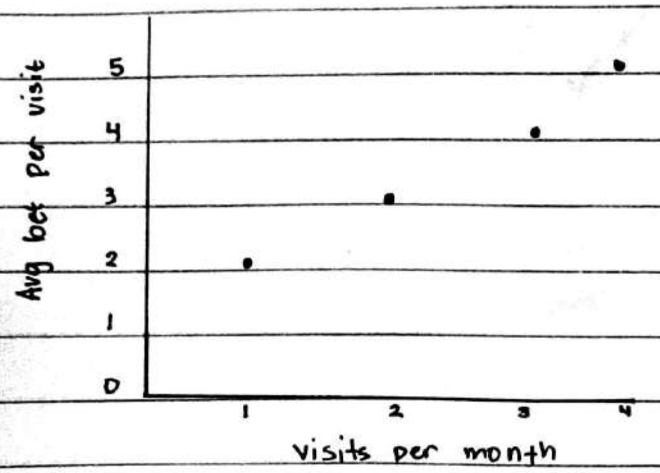
Almadronea, Kurt Justine M. CDM232

Casino Customers Dataset

Customer	Avg Bet per Visit	Visits per Month
A	1	2
B	2	3
C	3	4
D	4	5

1.) Plot the dataset on a 2 dimensional Plane.



2.) Compute the mean of each feature

$\mu_1 = \frac{1+2+3+4}{4} = 2.5 \quad \mu_2 = \frac{2+3+4+5}{4} = 3.5 \quad \mu = [2.5, 3.5]$

3.) Center the dataset by subtracting the mean from each sample

4.) Write down the centered data matrix

$$A_{centered} = \begin{bmatrix} 1-2.5 & 2-3.5 \\ 2-2.5 & 3-3.5 \\ 3-2.5 & 4-3.5 \\ 4-2.5 & 5-3.5 \end{bmatrix} \quad A_{centered} = \begin{bmatrix} -1.5 & -1.5 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \end{bmatrix}$$

5.) Compute the sample covariance matrix of the centered data

$$A_{centered} = \begin{bmatrix} -1.5 \\ -0.5 \\ 0.5 \\ 1.5 \end{bmatrix} \quad A^T_{centered} = \begin{bmatrix} -1.5 & -0.5 & 0.5 & 1.5 \\ -1.5 & -0.5 & 0.5 & 1.5 \end{bmatrix}$$

$$= \Sigma_1 = \frac{1}{n-1} A_c^T A_c = \frac{1}{n-1}\begin{bmatrix} -1.5 & -0.5 & 0.5 & 1.5 \\ -1.5 & -0.5 & 0.5 & 1.5 \end{bmatrix}\begin{bmatrix} -1.5 & -1.5 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \end{bmatrix}$$

$$= \Sigma_1 = \frac{1}{4-1}\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} = \Sigma_1 = \begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix}$$

6.) Compute the eigenvalues of covariance matrix

$\det(A - \lambda I) = 0$

$$= \left\| \begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\| = 0 = \left\| \begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right\| = 0$$

$$= \begin{vmatrix} 1.67-\lambda & 1.67 \\ 1.67 & 1.67-\lambda \end{vmatrix} = 0 = (1.67-\lambda)^2 - 2.79 = 0$$

$= (1.67-\lambda)(1.67-\lambda) - 2.79 = 0$

$\cancel{2.7889} - 1.67\lambda - 1.67\lambda + \cancel{2.7889} / \cancel{2.79} = 0$

$\lambda^2 - 3.34\lambda = 0$

$= (\lambda - 0)(\lambda - 3.34) = 0$

$= \lambda_1 = 3.34$

$\lambda_2 = 0$

7.) Compute the corresponding eigenvectors

$= (A - \lambda I)v = 0$

$$\left(\begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0 = \left(\begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$$= \left(\begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix}\right) - \begin{bmatrix} 3.34 & 0 \\ 0 & 3.34 \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0 = \begin{bmatrix} -1.67 & 1.67 \\ 1.67 & -1.67 \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$$= \begin{bmatrix} -1.67V_1 + 1.67V_2 = 0 \\ 1.67V_1 - 1.67V_2 = 0 \end{bmatrix} = \begin{bmatrix} 1.67V_2 = 1.67V_1 \\ 1.67V_1 = 1.67V_2 \end{bmatrix} = V\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

8.) Normalize the eigenvector

$$u = \frac{V}{\|V\|} \quad \|V\| = \sqrt{1^2 + 1^2} = \sqrt{2}$$

$$V_{normal} = \frac{V}{\|V\|} = \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

9.) Check if the eigenvectors and eigenvalues are correct using $Av = \lambda v$

$Av = \lambda v$

$$\begin{bmatrix} 1.67 & 1.67 \\ 1.67 & 1.67 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3.34 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3.34 & 3.34 \\ 3.34 & 3.34 \end{bmatrix} = \begin{bmatrix} 3.34 & 3.34 \\ 3.34 & 3.34 \end{bmatrix}$$

10.) Identify the first principal component

- Explain why this component is chosen over the others

3.34, was chosen since it explains the maximum variance based on the dataset that was given, ~~man~~ compared to $\lambda = 0$ it explains little to ~~some~~ no variance

11.) Project the centered data onto the first principal component

- new projection = Acentered x u

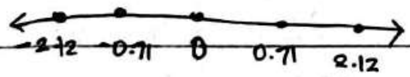
$$\begin{bmatrix} -1.5 & -1.5 \\ -0.5 & -0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$$

1st $= \dfrac{-1.5 - 1.5}{\sqrt{2}} = -2.12$ 3rd $= \dfrac{0.5 + 0.5}{\sqrt{2}} = 0.71$

2nd $= \dfrac{-0.5 - 0.5}{\sqrt{2}} = -0.71$ 4th $= \dfrac{1.5 + 0.5}{\sqrt{2}} = 2.12$

12.) Write down the resulting 1-dimensional representation of each sample

new projection $= \begin{bmatrix} -2.12 \\ -0.71 \\ 0.71 \\ 2.12 \end{bmatrix}$



6