机器学习

第五章: 概率参数估计

2020 年秋季

指导老师: 顾晓东



今天

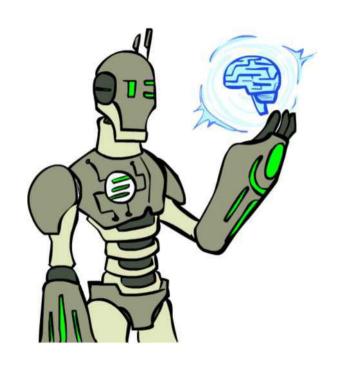


机器学习的概率视角

如何估计概率模型的参数?

最大似然估计

-贝叶斯估计

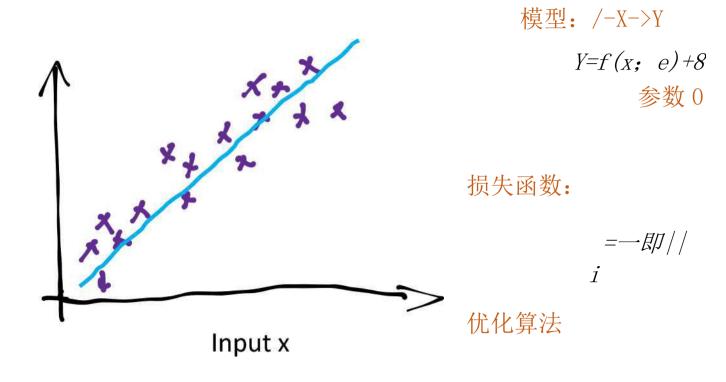


顾晓东

回忆: 机器学习作为功能近似



我们对这些数据做了什么假设?

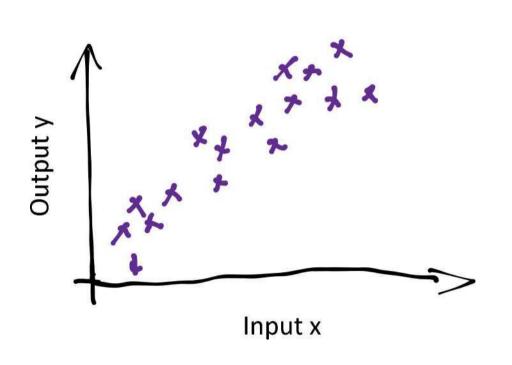


概率密度估计

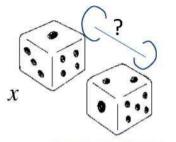


y

再次: 我们对这些数据做了什么假设?



模型: p(x, y)



parameters

x	у	p(x,y)
1	1	0.4
1	2	0.1

• loss function?

优化算法?

顾晓东

二项分布的概率密度估计

SE TONG UNITED SECTION OF THE PARTY OF THE P

我们观察到 N 个独立的、相同分布的(i.i.d) 抛硬币

(Bernoulli),并想估计硬币的形状



nmi s

数据: D={H, T, H, H, T, T}

模型: 带参数3的伯努利分布

设 x={0=尾巴, 1=头), p(x=1), 3, p(x=0)=1-0

这个问题的最优 3 是什么?

/ 1

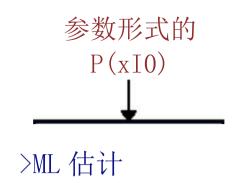
目的:找到最符合观察 D 的 3

顾晓东

找到一个 0, 它导致观察到的 D 的最大似然(概率

最大似然估计 =argmax.p(0|0)





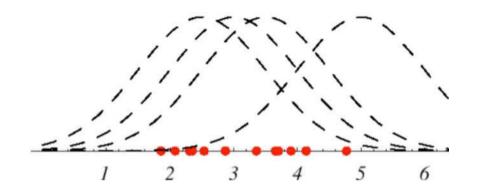
数 据 准 备 好

最好的参数 〉®毫升 估计数

顾晓东

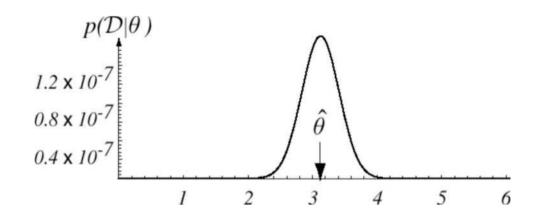
最大似然估计

Example:



顾晓东

最大似然估计



顾晓东

二项分布的 MLE

SAN TIBBE OF THE SECOND OF THE

■ 数据:观察到的集合 D 的 a_h 头和 a_t 指甲

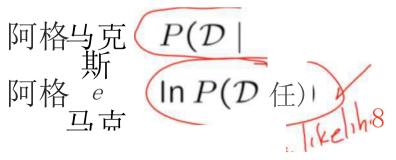
■ 假设: 二项分布

■ 学习 0 是一个优化问题

□什么是目标函数?

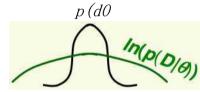
■MLE:选择 0,使之最大化

观测数据:



probability of

而不是 最大化 p(D|0), 通常更 容易最小化 -LN(p(D|0), 因为日志 是单调的



顾晓东

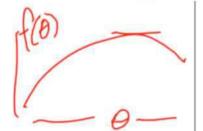
二项分布的 MLE



6=Arg max in F(7?)

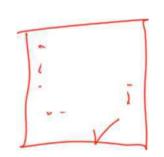
(0)

 $= argmax 在 0^{aH} (1-0)^{aT}$

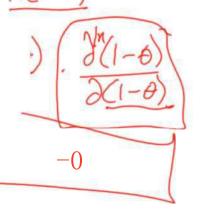


■将导数设为零:





又 k 凱. 十/&



/ 5项"

$$\partial = \frac{Xh}{Xi+}$$

顾晓东



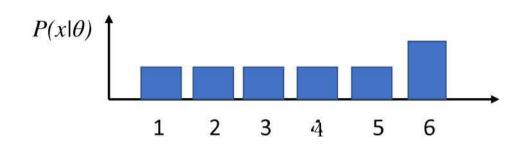
$$\hat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H + \alpha_T}$$

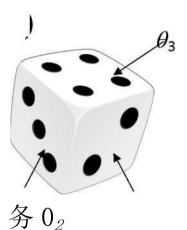
$$\frac{3}{3+2} = .6 = MLE\theta$$
 $\frac{300}{300+200} = .6 = MLE\theta$

多项式分布的 MLE



从硬币到骰子(K的结果而不是2)





每个翻转产生一个 K 值变量 x==分类(仞, 0={0 ៣果

i.i.d 的数据集 D。 翻转产生一个(x=1), a₂ (x=2), 。。。

顾晓东

多项式分布的 MLE



k

伽)=(a ".")冬館.破 "=去口 目的:

最大€(0)三日志(N!)么日志+么日志

s. T. Z&IQ=1

双重问题:

k

最小g(0,2)三日志(N!)-}日志(时)+}%日志色久(1一}劣)
" "我------ 我=我

顾晓东

多项式分布的 MLE



k

$$dg \ a_t$$
 ---- a

让

籌羞。

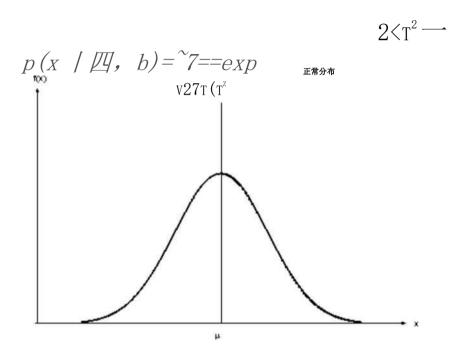
d3i Bi

顾晓东

高斯(正常)分布置薯的 MLE



- 9 连续随机变量 x 遵循单变量正态分布 cr² 具有平均川和方差 a².
- 9 概率密度函数(参数为 0=(X_")2-



顾晓东

高斯(正态)分布的 MLE

顾晓东



Inpg1。) =—:
$$\{\ln(2? \text{ r}) + \ln a^2 + \frac{\#}{m} \text{ n } A = 0 \Rightarrow \$, \text{ MP}(XA|\circ)\} = 0$$

$$\sum_{k=i}^{n} \left[\frac{1}{\sigma^2} - \frac{(Xk-")2a^4}{\sigma^2_{ML}}\right] + \sum_{k=1}^{n} \left[\frac{1}{n} \sum_{k=1}^{n} (x_k - \mu)^2\right]$$

$$\Rightarrow \sigma^2_{ML} = \frac{1}{n} \sum_{k=1}^{n} (x_k - \mu_{ML})^2$$

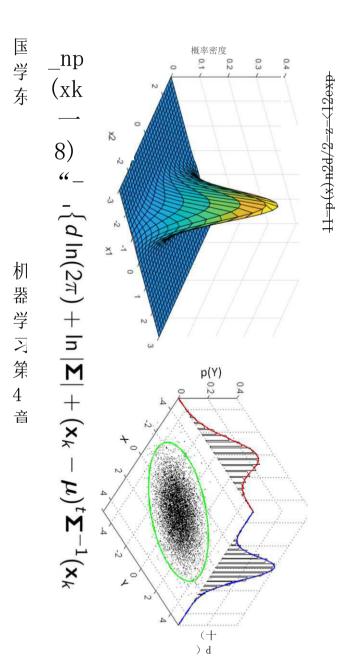
$$\boxed{\text{用 Pml 做 P}} \qquad \sigma^2_{ML} = \frac{1}{n} \sum_{k=1}^{n} (x_k - \mu_{ML})^2$$

顾晓东

用于高斯 正 常 Di sf fi bu Но n 的 ML Е



穆希瓦丘 e 高斯



高斯(正态)分布的 MLE



多元高斯

在 p(xa|0)中=-{d ln(27r)+在|E中(xr-
$$^{1}(x/\sqrt{-})$$
}

(使用标识 f 轴】=2Ax)

 n
=0 今 E^[(xA-G=T(xA-G]=0)
 $^{k=1}$
£T(xA-M)=0

高斯(正态)分布的 MLE

 \sim n

, xk 样本平均 k=1

多元高斯

同样,可以表明:

-h) (xk-

k=1

n

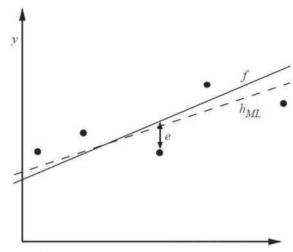
。cf。 真协方差矩阵: E(x-Ex)(x-Ex)')

顾晓东

ML 和最小二乘误差



目标函数: 任何实值函数 f



培训实例(x; d/)

9DJ 很吵: DJ-f(x ')+e;

为每个 x 独立绘制的随机变量;根据对于平均=为 0(e; A/(0, a) 的高斯分布 $^2)$) 独立和相同分布的 0(i.i.d.o.)

顾晓东

ML 和最小二乘误差



h. H

V2? ra² 2°

昵以月£(击一/? (x

我=我

最大似然假设是最小化的假设

平方误差之和

顾晓东



贝叶斯估计

ML 与贝叶斯



```
ML 估计:
```

假设是固定的,但未知 ®想找到 0ml=argmax.p{D\0)

```
的参数形式
                             p(xi9)
           数据
                  ◆ML 估计
          准备
贝叶斯学
 假设 90 是一个随机变量,具有先验分布 P(。)
 想要找到 p(3D)(后验分布)
                       的参数形式
                      贝叶斯
              d5t D
                              >p (o | d)
                                   分配
                      估计数
                        P(xi0)
                           P(e)
```

事先分发

顾晓东 机器学习:第四章

贝叶斯学习

使用 Bayes 规则:

MLE = argmaxP (@|0)



$$p(o|q) = \frac{CD(0)p(0)}{e}$$

$$= A \quad p \cdot \frac{A}{p} \quad \frac{A}{rgmax} \quad \frac{A}{rgm$$

顾晓东 机器学习:第四章

或等效:

$$P(0|Q) \text{ oc} P(7)|0) P(6Q)$$

最大后效(MAP)



MLE 和 MAP

最大似然估计 (MLE): 选择 0, 使观测数据的概率最大化 2? $0=\arg$ 最大 F(P|0)

最大后向估计(MAP):选择 0,这是最有可能的先前给定概率和数据

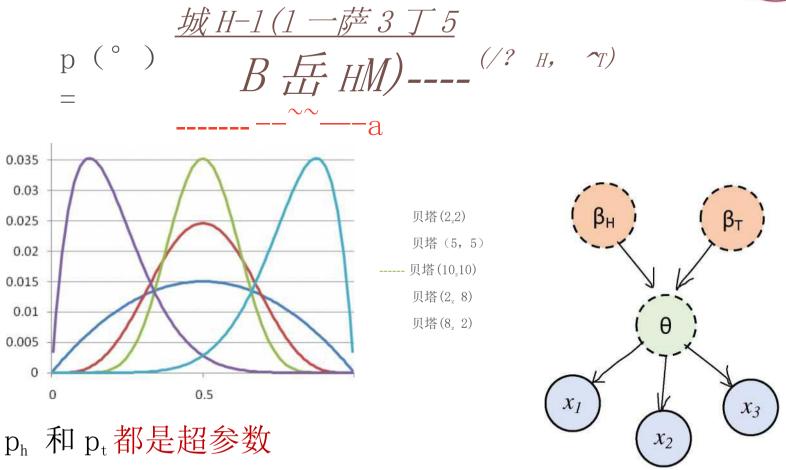
$$0= \arg \max_{o} p(0i p)$$

$$= Argmax=o \frac{p(p i0)p(0)}{p(q)}$$

顾晓东

贝塔事先分配一P(0)



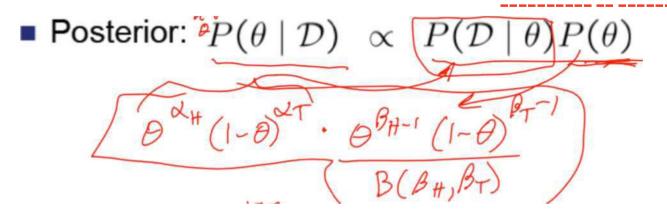


顾晓东

后向分布



■似然函数: 切 7? | 爲 1 § 叫 (1 一 o 而



后位作为0的函数与前位具有相同的形式(Beta分布。

顾晓东

贝塔分布地图



arsmax 8 4+BH-1

不/业寸一__

[C. Guestrin]

▶ 贝塔之前相当于额外的缩略图翻转

- ●作为 N-8, 先验被"遗忘"
- 对于小样本大小,优先是重要的! r.。

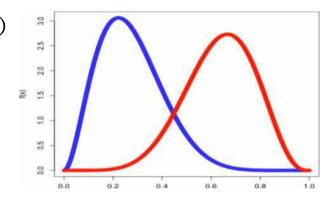
但



0是硬币抬头的概率

模型 0 与 2 种不同的先验:

- 。 P^OJis Beta (3, 8)(蓝色)
- p₂ (0) 是贝塔(7, 4) (红色) 他们看起来很不一样!

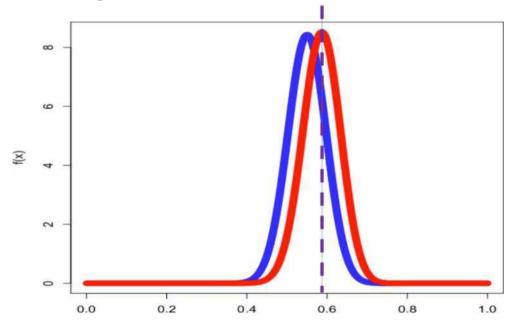


- 。 现在翻转 100 个硬币;得到 58 个正面和 42 个尾巴
 - 。 后辈长什么样?

顾晓东



argmax 返回模式



只要我们收集足够的数据,后辈就会收敛到真正的价值!

顾晓东

联合牧师



P(0)和P(0D)的形式相同

例如。 1 硬币翻转问题 似然是 $^{\sim}$ 二项式 p(q=)叫(1-0)皿 i

如果先验是贝塔分布[^]

然后^^^后辈献礼

 $^{\sim}$ BetQ(/3ff+Qh, $+^{a}t$)

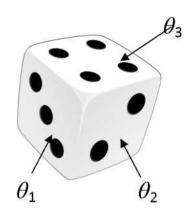
对于二项式, 共轭先验是贝塔分布。



联合牧师



多项式分布的共轭先验是什么?



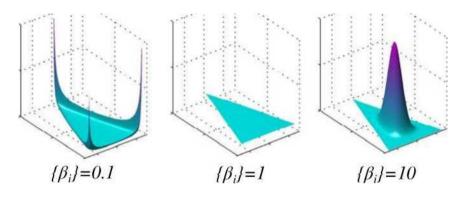
顾晓东

Dirichlet 分布



=(务、...>Z&I 饥=。

。 乙(6)是用伽马函数表示的 nomalizing



顾晓东 机器学习: 第四章



Peter Gustav Lejeune



Peter Gustav Lejeune Dirichlet

天生的 Johann Peter Gustav Lejeune

迪里切莱特

131805年2月杜伦, 法兰西帝国

死了 1859年5月5(54岁)

哥廷根,汉诺威王国

国籍 德语

已经知道了 见完整清单

奖项 博士 (荣誉):

波恩大学 (1827年)

常数:

联合牧师

例如。2骰子滚动问题(6个结果而不是2个)

似然是 $^{\sim}$ 多项式(0={ $0_{19} \ 0_{2}, \ldots, \ 0_{k}$ })

我=戒 i 彼 2...0 徉

如果之前是迪里切莱特分布,

 **rk
 0 依-1

 **P中 转液 5 ** 婆岫 **...,附

然后是后验是 Dirichlet 分布

P(6>|D)-Dirichlet(/3i+顷,以Q...

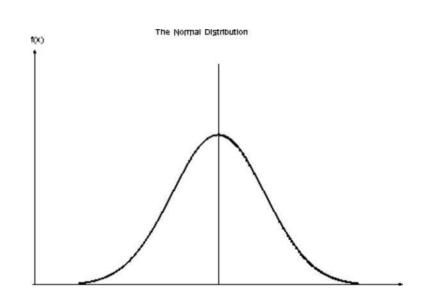
对于多项式,共轭先验是狄里克莱分布。



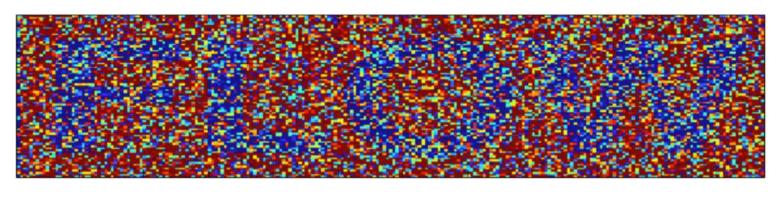
联合牧师



高斯的先验共轭是什么?







²观察图像⁴⁰

40 如何去除噪声,恢复原图像?

顾晓东 机器学习: 第四章



假设传感器模型为

$$\mathcal{F}=UI+Wi$$

在哪里

$$uie \{0, 1\} wizN(0, cr^2)$$

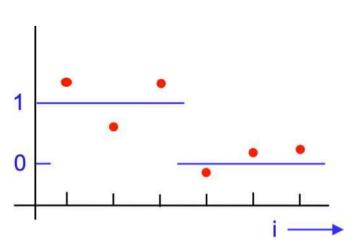
那么可能性是

 $\chi Z|U)$ 他一(饲_啊)勺(2成)i假设前面有表格

P(U)z

然后后 p(u|z)的负日志是

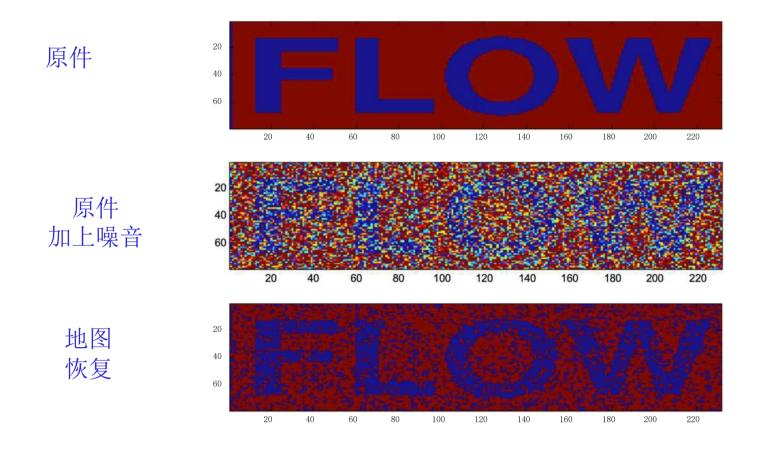
""可能事先



订 MAP=ArgminC(U)



d=10

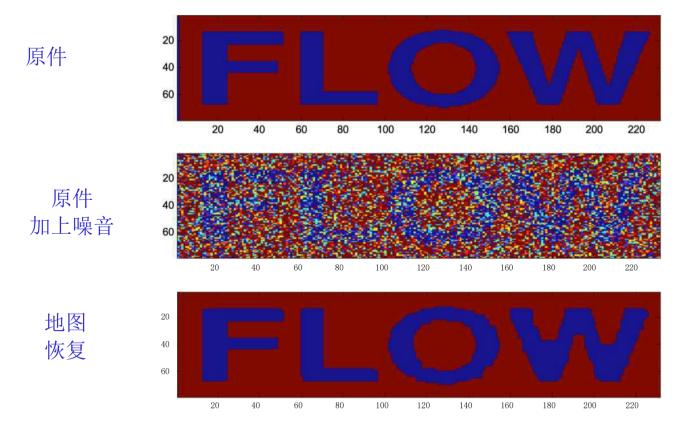


顾晓东

机器学习: 第四章



d = 60



顾晓东 机器学习:第四章