Q-3.1] - Data = $\{x^1, x^2 \ldots x^n\}$.

$C_1 = \{x^1, x^2 \ldots x^m\}$ $C_2 = \{x^{m+1}, x^{m+2}, \ldots x^n\}$.

Let ~~the~~ cluster centre be $O_1, O_2$ respectively s.t.,

$$O_1 = \frac{\sum_{j=1}^{m} x^j}{m} \qquad O_2 = \frac{\sum_{j=m+1}^{n} x^j}{(n-m)}$$

~~Let~~ ~~there~~ ~~is~~
We want to prove existence for a plane $a^Tx + b = 0$

s.t $a^Tx + b < 0$ $\forall x \in C_1$ $\qquad a^Tx + b > 0$ $\forall x \in C_2$. — (1)

Now, we also know that for any $x \in C_1$ since our solution of $O_1, O_2$ is optimal,

$$\|x - O_1\|^2 < \|x - O_2\|^2$$

and for any $x \in C_2$ — (2).

$$\|x - O_1\|^2 > \|x - O_2\|^2.$$

In order to find a plane, we need to find all such points that are equidistance from $O_1$, $O_2$.

$\therefore$ For any $x$ lying on our plane.

$$\|x - O_1\|^2 = \|x - O_2\|^2$$

$$(x - O_1)^T (x - O_1) = (x - O_2)^T (x - O_2)$$

$$\therefore \cancel{x^T x} - x^T O_1 - O_1^T x + O_1^T O_1 = \cancel{x^T x} - x^T O_2 - O_2^T x + O_2^T O_2$$

Note: $O_1^T x = x^T O_1$ and $O_2^T x = x^T O_2$

$$-2x^T O_1 + O_1^T O_1 = -2x^T O_2 + O_2^T O_2.$$

$$2\left(x^T (O_1 - O_2)\right) = O_1^T O_1 - O_2^T O_2$$

$$x^T (O_1 - O_2) = \frac{\|O_1\|^2 - \|O_2\|^2}{2}.$$

$$\therefore x^T (O_1 - O_2) - \left(\frac{\|O_1\|^2 - \|O_2\|^2}{2}\right) = 0.$$

where $O_1 = \dfrac{\sum_{i=1}^{m} x^i}{m}$ and $O_2 = \dfrac{\sum_{i=m+1}^{n} x^i}{n-m}$

Hence proved, there exists a plan. $a^T x + b$ which separates the two clusters.

~~Q-3.2] (i). Image 1 Cubes~~

~~K = 2 → Since we have two clusters, we can just make out the position of cubes, but their orientation / lighting is unclear.~~

~~K = 5 → Since the cubes image has a very small number of pixels,~~

Q-3.2] As the number of clusters increase, the
ii) image becomes more and more similar to the original image. At K=2, we just have 2 clusters and this gives a very crude segmentatn ie. cube or no cube (image-1)
• foreground vs background (image 2,3)
As K increases to 5, more aspects in the image becom clear, like orientatn of cubes in image 1, race track in image 3.
At K=10, the 1st image is nearly restored while the others give a somewhat 'noisy' segmation of the image components.

iii). Some Images (like image-1) have nearly discrete colours, (and fewer colours), and hence, having less clusters can preserve the information in the images.
But some images (like image-2) have continuous and a huge spectrum of colours. Hence it becomes difficult to preserve information in these images with less clusters and we need more clusters for preserving image-data.