

2022 年第二届长三角高校数学建模竞赛

题 目 长三角高教现状分析与发展预测的统计学应用

摘 要:

长江三角洲地区位于江海交汇之地,沿江、沿海的城市众多,是中国经济发展最活跃、开放程度最高、创新能力最强的区域之一。处于长江三角洲地区的城市不仅经济实力雄厚,而且高等教育的格局和优质高等教育的实力全国领先。应用客观赋权法对长三角高等教育资源的统计信息进行数据量化,并做出分析及预测,将对长江三角洲地区的高等教育的发展及规划及有一定的参考意义。

对于问题一,根据互联网中各个权威平台发布的高校排名,筛选出前 50 所长三角地区的高校。在对互联网中权威的统计数据进行归一化处理后,运用熵权法和主成分分析法来处理数据对各个高校的各项评分进行数据分析,通过 Matlab、Python 软件得出综合评价的排名最靠前的 20 所长三角地区的高校。结果见表 1。

对于问题二,根据浙江高考 2018 年-2021 年的一分一段表,运用 Excel 软件绘出折线图来估测小李弟弟的分数段以及位次范围;再根据小李同学的爱好估测小李同学的选科及志愿专业;根据 2018 年-2021 年高校的招生计划数浮动情况、录取情况及小李的志愿专业建立预测模型来预测 2022 年的各专业投档位次;使用估测出的位次范围、选科、爱好、以及问题一中的 20 所高校,运用 TOPSIS 分析方法及线性规划来制定 80 个志愿的方案,再分析其合理性。方案见表 2。

对于问题三,将高等教育的发展分为地区的经济实力的强弱、国家或地区的政策影响、历史文化底蕴、地理与人口环境、教学与科研仪器设备资产价值、师资及生源等几个方面,并运用主成分分析法来量化考量。运用灰色关联度模型及 SPSS 量化分析上述各个因素与长三角高等教育的相关性的强弱。最后对长三角的高等教育发展现状及高等教育的相关因素进行阐述。

对于问题四,根据问题三得出的长三角的高等教育发展现状量化数据及各个因素,收集近 20 年的长三角高等教育资源数据,并建立 Elman 神经网络模型。通过此神经网络模型,得出长三角高等教育的未来一段时间的发展状况,最后为其将来的发展和规划提出建议。

关键词: 熵权法 主成分分析法 TOPSIS 分析 灰色关联度模型 神经网络模型

目 录

一、 问题重述	1
1.1 问题背景	1
1.2 目标任务	1
二、 模型假设及符号说明.....	2
2.1 模型假设	2
2.2 主要符号说明	2
三、 问题分析	3
3.1 针对问题一	3
3.2 针对问题二	3
3.3 针对问题三	3
3.4 针对问题四	3
四、 问题一模型的建立与求解.....	4
4.1 熵权法应用于各高校综合评价	4
4.2 长三角 20 所高校综合实力排名表	5
五、 问题二模型的建立与求解.....	5
5.1 TOPSIS 法及线性规划应用于志愿筛选与填报	5
5.2 小李弟弟的 80 个志愿方案.....	7
5.3 志愿填报方案的合理性分析.....	10
六、 问题三模型的建立与求解.....	10
6.1 主成分分析法应用于权重与指标处理	10
6.2 灰色关联度模型应用于分析高等教育相关因素	11
6.3 长三角高等教育与各因素的灰色关联系数及结论	12
七、 问题四模型的建立与求解.....	14
7.1 Elman 神经网络模型应用于长三角高教发展预测	14
7.2 对长三角高教未来发展的预测	14
7.3 对长三角高教未来规划的建议	14
八、 模型的评价与改进.....	14
8.1 模型的优点	14
8.2 模型的改进	15
九、 参考文献	15
十、 附录	15

一、 问题重述

1.1 问题背景

长江三角洲地区位于长江下游，处于江海交汇之地，在国家现代化建设大局和全方位开放格局中具有举足轻重的战略地位，是中国经济发展最活跃、开放、创新能力最强的区域之一。在长江三角洲的 41 个城市中，拥有全国 17.47% 的本科院校，占全国 25.55% 的“双一流”建设大学和 20.81% 的“双高计划”学校，前五百名高校占全国的 35.29%。可见，处于长江三角洲地区的城市不仅经济实力雄厚，而且高等教育的格局和优质高等教育的实力全国领先。由此，对长三角高等教育资源量化分析及预测，将对其高等教育的发展有一定的价值。

1.2 目标任务

问题一：筛选出权威平台的高校排名中前 50 所长三角地区的大学。运用归一法归纳和整合一系列的统计数据，对各个高校的各项评分进行数据分析，最终通过 Matlab、Excel 排序得出综合评价排名最靠前的 20 所长三角地区的高校。

问题二：通过小李同学的爱好来估测小李同学的选科；再根据运用 Excel 绘出的浙江 2018-2021 年的一分一段表的图表来估测小李的弟弟的分数段以及位次范围。建立关于各相关高校的招生计划数和自 2021 年起前 3 年的相关高校录取情况（位次、实际招生数等）的预测模型，并运用 TOPSIS 分析方法预测 2022 年的相关专业投档位次。通过估测出的位次范围、选科、爱好、以及**问题一**中的 20 所高校等信息进行线性规划，来制定志愿填报的方案。

问题三：分出地区经济实力、国家或地区的政策影响、历史文化底蕴、地理与人口环境、教学与科研仪器设备资产价值、师资及生源等几个方面来量化考量。运用灰色关联度模型来分析上述各个因素与长三角高等教育的相关性的强弱。最后整合并量化分析出长三角的高等教育发展现状，得出各因素与高等教育的发展关联性的强弱。

问题四：收集长三角地区 2000 年-2021 年高等教育资源的数据，并建立关于长三角的高等教育发展现状量化数据及各个因素与长三角高等教育的相关性（问题三的权重与指标）的 Elman 神经网络模型。再根据此神经网络模型，通过差分得出未来一段时间的发展状况。最后给出建议。

二、 模型假设及符号说明

2.1 模型假设

为了便于考虑问题，我们在不影响模型准确性的前提下，作出以下假设：

- (1) 假设小李的弟弟的成绩稳定，位次在全省前 10%上下；
- (2) 假设 2022 年的招生计划相对于去年变化不大。
- (3) 假设小李的弟弟爱好广泛多元，可选专业不限。
- (4) 假设小李的弟弟所填的志愿均符合各个专业的选科限制。
- (5) 假设小李的弟弟的条件均符合所填志愿高校的招生简章的要求。

2.2 主要符号说明

注：此为本文的主要符号说明，其它符号解释详见正文部分。

符号	意义
w_j	权重
x_{xj}	高校数据
E_j	熵值
d_j	差异系数
s_j	高校综合评分
z_j	高校综合指数值
t	小李弟弟高考位次
b_{ji}	大学位次
C_j	是否选择高校
F_p	主成分
$Y_{(k)}$	社会指标
$E_{(w)}$	教育资源指标

三、 问题分析

3.1 针对问题一

问题一的目的在于通过分析得出实力最强的 20 所长三角地区的高等学校。首先，根据互联网中各个权威平台发布的高校排名等数据，筛选出长三角地区的前五十所高校。其次，再通过信息归纳、整合得出进一步的数据。最后，通过 Matlab、Excel 软件，应用熵权法对各个高校的各项评分进行数据分析，在排序后最终得出综合评价的排名最靠前的 20 所长三角地区的高校。

3.2 针对问题二

本题的目的是帮小李弟弟填报 80 个志愿。首先，根据浙江高考 2018 年-2021 年的一分一段表，运用 Excel 软件绘出折线图来估测小李弟弟的分数段以及位次范围。其次，根据小李同学的爱好（运动、网游、动漫、科幻、智能汽车、无人机等）估测小李同学的选科及志愿专业。然后，根据 2018 年-2021 年高校的招生计划数浮动情况、录取情况及小李的志愿专业建立预测模型来预测 2022 年的各专业投档位次。最后使用估测出的位次范围、选科、爱好、以及问题一中的 20 所高校，运用 TOPSIS 分析方法及线性规划来制定 80 个志愿的方案^[3]。

3.3 针对问题三

本题的要求是量化分析长三角的高等教育发展现状，并研究高等教育的发展与哪些因素有关联性。首先，将高等教育的发展分为地区经济实力、国家或地区的政策影响、历史文化底蕴、地理与人口环境等几个方面、教学与科研仪器设备资产价值、师资及生源来运用主成分分析法量化考量。其次，运用灰色关联度模型来分析上述各个因素与长三角高等教育的相关性的强弱。最后，整合信息并得出结论。

3.4 针对问题四

问题四要求我们预测长三角高等教育未来一段时间的发展情况，并为其将来的发展和规划提出建议。首先，我们根据问题 3 得出的长三角的高等教育发展现状量化数据及各个因素与长三角高等教育的相关性建立时间序列模型。其次，根据此时间序列模型，通过差分得出未来一段时间的发展状况。最后，调整模型，提出建议。

四、 问题一模型的建立与求解

4.1 熵权法应用于各高校综合评价

权重是指某一因素或指标对事物的重要程度,强调该因素在整个整体中的贡献度或者重要性,而权重、指标等的选取方法的主观性仍然较强,它们直接影响到评价体系的公正性以及预测的准确性,因此我们运用主成分分析法和熵权法。

假设有 n 条样本, m 个维度,表示每个随机变量的取值:

$$x_{xj}, i = 1, \dots, n, j = 1, \dots, m$$

网络中收集到的高校的经济、师资、生源、学科、就业等数据存在较大差异,为了不影响其量纲,需要对指标进行归一化,使新序列均落在 $[0,1]$ 之间,消除量级以及量纲:

$$x_{xj} = \frac{\max(\sum x_j) - x_{ij}}{\max(\sum x_j) - \min(\sum x_j)} \quad (2.1)$$

$$x_{xj} = \frac{x_{ij} - \min(\sum x_j)}{\max(\sum x_j) - \min(\sum x_j)} \quad (2.2)$$

计算每个维度的信息熵:

$$E_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij})$$

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, i = 1, \dots, n, j = 1, \dots, m$$

$$k = \frac{1}{\ln(n)} > 0, E_j \geq 0$$

计算差异系数:

$$d_j = 1 - E_j$$

计算权重:

$$w_j = \frac{d_j}{\sum_j d_j}$$

计算综合评分:

$$s_j = \sum_j w_j x_{ij}$$

问题一代码见附录

4.2 长三角 20 所高校综合实力排名表

排名序号	高校名称	综合得分	学校所在地
1	复旦大学	1.916	上海
2	浙江大学	1.874	杭州
3	南京大学	1.704	南京
4	上海交通大学	1.645	上海
5	中国科学技术大学	1.634	合肥
6	同济大学	1.615	上海
7	东南大学	1.584	南京
8	华东师范大学	1.547	上海
9	华东理工大学	1.524	上海
10	宁波大学	1.518	宁波
11	南京理工大学	1.504	南京
12	南京航空航天大学	1.493	南京
13	江南大学	1.486	无锡
14	温州医科大学	1.486	温州
15	上海中医药大学	1.482	上海
16	安徽大学	1.48	合肥
17	中国药科大学	1.475	南京
18	杭州电子科技大学	1.473	杭州
19	温州大学	1.47	温州
20	华东政法大学	1.468	上海

表 1 – 20 所高校综合得分

五、 问题二模型的建立与求解

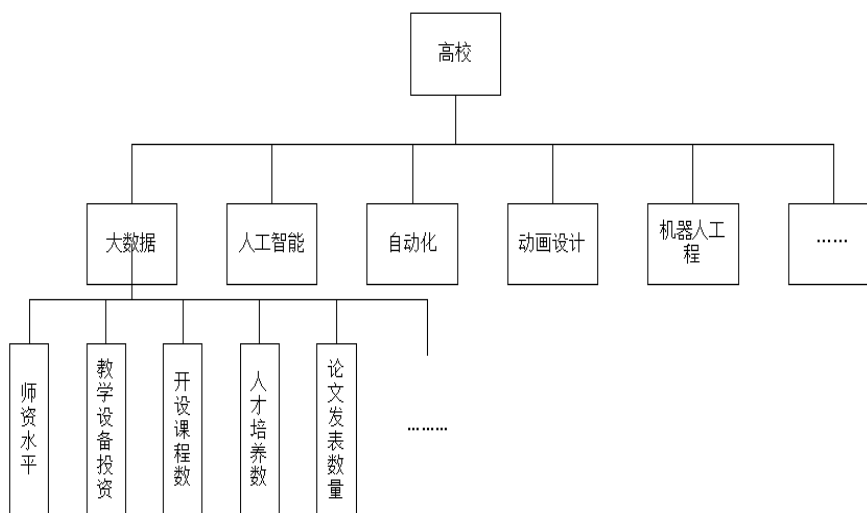
5.1 TOPSIS 法及线性规划应用于志愿筛选与填报

根据浙江高考 2018 年-2021 年的一分一段表,运用 Excel 软件绘出折线图,将分数段与位次段一一对应,估测小李弟弟的位次范围为 17500-42500.

小李弟弟爱好广泛,喜欢运动、网游、动漫、科幻、智能汽车、无人机等。网游、动漫、科幻、智能汽车、无人机等与物理、技术学科相关。由此,我们推测其选科为思想政治、物理、技术。并推测出其志愿专业与互联网、大数据、人工智能、自动化、计算机、数学、动画、电子技术、机器人工程等相关。

根据 2018 年-2021 年高校的招生计划数浮动情况、录取情况及小李的志愿专业建立预测模型来预测 2022 年的各专业投档位次；使用估测出的位次范围、选科、爱好、以及问题一中的 20 所高校，运用 TOPSIS 分析方法来制定 80 个志愿的方案。

我们首先把各高校的师资、教学设施投资、课程数、人才培养数、毕业论文数等量化为三级指标，再运用问题一的模型，对各个指标进行权重配比，最终加权求和得出各高校综合评分。



1. 把各专业质量的评定标准作为三级指标量化为 x_{3i} ，并得到权重 w_{3i} 。

同时，我们把各个高校的各个专业作为二级指标。

接着，对三级指标进行加权求和，进一步得到二级指标 $x_{2j} = w_{3i}x_{3i}$ 及权重为 w_{2j} 。

最后，我们得出高校综合指数值 z_j ：

$$z_j = \sum w_{2j}x_{2j} \text{ (高校综合指数值)}$$

2. 查询高校专业 2021 年位次为 b_{ji} (第 j 个学校 第 i 个专业)

并设小李弟弟的高考成绩位次为 t ，再通过查询资料对 2022 年各高校专业进行统计，并建立以下优化模型：

$$\max z = C_j z_j \text{ (} C_j \text{ 表示 } 0 / 1 \text{ , 即高校是否入选)}$$

$$s. t. \begin{cases} C_j b_{ji} \geq t \\ \sum C_j = 80 \\ C_j = 0 \text{ or } 1 \end{cases}$$

5.2 小李弟弟的 80 个志愿方案

志愿排序	院校名称	专业名称	省份	所在地	预测位次
1	江南大学	微电子科学与工程	江苏	无锡	17436
2	南京邮电大学	(新增)人工智能	江苏	南京	17486
3	江南大学	自动化	江苏	无锡	17591
4	中国矿业大学	计算机类	江苏	徐州	17648
5	合肥工业大学	信息与计算科学	安徽	合肥	17678
6	上海理工大学	工科试验班	上海	上海	17713
7	安徽大学	电子信息类	安徽	合肥	17720
8	合肥工业大学	机器人工程	安徽	合肥	17731
9	南京师范大学	(新增)能源动力类	江苏	南京	17756
10	安徽大学	数学类	安徽	合肥	17761
11	南京邮电大学	网络工程	江苏	南京	17849
12	上海大学	信息工程	上海	上海	17964
13	南京农业大学	环境科学与工程类	江苏	南京	20291
14	南京财	人力资源管理	江苏	南京	20370
15	南京航空航天大学	交通运输	江苏	南京	20373
16	江苏大学	信息安全	江苏	镇江	20652
17	上海对外经贸大学	物流管理	上海	上海	20754
18	合肥工业大学	能源与动力工程	安徽	合肥	20802
19	合肥工业大学	国际经济与贸易	安徽	合肥	20852
20	中国矿业大学	(新增)遥感科学与技术	江苏	徐州	20869
21	合肥工业大学	交通工程	安徽	合肥	20907
22	宁波大学	创新班(理工)	浙江	宁波	20950
23	南京农业大学	应用化学	江苏	南京	20951
24	上海师范大学	人力资源管理	上海	上海	21589
25	安徽大学	互联网金融	安徽	合肥	22076
26	中国矿业大学	安全科学与工程类	江苏	徐州	22130
27	宁波大学	通信工程	浙江	宁波	22420

28	南京审计大学	计算机科学与技术	江苏	南京	22507
29	杭州电子科技大学	网络空间安全	浙江	杭州	22610
30	南京工业大学	自动化	江苏	南京	22623
31	南京信息工程大学	电子信息类	江苏	南京	22680
32	南京农业大学	机械类	江苏	南京	22734
33	南京农业大学	交通运输	江苏	南京	22869
34	宁波大学	(新增)工程力学	浙江	宁波	22905
35	江苏大学	(新增)数据计算及应用	江苏	镇江	22907
36	杭州师范大学	数学与应用数学	浙江	杭州	23014
37	南京工业大学	电子信息类	江苏	南京	23177
38	合肥工业大学	食品科学与工程类	安徽	合肥	23212
39	杭州师范大学	口腔医学	浙江	杭州	23846
40	浙江理工大学	启新卓越实验班	浙江	杭州	24114
41	上海理工大学	电气工程及其自动化	上海	上海	24866
42	浙江工业大学	计算机类	浙江	杭州	25028
43	上海电力大学	自动化	上海	上海	26053
44	南京信息工程大学	地理空间信息工程	江苏	南京	26352
45	浙江工业大学	工业设计	浙江	杭州	26494
46	上海第二工业大学	数字媒体技术	上海	上海	27253
47	南京工程学院	电气工程及其自动化	江苏	南京	27643
48	上海海事大学	交通运输	上海	上海	27928
49	南京信息工程大学	光电信息科学与工程	江苏	南京	28131
50	杭州电子科技大学	电子信息类	浙江	杭州	28201
51	浙江工业大学	机器人工程	浙江	杭州	28278
52	南京工业大学	能源动力类	江苏	南京	28359
53	南京信息工程大学	机械电子工程	江苏	南京	28367
54	浙江工业大学	电子信息类	浙江	杭州	28443
55	浙江师范大学	法学类	浙江	金华	28615
56	上海海事大学	电子信息类	上海	上海	28908

57	上海海事大学	交通工程	上海	上海	29051
58	南京信息工程大学	信息管理与信息系统	江苏	南京	29345
59	南京工程学院	电气工程及其自动化	江苏	南京	29770
60	杭州电子科技大学	电子信息类(通信学院)	浙江	杭州	29862
61	浙江工业大学	电气类	浙江	杭州	29877
62	上海电力大学	计算机科学与技术	上海	上海	29899
63	江苏大学	微电子科学与工程	江苏	镇江	30395
64	浙江师范大学	理科试验班类	浙江	金华	30400
65	上海电力大学	电子信息工程	上海	上海	31063
66	杭州电子科技大学	计算机科学与技术	浙江	杭州	32146
67	南京工程学院	电气工程及其自动化	江苏	南京	32296
68	扬州大学	微电子科学与工程	江苏	扬州	32481
69	上海电力大学	集成电路设计	上海	上海	32570
70	浙江工业大学	安全工程	浙江	杭州	32698
71	南京信息工程大学	应用物理学	江苏	南京	32807
72	中国美术学院	工业设计	浙江	杭州	32889
73	南京信息工程大学	软件工程	江苏	南京	33308
74	江苏大学	测控技术与仪器	江苏	镇江	33383
75	上海工程技术大学	信息管理与信息系统	上海	上海	37062
76	苏州科技大学	地理信息科学	江苏	苏州	37183
77	浙江工业大学	软件工程	浙江	杭州	41326
78	浙江工商大学	计算机类	浙江	杭州	41359
79	杭州电子科技大学	自动化(中外合作办学)	浙江	杭州	41435
80	浙江工商大学	物联网工程	浙江	杭州	41538

表 2 - 小李弟弟的 80 个志愿方案

5.3 志愿填报方案的合理性分析

1. 志愿填报按照“冲、稳、保”方法填报。“冲”20个志愿，“稳”40个志愿，“保”20个志愿。位次范围17500-42500与小李弟弟成绩相符。
2. 小李弟弟爱好智能汽车、无人机等，与机器人工程、人工智能等专业相关。我们为其填报的志愿多数与此相关。
3. 填报高校所在城市多为经济发达城市，毕业会将有良好的就业环境。

六、 问题三模型的建立与求解

6.1 主成分分析法应用于权重与指标处理

首先，将高等教育的发展分为地区经济实力、国家或地区的政策影响、师资、教学设施投资、毕业论文数、历史文化底蕴、教学与科研仪器设备资产价值等几个方面，使用主成分分析法^[2]来分析各权重与指标，并量化考量。

(1) 下图为长三角地区各省教育指标数据：

指标分类	指标名称	2013-2018/2019年期间				
		安徽	上海	江苏	浙江	全国
高校资源	本科学校比重	37.66%	57.62%	46.70%	54.98%	47.08%
	每十万人高等学校平均在校生数	2279.7100	3442.3621	3009.2172	2393.6955	2576.0723
在校生培养层次	在校研究生比重	4.82%	24.24%	9.50%	6.99%	7.62%
	在校博士生比重	0.59%	4.60%	1.48%	1.09%	1.19%
教师资源	副高及以上专任教师比重	35.94%	50.32%	47.08%	44.93%	42.05%
	正高级专任教师比重	8.69%	18.25%	13.38%	14.25%	12.51%
	硕士及以上学位专任教师比重	58.90%	80.69%	64.20%	64.20%	58.90%
	博士学位专任教师比重	15.89%	49.89%	30.99%	30.19%	22.72%
高校资产	图书数量(万册)	9140.6800	7403.9508	17550.1056	10786.0024	11913.35293
	教学、科研仪器设备资产价值(万元)	1444121.8100	2714949.0028	3826922.7965	2252830.0236	2931567.274
	教室数量(间)	2682087.6200	1448327.3148	4184056.6872	2333416.2294	2655266.744
教育经费	教育经费支出(万元)	2682088.0000	5444638.6400	6892966.1600	4532728.7200	5623444.507
	生均教育事业费支出(元/人)	13105.5800	31977.6152	18347.8120	18347.8120	18872.0352
	生均公用经费支出(元/人)	6952.6300	18772.1010	8134.5771	7925.9982	8343.156
科学研究与实验发展	R&D人员数量(人)	27080.5000	44412.0200	61472.7350	46849.2650	50911.34
	R&D经费内部支出数量(万元)	302582.8300	928929.2881	995497.5107	571881.5487	832102.7825
	发表科技论文数量(篇)	38317.8300	79701.0864	116103.0249	47514.1092	81106.0735
	出版科技著作数量(种)	1136.1700	2692.7229	3010.8505	1806.5103	2503.361233
	专利申请数量(件)	9044.1700	10943.4457	33825.1958	16460.3894	20409.67697

表 3 – 各省教育指标数据

(2) 我们运用主成分分析法求解

首先运用 SPSS 软件将指标数据标准化，进行指标之间的相关性判断，然后确定七个对教育发展影响最大的指标作为主成分，表达式如下：

$$F_p = a_{1i} * Z_{X1} + a_{2i} * Z_{X2} + \dots + a_{pi} * Z_{Xp}$$

并计算出各个指标的权重：

四个省份的影响因素	得出的权重
本科学校比重	0.34
在校研究生比重	0.38
在校博士生比重	0.41
正高级专任教师比重	0.42
图书数量（万册）	0.35
教育经费支出（万元）	0.38
发表科技论文数量（篇）	0.44

表 4 – 各影响因素及其权重

加权求和后得到长三角地区四个省的教育质量综合得分来分析教育发展现状：

省份	综合得分
安徽	63.44
上海	86.35
江苏	75.83
浙江	73.56

表 5 – 各地区综合得分

6.2 灰色关联度模型应用于分析高等教育相关因素

地区的经济实力的强弱、国家或地区的政策影响、历史文化底蕴、地理与人口环境^[5]、教学与科研仪器设备资产价值、师资及生源^[6]等几个方面均对高等教育的发展有一定的影响。运用灰色关联度模型及 SPSS 量化分析上述各个因素^[1]与长三角高等教育的相关性的强弱。

$$Z_{nm} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix}$$

以横轴为指标，纵轴为样本编号的矩阵经每个指标规范化后为虚拟母样本为 Y，其每个元素取对应指标的最优值

$$Y = [y_1, y_2, \cdots, y_m]$$

关联系数

$$y_j = \max(z_{1j}, z_{2j}, \dots, z_{nj})$$

记差值矩阵为 K

$$K = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1m} \\ k_{21} & k_{22} & \cdots & k_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nm} \end{bmatrix} = \begin{bmatrix} |z_{11} - y_1| & |z_{12} - y_2| & \cdots & |z_{1m} - y_m| \\ |z_{21} - y_1| & |z_{22} - y_2| & \cdots & |z_{2m} - y_m| \\ \vdots & \vdots & \ddots & \vdots \\ |z_{n1} - y_1| & |z_{n2} - y_2| & \cdots & |z_{nm} - y_m| \end{bmatrix}$$

$$a = \min_i \min_j |k_{ij}|, \quad b = \max_i \max_k |k_{ij}|$$

计算指标关联度

取 $\rho = 0.5$ ，记得到第 i 个样本的灰色关联度，即评价结果为 r_i 为

$$r_i = \frac{1}{m} \sum_{j=1}^m \xi_{ij} \quad \text{或} \quad r_i = \sum_{j=1}^m w_j \xi_{ij}$$

w_j 为指标的权重，以此作为各个观测值的评价结果。

6.3 长三角高等教育与各因素的灰色关联系数及结论

长三角地区GDP与教育资源的灰色关联系数（前10个地区）

排名序号	灰色关联系数	地区	评价
1	0.812	上海	符合
2	0.76	南京	符合
3	0.72	苏州	符合
4	0.685	杭州	较为符合
5	0.674	合肥	较为符合
6	0.643	宁波	较为符合
7	0.625	无锡	较为符合
8	0.614	温州	较为符合
9	0.585	盐城	较为符合
10	0.546	嘉兴	较为符合

表 6 - 长三角 GDP 与教育资源的灰色关联系数

长三角地区人口数量与教育资源的灰色关联系数（前10个地区）

排名序号	灰色关联系数	地区	评价
1	0.715	上海	符合
2	0.708	南京	符合
3	0.685	苏州	符合
4	0.671	杭州	符合
5	0.657	合肥	较为符合
6	0.634	无锡	较为符合
7	0.622	宁波	较为符合
8	0.583	温州	较为符合
9	0.542	盐城	较为符合
10	0.518	南通	较为符合

表 7 - 长三角人口数量与教育资源的灰色关联系数

影响因素	$y(x_0(k), x_1(k))$ 安徽省	$y(x_0(k), x_2(k))$ 江苏省	$y(x_0(k), x_3(k))$ 上海	$y(x_0(k), x_4(k))$ 浙江省
经济发展水平	0.4751	0.6586	0.8922	0.6725
历史文化因素	0.4299	0.5733	0.7680	0.5845
人口数量	0.6356	0.5462	0.5766	0.6474
人均GDP	0.7520	0.8985	0.7753	0.6685
科技发展水平	0.4224	0.6657	1.0000	0.7468
人才引进政策	0.3356	0.4035	0.5317	0.5845

表 8 - 各影响因素关联度

1. 从评分结果来看，高等教育综合资源分配大致为上海市最优，江苏省第二位，浙江省第三位，安徽省第四位。
2. 高等教育的发展与地区经济水平、国家及地区政策影响、师资分配等因素均有关。
3. 对高等教育发展具有较大影响的因素有：地区经济实力、国家及地区财政资源分配等。
4. 对高等教育发展具有相对较小影响的因素有：自然环境、历史文化底蕴等。

七、 问题四模型的建立与求解

7.1 Elman 神经网络模型应用于长三角高教发展预测

首先收集长三角各地区 2000 到 2021 年高等教育资源数据（以第三问指标作为依据），并由第三题模型得到各个指标（社会指标）对教育资源的影响（通过相关度量化），然后将社会指标最为输入 $y(k)$ ，通过 Elman 神经网络模型对未来五年社会指标进行预测，输出教育资源指标为 $E(w)$ 。

$$\begin{aligned}y(k) &= g(w^3 x(k)) \\x(k) &= f(w^1 x_c(k) + w^2(u(k-1))) \\x_c(k) &= x(k-1)\end{aligned}$$

u 为 r 维以经济发展水平、人口数量、人均 GDP 等社会指标组成的输入向量； x 为 n 维中间层结点单元向量； y 为 m 维输出结点向量；为 n 维反馈状态向量，为中间层到输出层连接全取；为输入层到中间层连接权值；为承接层到中间层到连接权值； $g(*)$ 为输出神经元的传递函数，是中间层输出的线性组合； $f(*)$ 为中间层神经元的传递函数，采用 S 函数。

最后采用误差平方和函数对结果进行权值修正：

$$E(w) = \sum (y_k(w) - \widetilde{y}_k(w))^2$$

7.2 对长三角高教未来发展的预测

1. 长三角地区高等教育发展总体趋势是向上的。
2. 将来高等教育资源的分配在长三角地区会有一定的协调，如安徽省、浙江省的高等教育资源日渐增加等。
3. 长三角高教将提升各大城市整体高教质量，共享优质资源，并进一步优化高教格局。

7.3 对长三角高教未来规划的建议

1. 注重区域协调发展^[4]，适当地将一些资源向高等教育资源相对较少的地区分配，力求共同发展。
2. 建立集群发展体制机制，各高教团体协作发展。
3. 注重各区域互帮互联，强强联合、强弱区域发展一体化。

八、 模型的评价与改进

8.1 模型的优点

1. 建立的模型预测的是位次而非分数线，考虑到了同分不同位次的情况，更精准。

2. 该模型的采用权威数据，兼容性高，可得出较为完善的结果。
3. 采用 Matlab、Excel、SPSS 等统计软件来统计，使得结果较为准确。
4. 建立的模型涵盖的数据丰富，时间跨度大，所得结果更为准确，误差更小。

8.2 模型的改进

1. 根据招生计划等政策的更新，可微调所填报的 80 个志愿的数据。
2. 根据地区发展不平衡等因素，可进一步搜索文献，调整各个因素的权重。
3. 可进一步优化我们的模型，尽可能在预测的时间轴上作更精确的分析。

九、 参考文献

- [1] 王爽 赵鹏， 基于 TOPSIS 的出行方案评价指标权重的确定方法[J]. 交通运输系统工程与信息, 2009, 9 (2): 122-128.
- [2] 张浩 冯林， 主成分分析法在高校科技创新能力评价中的应用 [J]. 武汉理工大学学报 (信息与管理工程版), 2004, 26(6): 157-161.
- [3] 鲍学英 李海连 王起才， 基于灰色关联分析和主成分分析组合权重的确定方法研究[J]. 小数学的实践与认识, 2016, 46 (9): 129-134.
- [4] 段从宇， 资源视角的高等教育区域协调发展研究 大连理工大学, 2015.
- [5] 刘艳春 孙亮， 人口因素与区域经济的多重协整关系研究——基于沈阳市时间序列数据的分析[J]. 数理统计与管理, 2016, 35(2): 309-318.
- [6] 郑延福， 本科高校教师教学质量评价研究 中国矿业大学, 2012.

十、 附录

所用软件：Python、Matlab

问题一：

```
def cal_entropy_weight(x, label = dict()):
    """熵值法计算变量的权重"""
    print('原始数据: '+ str(x))
    # 获得表的列头
    x_header = x.keys()
    # 初始化默认的 label: 1 为正向指标 2 为负向指标
    if len(label) == 0:
```

```

        for key in x_header:
            label[key] = 1
    print('label:'+str(label))
    # 标准化
    for key in label:
        # 得到列最大和最小
        maxium = np.max(x[key], axis=0)
        minium = np.min(x[key], axis=0)
        # print('maxium:' + str(maxium))
        # print('minium:' + str(minium))
        if label[key] == 1:# 正向指标
            x[key] = x.loc[:, key].apply(lambda x: ((x - minium) / (maxium - minium)))
            # x[key].apply(lambda x: ((x - maxium) / (np.max(x) - np.min(x))))
        elif label[key] == 2: # 负向指标
            x[key] = x.loc[:, key].apply(lambda x: ((maxium - x) / (maxium - minium)))
        else:
            print('输入的数据标签有误! ')
            return
    # x_standardized.append(pd.DataFrame({key:temp_x}), ignore_index=True)

    # print('x_standardized:' + str(x))
    # 真向标准化
    # x = x.apply(lambda x: ((x - np.min(x)) / (np.max(x) - np.min(x))))
    # 负向标准化
    # x = x.apply(lambda x: ((np.max(x) - x) / (np.max(x) - np.min(x))))
    # 获得标准化的表中的值
    x_standardized_data = list(x.to_dict().values())
    print('标准化: ' + str(x))

    # 求 k
    rows = x.index.size # 行
    cols = x.columns.size # 列
    k = 1.0 / math.log(rows)
    lnf = [[None] * cols for i in range(rows)]
    # 矩阵计算--
    # 信息熵
    # p=array(p)
    x = array(x)
    lnf = [[None] * cols for i in range(rows)]
    lnf = array(lnf)
    # p_arr = array()

```

```

temp_p_list = list()
p_row_list = list()
for i in range(0, rows):
    for j in range(0, cols):
        if x[i][j] == 0:
            lnfi_j = 0.0
            p_row_list.append(0.0000)
        else:
            p = x[i][j] / x.sum(axis=0)[j]
            # print(type(p))
            p_row_list.append(p)
            lnfi_j = math.log(p) * p * (-k)
        lnfi[i][j] = lnfi_j
    temp_p_list.append(copy.deepcopy(p_row_list))
p_row_list

```

问题三:

主成分分析代码:

```

import numpy as np
import pandas as pd
from sklearn.decomposition import PCA
from sklearn import preprocessing

# 数据导入
csv_file = "3 (1).csv"
csv_data = pd.read_csv(csv_file, low_memory=False)
csv_df = pd.DataFrame(csv_data)

scaler = preprocessing.MinMaxScaler().fit(csv_df)
X_scaler = pd.DataFrame(scaler.transform(csv_df))

# 主成分分析建模
pca = PCA(n_components=None) # n_components 提取因子数量
# n_components= 'mle', 将自动选取主成分个数 n, 使得满足所要求的方差百分比
# n_components=None, 返回所有主成分
pca.fit(X_scaler)
pca.explained_variance_ # 贡献方差, 即特征根
pca.explained_variance_ratio_ # 方差贡献率
pca.components_ # 成分矩阵
k1_spss = pca.components_ / np.sqrt(pca.explained_variance_.reshape(-1, 1)) # 成分得分系数矩阵

```

```

# 确定权重
# 求指标在不同主成分线性组合中的系数
j = 0
Weights = []
for j in range(len(k1_spss)):
    for i in range(len(pca.explained_variance_)):
        Weights_coefficient = np.sum(100 * (pca.explained_variance_ratio_[i]) * (k1_spss[i][j])) /
np.sum(
        pca.explained_variance_ratio_)
    j = j + 1
    Weights.append(np.float(Weights_coefficient))
print('Weights',Weights)

Weights=pd.DataFrame(Weights)
Weights1 = preprocessing.MinMaxScaler().fit(Weights)
Weights2 = Weights1.transform(Weights)
print('Weights2',Weights2)

import pandas as pd

x = pd.read_excel('data.xlsx')
x = x.iloc[:, :].T
print(x)

# 1、数据均值化处理
x_mean = x.mean(axis=1)
for i in range(x.index.size):
    x.iloc[i, :] = x.iloc[i, :] / x_mean[i]

# 1、数据差值化处理
x = (x - x.min()) / (x.max() - x.min())
x = x.T

# 1、数据初值化处理
x_mean = x.mean(axis=1)
for i in range(x.index.size):
    x.iloc[i, :] = x.iloc[i, :] / x.iloc[i, 0]

# 2、提取参考队列和比较队列
ck = x.iloc[0, :]
print(" 参考队列: ", ck)

```

```

cp = x.iloc[1:, :]
print(" 参考队列: ", cp)

# 比较队列与参考队列相减
t = pd.DataFrame()
for j in range(cp.index.size):
    temp = pd.Series(cp.iloc[j, :] - ck)
    t = t.append(temp, ignore_index=True)

# 求最大差和最小差
mmax = t.abs().max().max()
mmin = t.abs().min().min()
rho = 0.4

# 3、求关联系数
ksi = ((mmin + rho * mmax) / (abs(t) + rho * mmax))

# 4、求关联度
r = ksi.sum(axis=1) / ksi.columns.size

# 5、关联度排序, 得到结果
result = r.sort_values(ascending=False)

print(r)

```

问题四:

matlab 代码:

```

function [test_ty]=shengjingwangluo(data,L,N)
price = data'
构造样本集
数据个数
n=length(price);

% 确保 price 为列向量
price=price(:);

% x(n) 由 x(n-1),x(n-2),...,x(n-L)共 L 个数预测得到.

% price_n: 每列为一个构造完毕的样本, 共 n-L 个样本

```

```

price_n = zeros(L+1, n-L);
for i=1:n-L
    price_n(:,i) = price(i:i+L);
end
划分训练、测试样本
trainx = price_n(1:L-N+1,:);
trainy = price_n(L-N+2:end,:);
[ww,mm]=size(trainx);
testx = price(n-ww+1:n, end);
创建 Elman 神经网络
% 包含 15 个神经元，训练函数为 traingdx
net=elmannet(1:2,15,'traingdx');

% 设置显示级别
net.trainParam.show=1;

% 最大迭代次数为 3000 次
net.trainParam.epochs=3000;

% 误差容限，达到此误差就可以停止训练
net.trainParam.goal=0.000001;

% 最多验证失败次数
net.trainParam.max_fail=5;

% 对网络进行初始化
net=init(net);
网络训练
%训练数据归一化
[trainx1, st1] = mapminmax(trainx);
[trainy1, st2] = mapminmax(trainy);

% 测试数据做与训练数据相同的归一化操作
testx1 = mapminmax('apply',testx,st1);

% 输入训练样本进行训练
[net,per] = train(net,trainx1,trainy1);
测试。输入归一化后的数据，再对实际输出进行反归一化
% 将训练数据输入网络进行测试
train_ty1 = sim(net, trainx1);

```

```

train_ty = mapminmax('reverse', train_ty1, st2);

% 将测试数据输入网络进行测试
test_ty1 = sim(net, testx1);
test_ty = mapminmax('reverse', test_ty1, st2);
显示结果
1.显示训练数据的测试结果
figure(1)
x=1:length(train_ty);

% 显示真实值
plot(x,trainy,'b-');
hold on
% 显示神经网络的输出值
plot(x,train_ty,'r--')

legend('指标真实值','Elman 网络输出值')
title('训练数据的测试结果');

% 显示残差
figure(2)
plot(x, train_ty - trainy)
title('训练数据测试结果的残差')

% 显示均方误差
mse1 = mse(train_ty - trainy);
fprintf(' mse = \n %f\n', mse1)

% 显示相对误差
disp(' 相对误差: ')
fprintf('%f ', (train_ty - trainy)./trainy );
fprintf('\n')
% 显示预测值
disp(' 预测值: ')
fprintf('%f ', test_ty );
fprintf('\n')

2.
function [Y1]=shengjingwangluo2(A,B,testx)
x=A';
y=B';

```

```

[n1,m1]=size(A)
[n2,m2]=size(B)
n=m1
m=m2

p=x;
t=y;
[pn,minp,maxp,tn,mint,maxt]=premnmx(p,t);
u=ones(n,1);
dx=[-1*u,1*u];
%BP 网络训练
net=newff(dx,[n,15,m],{'tansig','tansig','purelin'},'traingdx');
    %建立模型，并用梯度下降法训练.
net.trainParam.show=1000;
net.trainParam.Lr=0.05;
net.trainParam.epochs=5000;
net.trainParam.goal=0.65*10^(-3);
net=train(net,pn,tn);

an=sim(net,pn);
a=postmnmx(an,mint,maxt);

pnew=testx;

pnewn=tramnmx(pnew,minp,maxp);
anewn=sim(net,pnewn);
anew=postmnmx(anewn,mint,maxt)
Y1=anew';

```