

# 期末考试模拟题2（押题卷）

#ML/DS

## 题目 1

有两个罐子，罐子 A 和罐子 B。每个罐子里都有红球和蓝球。罐子 A 中红球的比例是  $p$ ，罐子 B 中红球的比例是  $q$ 。我们从这两个罐子中抽球，但不知道每次抽球是从哪个罐子抽的。现在有一个观察数据集：

数据点1: 红, 红, 蓝  
数据点2: 红, 蓝, 蓝  
数据点3: 蓝, 红, 红

从  $p = 0.6, q = 0.4$  开始迭代一次，估算  $p$  和  $q$  的值

## 题目 2

我们有三篇文档和初始化的主题：

文档一：苹果(A)，数学(B)，物理(A)  
文档二：苹果(B)，数学(C)，足球(A)  
文档三：物理(B)，足球(B)，数学(B)

如果我们对于每个主题都有  $\alpha_k = 0.01$ ，对于每个单词都有  $\beta_v = 0.02$ ，现在，在使用 Gibbs 抽样算法训练 LDA 模型的过程中，要为文档二中的“足球”重新分配主题，请计算“足球”属于 A、B、C 三个主题的概率。如果为文档一中的“苹果”重新分配呢？

## 题目 3

给出三个数据点和它的类别

(2,5)，正例  
(1,4)，正例  
(-3,-6)，反例

计算一个线性的支持向量机对这些数据点进行分类

## 题目 4

迈尔斯-布里格斯类型指标（Myers-Briggs Type Indicator）是由美国作家伊莎贝尔·布里格斯-迈尔斯和她的母亲凯瑟琳·库克·布里格斯共同制定的一种人格类型理论模型。该指标通过四个维度上的划分（E/I：外倾/内倾；S/N：实感/直觉；T/F：理智/情感；J/P：判断/理解）来将人格划分为 16 个类型。作为一名充满正义感的银河棒球侠，你认为星间列车上最近的气氛有些沉闷，于是你决定找找列车上哪些人的性格比较相近，以便让他们成为更好的朋友。你收集了五位列车乘员的四项指标得分如下：

名字缩写	I/内倾得分	N/直觉得分	T/理智得分	P/理解得分
TR	3.23	4.21	1.19	0.45
M	-5.42	-0.45	-2.29	0.13
W	2.13	1.03	-2.77	-0.45
D	3.09	0.21	1.35	-0.96
H	-1.46	0.71	-0.23	-0.43

- (1) 请你使用经过 Z-Score 标准化后的数据之间的哈密顿距离作为数据之间的距离度量，写出成对距离矩阵
- (2) 请你使用数据点间的平均距离作为类间距离的度量依据，利用（1）中求得的成对距离矩阵，完成凝聚的系统聚类，画出聚类谱系图
- (3) 现在，利用（1）中得到的距离矩阵，在距离小于 0.5 的两个人之间连接一条无向边，构成一张图。计算这张图的 Laplace 矩阵。

题目 5

假设有一个小镇，这个小镇的天气有两种状态：晴天 (S) 和雨天 (R)。然而，我们无法直接观测到天气状态。我们只能通过观察镇上的居民 Bob 的外出活动来推测天气状态。Bob 的活动有两种：散步 (W) 和购物 (M)。

天气状态的转移概率是：

$P(S | S) = 0.8, P(R | S) = 0.2$   
 $P(S | R) = 0.4, P(R | R) = 0.6$

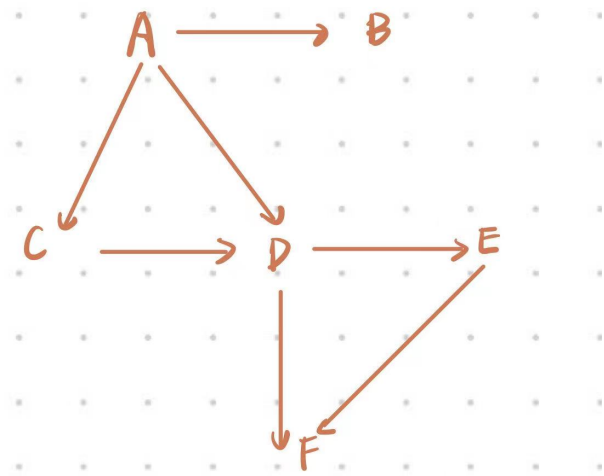
不同天气状态下观测到不同活动的概率是：

$P(W | S) = 0.7, P(M | S) = 0.3$   
 $P(W | R) = 0.4, P(M | R) = 0.6$

- 若 Bob 的活动序列是散步、购物、散步：
- (1) 使用后向算法求这个序列出现的概率
  - (2) 推测最可能的天气序列

题目 6

最近，第二波新型冠状病毒肺炎疫情来势汹汹，根据钟南山等人的模型预测，这波疫情将在六月底达到峰值。小航同学研究了疫情中几个因素之间的相互关系，这些因素包括：A——社会面阳性病例数量迅速增加；B——发热门诊就诊人数迅速增加；C——任课教师感染新冠；D——同班同学感染新冠；E——舍友感染新冠；F——自己感染新冠。他认为这些因素都可以使用 0-1 变量表征，且变量之间的概率相依关系可以使用一张有向图表示。



- (1) 如果 D 已知，判断 A 和 F 是否互相独立；如果 C 已知，判断 A 和 E 是否互相独立
- (2) 已知各个变量间的概率相依关系如下表所示：

-	A=0	A=1
B=1	0.2	0.7

-	A=0	A=1
C=1	0.1	0.6

-	A=0, C=0	A=1, C=0	A=0, C=1	A=1, C=1
D=1	0.03	0.15	0.35	0.87

-	D=0	D=1
E=1	0.16	0.43

-	D=0, E=0	D=1, E=0	D=0, E=1	D=1, E=1
---	----------	----------	----------	----------

-	D=0, E=0	D=1, E=0	D=0, E=1	D=1, E=1
F=1	0.01	0.65	0.49	0.93

求 A=0, B=1, C=1, D=0, E=1, F=1 的概率

(3) 现在已知 D=0, E=1, F=1, 求 A=0 的概率

题目 7

某个条件随机场有定义在边上的特征函数：

$$\begin{aligned}
t_1 &= t_1(y_{i-1} = 2, y_i = 1, x) \quad \lambda_1 = 3.4 \\
t_2 &= t_2(y_1 = 1, y_2 = 2, x) \quad \lambda_2 = 1.7 \\
t_3 &= t_3(y_2 = 2, y_3 = 2) \quad \lambda_3 = 0.45
\end{aligned}$$

有定义在节点上的特征函数

$$\begin{aligned}
s_1 &= s_1(y_i = 2) \quad \mu_1 = 1 \\
s_2 &= s_2(y_1 = 1) \quad \mu_2 = 0.875 \\
s_3 &= s_3(y_3 = 1) \quad \mu_3 = 0.025
\end{aligned}$$

现在给定某一观测序列  $x$ , 计算标记序列为{1,2,2}和{2,1,2}的未归一化的概率

题目 9

(1) 已知一个图像：

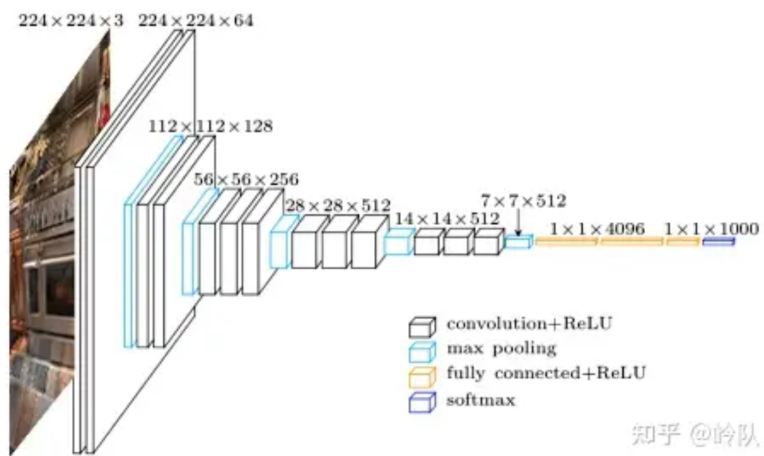
0	3	2	9	8
0	3	4	0	0
0	0	1	0	0
0	3	4	1	0
0	9	3	1	0

现在使用如下卷积核对图像进行卷积：

1	0	-1
1	0	-1
1	0	-1

要求  $stride = 1, padding = 1$ 。给出卷积后的特征图矩阵，并使用文字表述这个卷积核起到的作用。

(2) 计算下面这个网络的参数数目（只考虑权重，不考虑偏置）



(3) 列举防止过拟合的方法

(4) 绘制 LeNet、AlexNet、GoogleNet、ResNet 的基本结构

## 题目 10

一个循环神经网络用于预测股票价格，每次输入的维度是 20 维向量，输出 1 维标量，隐状态是 512 维。

(1) 画出这个网络的基本结构

(2) 计算网络参数的数量（只考虑权重，不考虑偏置）