

小明新入职的一家互联网公司通过员工的工作时长和客户满意度来评估员工的表现，并将员工分为高效员工和低效员工两类。如图 12 所示，以下是部分员工的数据。请你根据员工的工作时长和客户满意度数据，计算出判别函数，并解释判别分析中间变量的实际意义。小明的工作时长为 160 小时/月，客户满意度为 80，请问他是高效员工吗？

表 12 员工工作信息表

员工编号	工作时长（小时/月）	客户满意度（满分 100）	类别
1	160	85	高效
2	170	90	高效
3	150	78	高效
4	140	65	低效
5	155	60	低效
6	165	70	低效

对于这个问题，如果学生使用手动计算将按照如下步骤计算：

（1）首先对数据进行整理，根据题目给的数据可以将员工划为两类：

类别	工作时长 (小时/月)	客户满意度 (满分 100)
高效	160, 170, 150	85, 90, 78
低效	140, 155, 165	65, 60, 70

（2）计算两个类别的均值向量。

高效员工均值向量：

$$\mu_1 = \left(\frac{160 + 170 + 150}{3}, \frac{85 + 90 + 78}{3} \right) = (160.00, 84.33)$$

低效员工均值向量：

$$\mu_2 = \left(\frac{140 + 155 + 165}{3}, \frac{65 + 60 + 70}{3} \right) = (153.33, 65.00)$$

（3）计算散度矩阵 (S)。

散度矩阵是用来衡量类内数据的分布情况，定义为：

$$S = S_{w1} + S_{w2}$$

高效员工的散度矩阵 S_{w1} ：

$$S_{w1} = \sum_{x \in w_1} (x - \mu_1)(x - \mu_1)^T$$

低效员工的散度矩阵 S_{w2} ：

$$S_{w2} = \begin{bmatrix} 158.33 & 25.00 \\ 25.00 & 25.00 \end{bmatrix}$$

总散度矩阵 S :

$$S = \begin{bmatrix} 258.33 & 85.00 \\ 85.00 & 61.33 \end{bmatrix}$$

(4) 计算判别向量 w 和阈值 θ 。

判别向量 w 公式为:

$$w = S^{-1}(\mu_1 - \mu_2)$$

差值向量:

$$\mu_1 - \mu_2 = (160 - 153.33, 84.33 - 65) = (6.67, 19.33)$$

通过矩阵求逆:

$$S^{-1} = \begin{bmatrix} 0.015 & -0.021 \\ -0.021 & 0.063 \end{bmatrix}$$

代入公式可得:

$$w = S^{-1}(\mu_1 - \mu_2) = [-0.1432, 0.5137]$$

阈值 θ 的计算公式为:

$$\theta = -\frac{1}{2}(\mu_1 + \mu_2)^T w$$

均值和为:

$$\mu_1 + \mu_2 = (160 + 153.33, 84.33 + 65) = (313.33, 149.33)$$

可以计算出

$$\theta = -\frac{1}{2}[(313.33 \cdot 0.12) + (149.33 \cdot 0.28)] = -15.92$$

(5) 求解判别方程。

判别函数的方程为:

$$f(x) = w^T x + \theta$$

代入 w 和 θ 的值求得的判别方程为:

$$f(x) = [-0.1432, 0.5137]^T x - 15.9188$$

(6) 判别结果。

判别函数的方程为:

$$x = (165, 80)$$

将小明的数据代入方程：

$$f(x) = (-0.1432 \cdot 165) + (0.5137 \cdot 80) - 15.9188$$

计算得到：

$$f(x) = -23.628 + 41.096 - 15.9188 = 1.55$$

得到 $f(x) = 1.55 > 0$ ，可见小明为高效员工，问题求解完成。