关注微信公众号（爱数模的小驴）或加资料群392406441，群文件下载2023年美赛免费的答疑解析资料

2023年美赛C题第三部分框架.

1. 按照难度分类

① 量化各单词的难度

② 对附件的所有单词的难度系数进行计算.

③ 建立K均值聚类模型

2. 对于分类结果中的单词属性进行数据统计

① 统计对应的占比

3. 判断EERIE的难度如何.

① 根据问题二求解得到单词难度系数.

② 通过求解知该数组属于具体的类别.

4. 分类模型的准确性

① 使用类中心对各单词进行识别.即

② 判断识别结果准确性.

5. 列出数据集其他特征.

① 可将原始数据的得分统计,大部分大约有最高的分值为各分

② 超75%的人群需要需要5次以上的机次.

6. 书信.

① 总结结论.

② 陈述论文.

[本框无需体现在论文中,仅用于论文布局]

扫描全能王 创建

关注微信公众号（爱数模的小驴）或加资料群 392406441，群文件下载 2023 年美赛免费的答疑解析资料

C题第2问

问题分析

［为了避免重复，问题分析主要以下几点为准 具体听视频讲解］

① 以附件所提供的数据为依据，通过得分占比量化单词的难度系数

② 通过求解附件中所有单词的难度系数，以此建立K均值聚类模型

③ 使用Matlab软件对其求解得到具体的分类结果及类中心，

④ 对每一个类中的单词属性进行数理统计，分别统计各单词属性占比情况

⑤ 根据问题二的求解结果，得到单词"EERIE"的难度系数，通过计算该系数距离上述类中心的距离以此来判断该单词的难度

⑥ 通过对附件所有单词使用欧氏距离模型对其求解，以最小值所属的类为该单词的识别结果，验证分类模型的准确性

⑦ 分别统计得分人数的分布情况以及不同单词对应的众数分布情况．

扫描全能王 创建

关注微信公众号（爱数模的小驴）或加资料群 392406441，群文件下载 2023 年美赛免费的答疑解析资料

模型建立与求解

根据问题分析可知，要想对附件中单词按照难度分类，则首先需要根据每个单词的得分情况明确单词的难度系数，并结合问题二相关结论可知参与该游戏困难模式下人数及总结果数都会对得分结果产生影响，为了客观的描述单词得分的难易程度，

令单词在正常模式下的结果为A，困难模式下结果为B.

简单模式下：单词得分1～7的人数分别为 x_i（$i=1,2\cdots7$）

困难模式下：单词得分1～7的人数分别为：y_i（$i=1,2\cdots7$）

在同等概率下，玩家选择两种不同模式，单词1～7得分[redacted]的难易程度为：

$$P_i=\frac{x_i}{A}+\frac{y_i}{B}\quad(i=1,2\cdots7)$$

其中，P_i表示得分为1～7的难易程度，该数值越小时，难度则越大.

根据附件所提供的数据，不能直接得出上式结果，本文为了更好的量化其数值，可采用以下方式：

扫描全能王 创建

关注微信公众号（爱数模的小驴）或加资料群 477057192，群文件下载 2023 年美赛免费的答疑解析资料

$$P_i = \frac{X_i}{A} + \frac{y_i}{B}$$

$$= \frac{BX_i + Ay_i}{AB}$$

则存在不等式：

$$\frac{AX_i + Ay_i}{AB} > \frac{BX_i + Ay_i}{AB} > \frac{BX_i + By_i}{AB} \quad (\text{其中 } A > B)$$

对上述不等式进行化简可得：

$$\frac{X_i + y_i}{B} > P_i > \frac{X_i + y_i}{A}$$

则 P_i 近似等于区间均值，

即：$$P_i = \frac{1}{2}\left[\frac{X_i + y_i}{B} + \frac{X_i + y_i}{A}\right]$$

将附件中的数据代入上式，分别求得各单词的得分的难易程度，要对单词进行分类，可采用 K 均值聚类模型，具体步骤如下：

[百度插入 K 均值算法的步骤]

根据上述模型求解可知，分别对附件的单词分类结果如下：

④

关注微信公众号（爱数模的小驴）或加资料群 477057192，群文件下载 2023 年美赛免费的答疑解析资料

类序号	类中心	类元素
类1		
类2		
类3		
.....		
.....		

使用数理统计方法，分别对上表中各类元素的单词属性进行占比统计，结果如下：

项目	类单词属性1	单词属性2	单词属性3	.....
类1				
类2				
类3				
.....				

综合描述：（根据表中结果自行描述）

要对单词“EERIE”的难易程度定义，则首先根据问题二所预测的结果代入上述模型，从而求得得分1～7对应的隶值如下表所示：

扫描全能王 创建

关注微信公众号（爱数模的小驴）或加资料群 477057192，群文件下载 2023 年美赛免费的答疑解析资料

项目	1	2	3	4	5	6	7
P_i							

令待检测单词用 LLT 组合难度矩阵为 W 则

$$W=[P_1\ P_2\ P_3\ \cdots P_7]$$

上述聚类所得类中心分别为 H_i 其中 i 表示类序号

$$H_i=[P_{1i}, P_{2i}, P_{3i}, \cdots P_{7i}]$$

则存在 $\min d_i=\|W-H_i\|^2 \quad (i=1,2\cdots n)$

当 d_i 取小值时求对应的 i 值

即单词所所属类别为 i，且其整体难度为 H_i

将上表数据代入上述模型得单词 EERIE 的难度为：

单词	整体难度
EERIE	[ ]

同理对附件所有单词按照上述模型求解从而得到各单词的分类结果，具体如下表所示：

扫描全能王 创建

关注微信公众号（爱数模的小驴）或加资料群477057192，群文件下载2023年美赛免费的答疑解析资料

单词	K均值聚类结果所属类别	该列所属类别	是否准确
合计			

通过上表得到K均值聚类的正确率______.

使用统计学原理分别对不同单词对应的差分值，以了组合整体分布统计，得到

① 超75%的人群基本需要2步以上完成

② ______

【开放性答题，自行发挥】

扫描全能王 创建

关注微信公众号（爱数模的小驴）或加资料群 392406441，群文件下载 2023 年美赛免费的答疑解析资料

书信框架

① 第一段 综合描述本文主要研究的问题

② 第二段 有接说研究各题的结果

1. 注意只表述主要问题

2. 注意适当的进行问题成果顺序调整

3. 强调结果的重要性及有效性

4. 对于预测与聚类结果强调准确性

扫描全能王 创建