Q2

X_1	X_2
4	11
8	4
13	5
7	14

Approach: PCA

Normalizing the data

↓

This is important because of a theorem that says that the line we obtain ~~after~~ as the principal component axis passes through mean of data [not through ~~origin~~ (0,0)].

$X_1^{(\text{modified})}$	$X_2^{(\text{modified})}$
$4-8=-4$	$11-8.5=2.5$
$8-8=0$	$4-8.5=-4.5$
$13-8=5$	$5-8.5=-3.5$
$7-8=-1$	$14-8.5=5.5$

$$\mu_{X_1} = \frac{4+8+13+7}{4} = 8$$

$$\mu_{X_2} = \frac{11+4+5+14}{4} = 8.5$$

Let $X = \begin{bmatrix} -4 & 2.5 \\ 0 & -4.5 \\ 5 & -3.5 \\ -1 & 5.5 \end{bmatrix}$. $\rightarrow$ [Our choice of X indicates that we are working in a transformed basis:
$x_1 \rightarrow x_1 - \mu_{x_1}$
$x_2 \rightarrow x_2 - \mu_{x_2}$.]

$$X^T X = \begin{bmatrix} -4 & 0 & 5 & -1 \\ 2.5 & -4.5 & -3.5 & 5.5 \end{bmatrix} \begin{bmatrix} -4 & 2.5 \\ 0 & -4.5 \\ 5 & -3.5 \\ -1 & 5.5 \end{bmatrix}$$

$$= \begin{bmatrix} 42 & -33 \\ -33 & 69 \end{bmatrix}$$

<u>Eigenvalues of $X^T X$</u> $\rightarrow$ $\lambda_1 = 19.845$
$\lambda_2 = 91.154$

As $\lambda_2 > \lambda_1$, ~~eigent~~ eigenvector corresponding to λ_2 indicates principal component axis.

$$(XX^T) \begin{bmatrix} x \\ y \end{bmatrix} = \lambda_2 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Rightarrow \begin{aligned} 42x - 33y &= 91.154x \\ -33x + 69y &= 91.154y \end{aligned}$$

$$\Rightarrow \begin{aligned} x &= 0.557 \\ y &= -0.830. \end{aligned}$$

Values of each datapoint in X projected along principal axes is given by

$$X \begin{bmatrix} 0.557 \\ -0.830 \end{bmatrix} = \begin{bmatrix} -4.305 \\ 3.736 \\ 5.692 \\ -5.124 \end{bmatrix}.$$