机器学习工程师微专业
作业及答案

03 树模型初步与进阶

第三章 最大熵与 EM 算法

讲师：加号

选择题（2 分/题）

1.关于贝叶斯公式 P(y|x)=(P(x|y) * P(y))/P(x)下列表述错误的是：B

A.P(y|x)是后验概率，一般是我们求解的目标

B.P(x|y)是后验概率，可以统计得到

C.P(y)是先验概率，一般都是人主观给出的。

D.P(x)是先验概率，往往用全概率公式计算得到

2.关于最大似然下列表述错误的是：A

A.对数似然函数可导的情况下，让导函数等于 0 的就是我们要求解的参数

B.极大似然的思想是利用已知的样本结果，反推最有可能（最大概率）导致这样结果的参数值

C.对数似然行数导数为 0 的驻点可能不是极值点

D.极大似然估计通过若干次试验，观察其结果，利用试验结果得到某个参数值能够使样本出现的概率为最大。

3.下列描述不正确的是：D

A.信息熵是对随机变量平均不确定性的度量

B.分布越不确定信息熵越大

C.最大熵原理认为，学习概率模型时，在所有可能的概率模型中，熵最大的模型是最好的模型。

D.随机变量取所有值的概率都相等的时候，信息熵为 0

判断题（2 分/题）

1.分布越均匀，熵越大：对

2.互信息取值可以为负数：**错**

3.EM 算法中的 E 步骤估计未知参数的期望值，给出当前的参数估计；M 步骤重新估计分布参数，以使得数据的似然性最大，给出未知变量的期望估计。：对

计算题 1（10 分/题）

已知一个箱子里有 10 个球，两个白的，八个红的。请问，从这个箱子里有放回地一次次抽球的熵是多少?

$$H = -x\log(x)-(1-x)\log(1-x)$$

其中 x=0.2

计算题 2（10 分/题）

1）在自然语言处理中，已知：

a.“学习”可能是动词，也可能是名词。

b.“学习”可以被标注为主语、谓语、宾语、定语……

未知量：x1、x2 分别表示“学习”是动词和名词，y1、y2、y3 和 y4 分别表示“学习”是主语、谓语、宾语和定语。

请问，在无任何偏颇的情况下，每个未知量的概率是？

答案：p(x1)=p(x2)=0.5,p(y1)=p(y2)=p(y3)=p(y4)=0.25

2）若又已知 c：“学习”被定义为定语的可能性很小，p(y4)=0.05,

请问，那么剩下的概率坚持无偏原，分别是：

答案：p(y1)=p(y2)=p(y3)=0.95/3

3）又已知 d：当“学习”被标作动词的时候，它被标作谓语的概率是 0.95,p(y2|x1)=0.95。

请根据以上所有条件，写出按照最大熵原理计算 X 和 Y 的分布并使 H(Y|X)达到最大值时，需要满足的条件：

答案：

p(x1)+p(x2)=1

p(y1)+p(y2)+p(y3)+p(y4)=1

p(y4)=0.05

p(y2|x1)=0.95

计算题 3（10 分/题）

假设古今中外练成降龙十八掌所需要花费的时间符合高斯分布。已知，洪七公是练武奇才，花了 9 周，黄蓉天资聪颖，花了 9.5 周，而郭靖比较笨，花费了 11 周。现在，天天被大武小武欺负的杨过想偷学这门武功，他对自己的资质并不自信，想利用 MLE 估

计出这个高斯分布，并看看普通人学降龙十八掌所需的平均时间（期望）是多少？（请使用 MLE 步骤解答）

Hint：

高斯分布（正态分布）公式如下：

$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

包含两个参数，μ（期望）和 σ （方差）。我们要求的就是这个期望值μ。另，当我们对一个未知数求导的时候，另一个未知数被看做常数。

答案：

1）把 9， 9.5 和 11 带入公式，我们得到联合概率公式：

$$P(9,9.5,11;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(9-\mu)^2}{2\sigma^2}\right)\times\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(9.5-\mu)^2}{2\sigma^2}\right)\times\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(11-\mu)^2}{2\sigma^2}\right)$$

跟课上讲的 MLE 步骤一样，我们只要最大化μ就可以了。

2）但是这个公式最大化本身比较难算，所以我们两边取对数：

$$\ln(P(x;\mu,\sigma))=\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{(9-\mu)^2}{2\sigma^2}+\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{(9.5-\mu)^2}{2\sigma^2}+\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{(11-\mu)^2}{2\sigma^2}$$

化简得：

$$\ln(P(x;\mu,\sigma))=-3\ln(\sigma)-\frac{3}{2}\ln(2\pi)-\frac{1}{2\sigma^2}\left[(9-\mu)^2+(9.5-\mu)^2+(11-\mu)^2\right]$$

这就是我们的对数似然值。

3）为了求这个时候的最大值，接下来我们直接求导：

$$\frac{\partial \ln(P(x;\mu,\sigma))}{\partial \mu} = \frac{1}{\sigma^2}\left[9 + 9.5 + 11 - 3\mu\right].$$

4）导数等于零，得到极值点：

$$\mu = \frac{9 + 9.5 + 11}{3} = 9.833$$

答：平均时间为 9.833 周

主观题（10 分/题）

分别用一句话阐述 EM 算法中，E 步骤和 M 步骤的作用：

E：计算期望，检测是否是最大值

M：最大化参数

课程链接：http://course.study.163.com/400000002658002/learning

如有问题，请咨询稀牛学院客服微信