

题目

在对高维数据降维之前应先进行"中心化", 常见的是将协方差矩阵 XX^T 转化为 XHH^TX^T , 其中 $H = I - \frac{1}{m}11^T$, 试析其效果。

解答

答: 相当于将 X 变为 X' :

$$\begin{aligned}X' &= XH \\&= X \left(I - \frac{1}{m}11^T \right) \\&= X - \frac{1}{m}X11^T \\&= X - \frac{1}{m} \begin{bmatrix} x_1 & x_2 & \cdots \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots \end{bmatrix} \\&= X - \frac{1}{m} \sum_i x_i \begin{bmatrix} 1 & 1 & \cdots \end{bmatrix} \\&= X - \bar{x} \begin{bmatrix} 1 & 1 & \cdots \end{bmatrix} \\&= \begin{bmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \cdots \end{bmatrix}\end{aligned}$$

其效果便是中心化 $x'_i = x_i - \bar{x}$ 。