

Answers of Assignment 1

一、

假设有 n 个样本。

由题知，有四个指标，分别为 $V1$ (身高)、 $V2$ (体重)、 $V3$ (出生地)、 $V4$ (年级)。

可以表示为数据矩阵的形式，如下表：

样本	变量			
	V1	V2	V3	V4
1	V11	V12	V13	V14
2	V21	V22	V23	V24
...
n	Vn1	Vn2	Vn3	Vn4

(1) $V1, V2$ 为定量指标，可以取连续值。需进行标准化变换，如下式：

$$V_{ij}^* = \frac{V_{ij} - \bar{V}_j}{s_j}, i = 1, 2, \dots, n, j = 1, 2.$$

其中 \bar{V}_j 为列均值， s_j 为列标准差。

取 Euclidean 距离定义学生信息 $V1$ (身高)、 $V2$ (体重)的距离度量。

$$d_{ij}^{(1)} = \sqrt{\sum_{k=1}^2 (V_{ik}^* - V_{jk}^*)^2}, i, j = 1, 2, \dots, n.$$

(2) $V3, V4$ 为定性指标，取值作如下定义：

将定性变量称为项目，将定性变量的各种不同取“值”称为类目。对于第 i 个样本有，

$$x_{(i)} = (\delta_i(1,1), \delta_i(1,2), \dots, \delta_i(1,r), \delta_i(2,1), \dots, \delta_i(2,5))^T, i = 1, 2, \dots, n.$$

其中 n 为样本个数， r 为 $V3$ 指标所包含的城市/省份数。

$$\delta_i(k,l) = \begin{cases} 1, & \text{第 } i \text{ 个样本中第 } k \text{ 个项目的数据为第 } l \text{ 个类目时,} \\ 0, & \text{otherwise.} \end{cases}$$

若 $\delta_i(k,l) = \delta_j(k,l) = 1$ ，则称 i 和 j 样本在 (k, l) 上 1-1 配对，数量记为 m_1 ；

若 $\delta_i(k,l) \neq \delta_j(k,l)$ ，则称 i 和 j 样本在 (k, l) 上不配对，数量记为 m_2 ；

距离定义为：

$$d_{ij}^{(2)} = \frac{m_1}{m_1 + m_2}$$

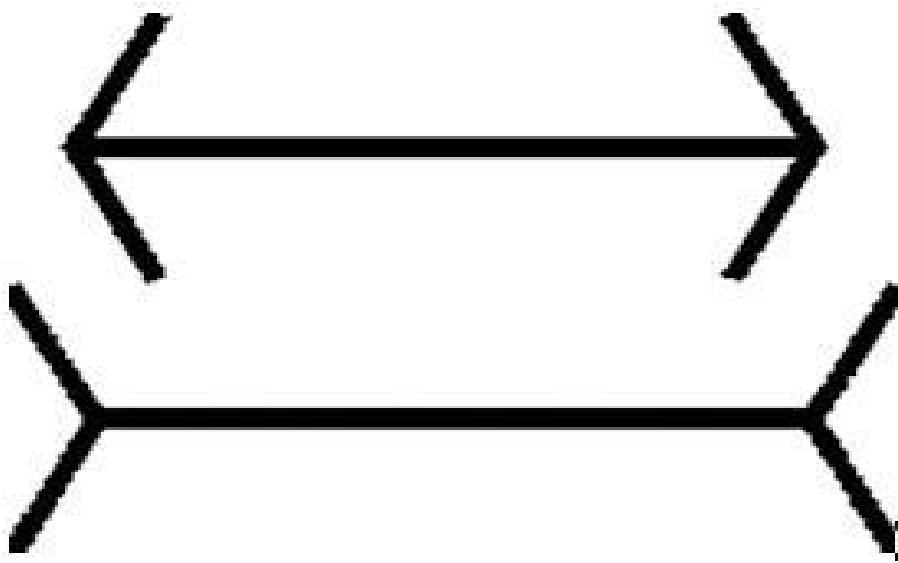
(3) 综上所述，学生信息的距离度量如下：

$$d_{ij} = d_{ij}^{(1)} + d_{ij}^{(2)}, \quad i, j = 1, 2, \dots, n.$$

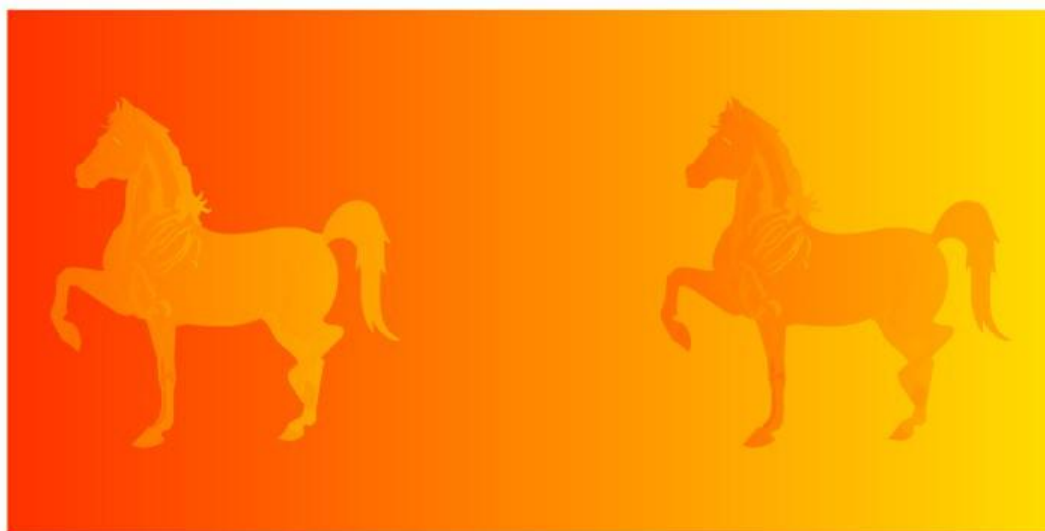
二、代码见.cpp 文件。

三、

图（一）：



图（二）：



第一幅图：人们在描述物体的长度的时候，通常会采用了另外一个物体作为参照物。在此图中会认为上方的线段比下方的线段短。因为上方线段与内扣的箭头相连，下方线段与向外

的箭头相连，视觉感知系统基于相对判断，会错误的认为上方的箭头更短。如果物体使用相同的参照物，则有助于作出更加准确的相对判断。

第二幅图：会认为左边的马和右边的马颜色不同。但是剪切一小部分比较可知，相同部位的颜色是相同的。因为左右两边的马的背景不同，导致了视觉错误。由此可得，相同的颜色在不同的背景下可能会使人产生完全不同的视觉感知。

视觉感知原理：人类感知系统的工作原理决定于对所观察事务的相对判断。感知系统基于相对判断，而非绝对判断。人类感知系统将可察觉的刺激强度的变化表达为一个目标刺激强度的百分比。