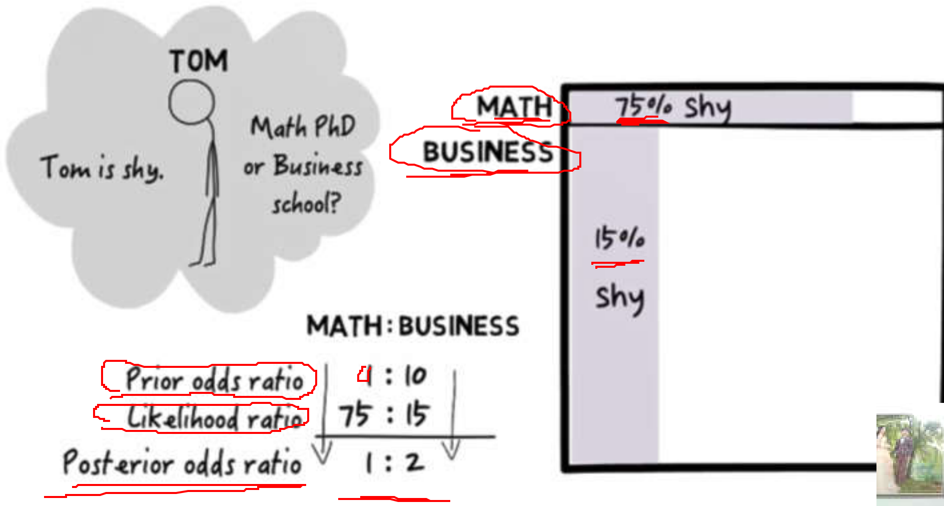
P (shy | math) 和 P (shy | business)

(大部分同学比较了这两个概率)



A Puzzle：
校园遇到一个shy的男同学，他是Math Phd还是Business School的？

A Puzzle：
校园遇到一个shy的男同学，他是Math Phd还是Business School的？

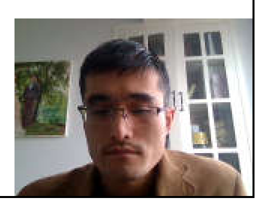


贝叶斯公式：

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

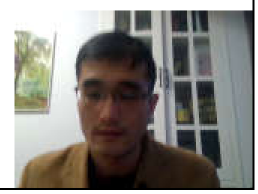
What Bayes tells us: Don't forget the prior!

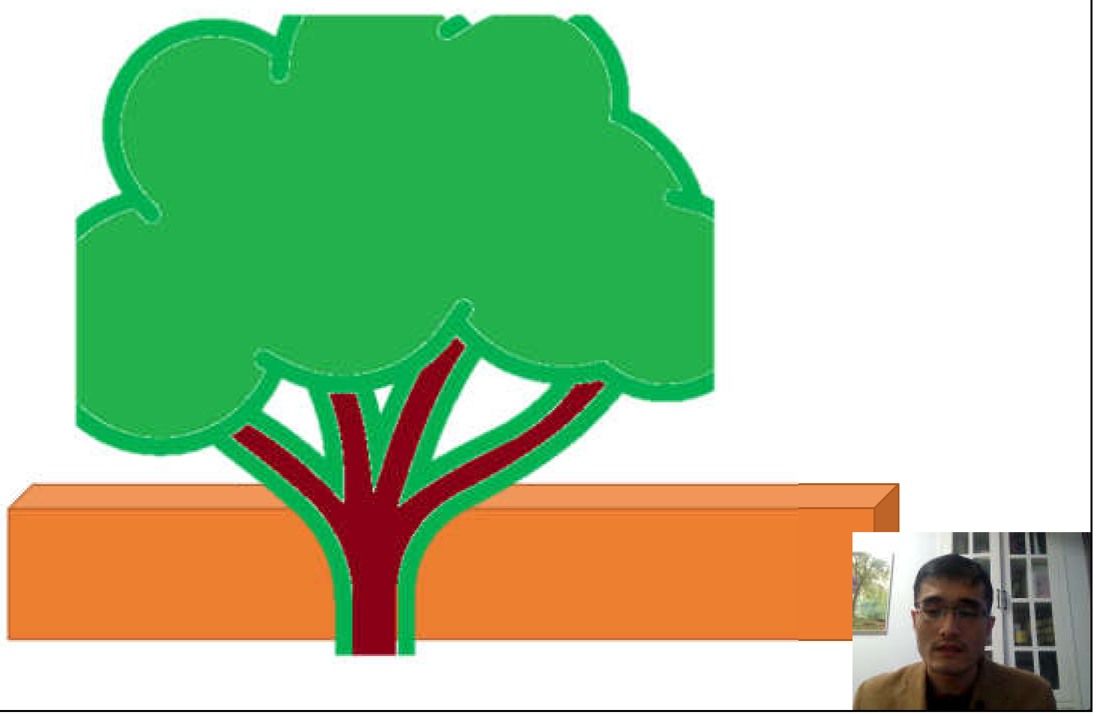


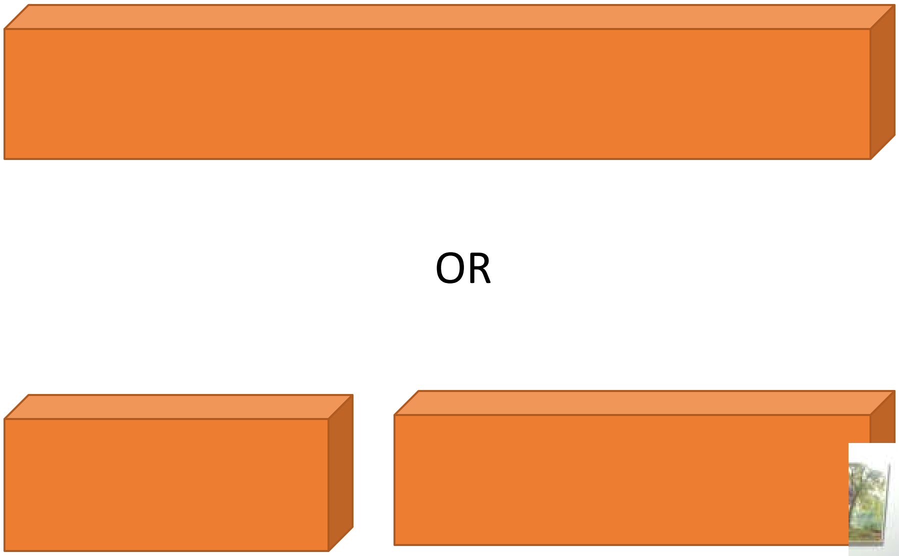
贝叶斯公式：

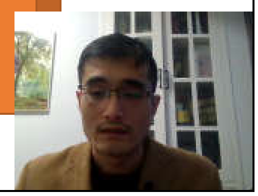
$$P(B|A) = \frac{P(A|B)\boxed{P(B)}}{P(A)}$$

$$\text{Posterior} = \frac{\text{likelihood} \times \boxed{\text{prior}}}{\text{evidence}}$$





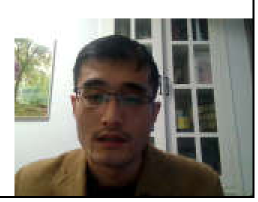




贝叶斯公式：

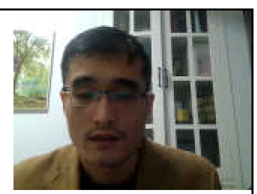
$$P(B|A)=\frac{\boxed{P(A|B)}P(B)}{P(A)}$$

$$\text{Posterior}=\frac{\boxed{\text{likelihood}}\ \text{x}\ \ \text{prior}}{\text{evidence}}$$



似然 (likelihood)

$$L(\Theta|x)=P(X=x|\Theta)$$



极大似然估计 (MLE)

$$L(\Theta|x)=P(X=x|\Theta)$$

Maximum Likelihood Estimate (MLE): 找出参数Θ，使得从中抽样所得的观测数据的概率最大

$$L(\theta)=L(x_1,x_2,\ldots,x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta)$$

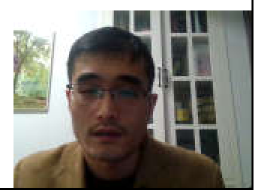
贝叶斯公式应用I – 分类器：

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

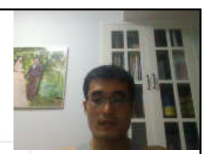
$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

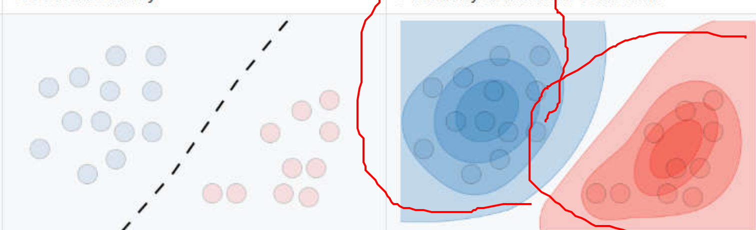
贝叶斯分类器：

$$p(类别|特征) = \frac{p(特征|类别)p(类别)}{p(特征)}$$



判别模型 vs 生成模型（support sampling）

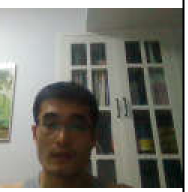


	Discriminative model	Generative model
Goal	Directly estimate $P(y\|x)$	Estimate $P(x\|y)$ to then deduce $P(y\|x)$
What's learned	Decision boundary	Probability distributions of the data
Illustration		
Examples	Regressions, SVMs	GDA, Naive Bayes

A **generative algorithm** models how the data was generated in order to categorize a signal. It asks the question: based on my generation assumptions, which category is most likely to generate this signal?
A **discriminative algorithm** does not care about how the data was generated, it simply categorizes a given signal.

生成模型对数据的生成方式进行建模。它提出一个问题：根据我的生成方式假设，哪个类别(y)最有可能产生当前的特征（X)？需要对P(X|Y)进行建模。
判别模型不关心数据是如何生成的，它只是对给定的特征进行判别/分类。直接对P(Y|X)进行建模。

判别模型 vs 生成模型（support sampling）



贝叶斯公式应用 II - 医学诊断：

$$p(\text{病因}|\text{症状}) = \frac{\mathrm{p}(\text{症状}|\text{病因})p(\text{病因})}{\mathrm{p}(\text{症状})}$$

	Well	Cold	Allergy
P(d)	0.9	0.05	0.05
P(sneeze|d)	0.1	0.9	0.9
P(cough|d)	0.1	0.8	0.7
P(fever|d)	0.01	0.7	0.4

贝叶斯公式应用 II - 医学诊断：

$$p(\text{病因}|\text{症状}) = \frac{\mathrm{p}(\text{症状}|\text{病因})p(\text{病因})}{\mathrm{p}(\text{症状})}$$

	Well	Cold	Allergy
P(d)	0.9	0.05	0.05
P(sneeze|d)	0.1	0.9	0.9
P(cough|d)	0.1	0.8	0.7
P(fever|d)	0.01	0.7	0.4

贝叶斯公式应用 II - 医学诊断：

$$p(\text{病因}|\text{症状}) = \frac{\mathrm{p}(\text{症状}|\text{病因})p(\text{病因})}{\mathrm{p}(\text{症状})}$$

早期AI：

An attempt to simplify probabilistic reasoning in 1960s medical diagnostic programs

贝叶斯公式应用 III - NLP：