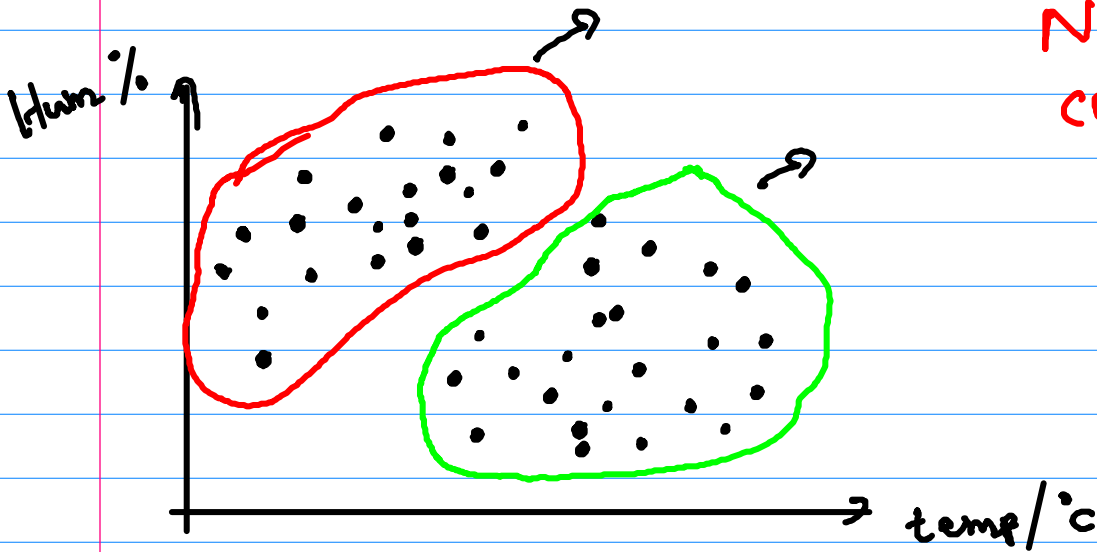


Clustering

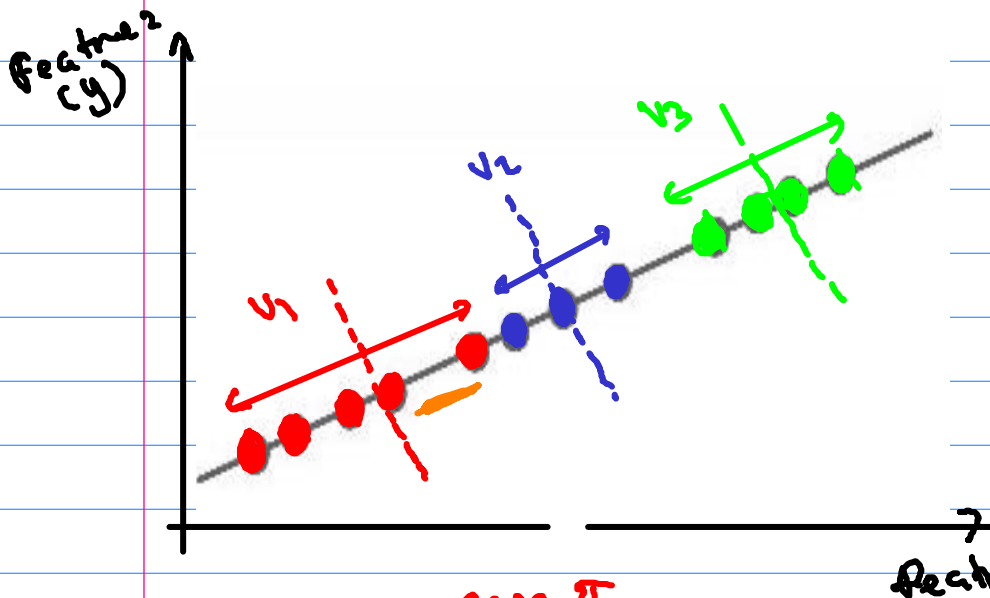
unsupervised learning
No Labels
Can labelled datasets?



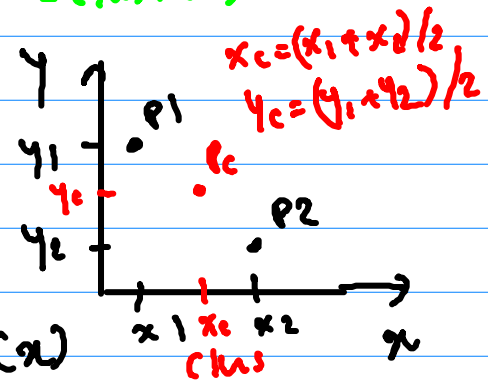
k means Clustering

$k=3$ (k - number of clusters required)

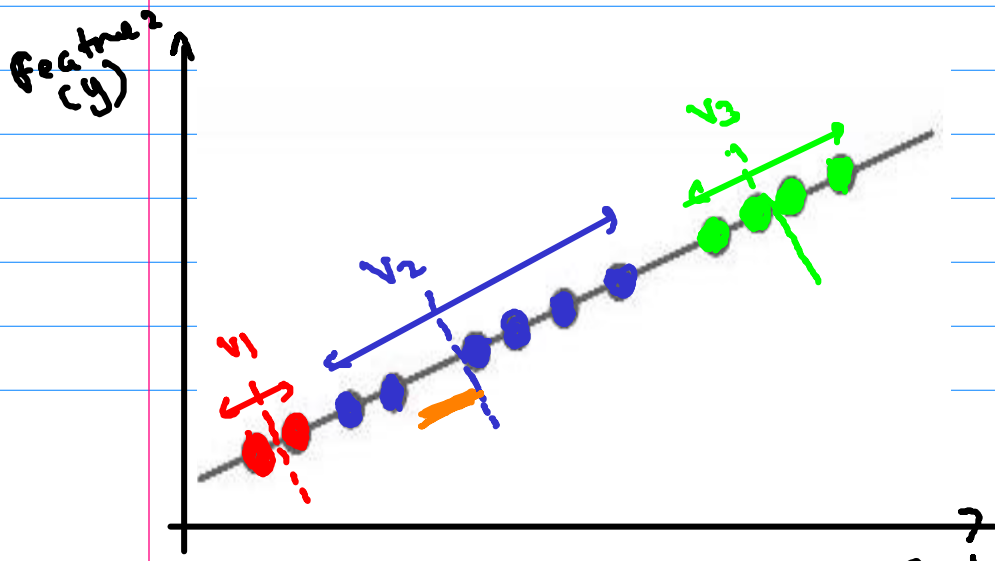
case I



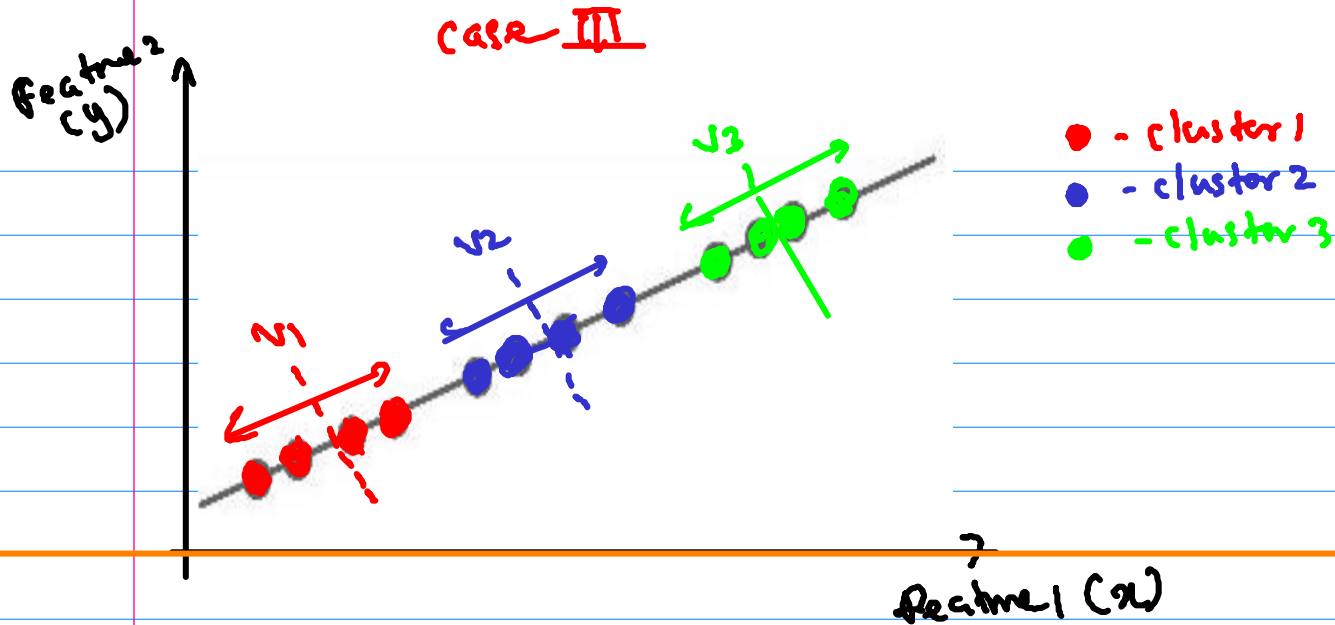
- - cluster 1
- - cluster 2
- - cluster 3



case II



- - cluster 1
- - cluster 2
- - cluster 3



Total Variance

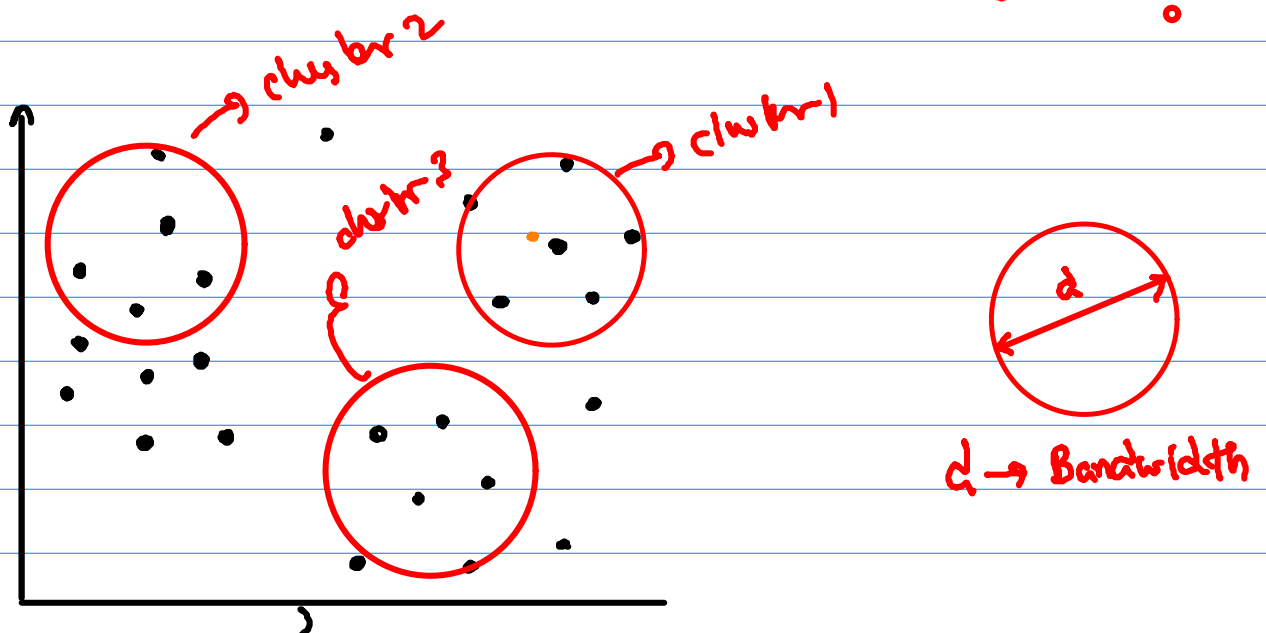
$$TV = V_1 + V_1 + V_3 + \dots - V_K$$

in this example, $TV = v_1 + v_2 + v_3$

* Best model \rightarrow which has the minimum TV

Mean Shift clustering

K is not given !



Bandwidth $\propto \frac{1}{\# \text{ clusters}}$

