

Evaluation model

Yunlong Cheng

2019 年 7 月 17 日

1 主成分分析 (PCA)

1.1 简单实例

由一组数据如下：

	x	y
	2.5	2.4
	0.5	0.7
	2.2	2.9
	1.9	2.2
Data =	3.1	3.0
	2.3	2.7
	2	1.6
	1	1.1
	1.5	1.6
	1.1	0.9

行代表案例，列代表特征，这里有 10 个案例，两个特征。

分析：

1. 第一步：求 x 和 y 的平均值，对于所有案例减去平均值，得：

	x	y
	0.69	0.49
	-1.31	-1.21
	0.39	0.99
	0.09	0.29
DataAdjust =	1.29	1.09
	0.49	0.79
	0.19	-0.31
	-0.81	-0.81
	-0.31	-0.31
	-0.71	-1.01

2. 第二步：求特征协方差矩阵，在这里为

$$cov = \begin{pmatrix} .6165 & .6154 \\ .6154 & .7165 \end{pmatrix}$$

3. 第三步，求协方差的特征值和特征向量：

$$eginvalues = \begin{pmatrix} .04908 \\ 1.28402 \end{pmatrix}$$

$$eginvectors = \begin{pmatrix} -.7351 & -.6678 \\ .6678 & -.7351 \end{pmatrix}$$

4. 第四步，将特征值从大到小排序，选择其中最大的 k 个。在这里我们选择 1.28402，特征向量为 $\begin{pmatrix} -.6678 \\ -.7351 \end{pmatrix}$ 。

5. 第五步，将样本点投影到选取的特征向量上，投影后的数据为：

Transformed Data (Single rigenvector)
x
-0.827970186
1.77758033
-0.992197494
-0.274210416
-1.67580142
-0.912949103
0.991094375
1.14457216
0.438046137
1.22382056

2 理想解方法 (TOPSIS)

问题：

为了客观地评价我国研究生教育的实际状况和各研究生院的教学质量，国务院学位委员会办公室组织过一次研究生院的评估。为了取得经验，先选 5 所研究生院，收集有关数据资料进行了试评估，下表是所给出的部分数据：

j	人均专著 x_1	生师比	科研经费 x_3	逾期毕业率 x_4
i	(本/人)	x_2	(万元/年)	(%)
A	0.1	5	5000	4.7
B	0.2	6	6000	5.6
C	0.4	7	7000	6.7
D	0.9	10	10000	2.3
E	1.2	2	400	1.8

权向量为 $\omega = [0.2, 0.3, 0.4, 0.1]$

解：

1. 指标同向化处理：因为有的指标是越小越好，有的指标是越大越好，同向化消除这些影响。

- 人均专著、科研经费为效应型指标（越大越好）
- 逾期毕业率为成本型指标（越小越好）
- 生师比为区间型指标

	人均专著	生师比	科研经费	逾期毕业率
院校 A	0.1	1.000000	5000	0.212766
院校 B	0.2	1.000000	6000	0.178571
院校 C	0.4	0.833333	7000	0.149254
院校 D	0.9	0.333333	10000	0.434783
院校 E	1.2	0.000000	400	0.555556

2. 构造归一化矩阵：例：院校 A 的人均专属： $\frac{0.1}{\sqrt{0.1^2+0.2^2+0.4^2+0.9^2+1.2^2}} = 0.0637$
得：

	人均专著	生师比	科研经费	逾期毕业率
院校 A	0.063758	0.597022	0.344901	0.275343
院校 B	0.127515	0.597022	0.413882	0.231092
院校 C	0.255031	0.497519	0.482862	0.193151
院校 D	0.573819	0.199007	0.689803	0.562658
院校 E	0.765092	0.000000	0.027592	0.718952

3. 构造加权规范阵: $c_{ij} = w_{ij} \cdot b_{ij}$ 。
4. 确定正理想解 C^* 和负理想解 C^0

	人均专著	生师比	科研经费	逾期毕业率
最优方案	0.765092	0.597022	0.689803	0.718952
最劣方案	0.063758	0	0.027592	0.193151

5. 计算各方案到正理想解与负理想解的距离:

- 到正理想解的距离为: s_i^*
- 到负理想解的距离为: s_i^0

6. 计算各方案的排序指标值 (综合评价指数):

$$f_i^* = \frac{s_i^0}{s_i^0 + s_i^*} \quad i = 1, 2, \dots, m$$

i	人均专著 x_1 (本/人)	生师比 x_2	科研经费 x_3 (万元/年)	逾期毕业率 x_4 (%)	最终得分	正理想解	负理想解	排序
院校A	0.1	5	5000	4.7	0.485830	0.406984	0.384552	4
院校B	0.2	6	6000	5.6	0.526483	0.368177	0.409360	3
院校C	0.4	7	7000	6.7	0.562158	0.315879	0.405565	2
院校D	0.9	10	10000	2.3	0.677571	0.239342	0.502968	1
院校E	1.2	2	400	1.8	0.400512	0.531357	0.354994	5

3 熵权法

问题:

某医院为了提高自身的护理水平, 对拥有的 11 个科室进行了考核, 考核标准包括 9 项整体护理, 并对护理水平较好的科室进行奖励。下表是对各个科室指标考核后的评分结果。

科室	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
A	100	90	100	84	90	100	100	100	100
B	100	100	78.6	100	90	100	100	100	100
C	75	100	85.7	100	90	100	100	100	100
D	100	100	78.6	100	90	100	94.4	100	100
E	100	90	100	100	100	90	100	100	80
F	100	100	100	100	90	100	100	85.7	100
G	100	100	78.6	100	90	100	55.6	100	100
H	87.5	100	85.7	100	100	100	100	100	100
I	100	100	92.9	100	80	100	100	100	100
J	100	90	100	100	100	100	100	100	100
K	100	100	92.9	100	90	100	100	100	100

解：

1. 对原始数据进行归一化： $y_{ij} = \frac{x_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)}$

科室	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
A	1.00	0.00	1.00	0.00	0.50	1.00	1.00	1.00	1.00
B	1.00	1.00	0.00	1.00	0.50	1.00	1.00	1.00	1.00
C	0.00	1.00	0.33	1.00	0.50	1.00	1.00	1.00	1.00
D	1.00	1.00	0.00	1.00	0.50	1.00	0.87	1.00	1.00
E	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00
F	1.00	1.00	1.00	1.00	0.50	1.00	1.00	0.00	1.00
G	1.00	1.00	0.00	1.00	0.50	1.00	0.00	1.00	1.00
H	0.50	1.00	0.33	1.00	1.00	1.00	1.00	1.00	1.00
I	1.00	1.00	0.67	1.00	0.00	1.00	1.00	1.00	1.00
J	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
K	1.00	1.00	0.67	1.00	0.50	1.00	1.00	1.00	1.00

2. 求信息熵：

$$p_{ij} = \frac{t_{ij}}{\sum_{i=1}^n y_{ij}} \quad E_{ij} = - \sum_{i=1}^n p_{ij} \log_n P_{ij}$$

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
信息熵	0.95	0.87	0.84	0.96	0.94	0.96	0.96	0.96	0.96

3. 确定各指标信息熵权重：

$$W_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)}$$

	W ₁	W ₂	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉
权重	0.08	0.22	0.27	0.07	0.11	0.07	0.07	0.07	0.07

4. 对各个科室进行评分：

$$a_i = \sum_{j=1}^k W_j x_{ij} \quad (i = 1, 2, \dots, n)$$

科室	A	B	C	D	E	F	G	H	I	J	K
得分	95.71	93.14	93.17	92.77	95.84	98.01	90.21	95.17	95.97	97.81	97.02