Algorithm:

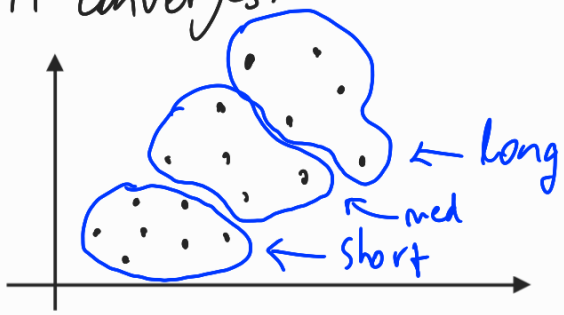
1st: Random n points (n = number of clusters) → cluster centroids

2nd: Go through all training examples.
Calculate distance, assign each to the closest cluster.

3rd: Move centroids to mean location of points closer to each.

4th: Go to 2nd until it converges.

- Non-separated clusters:



Distortion

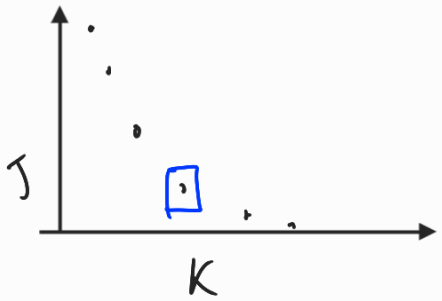
$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - \mu_{c^{(i)}} \|^2$$ (straight line)

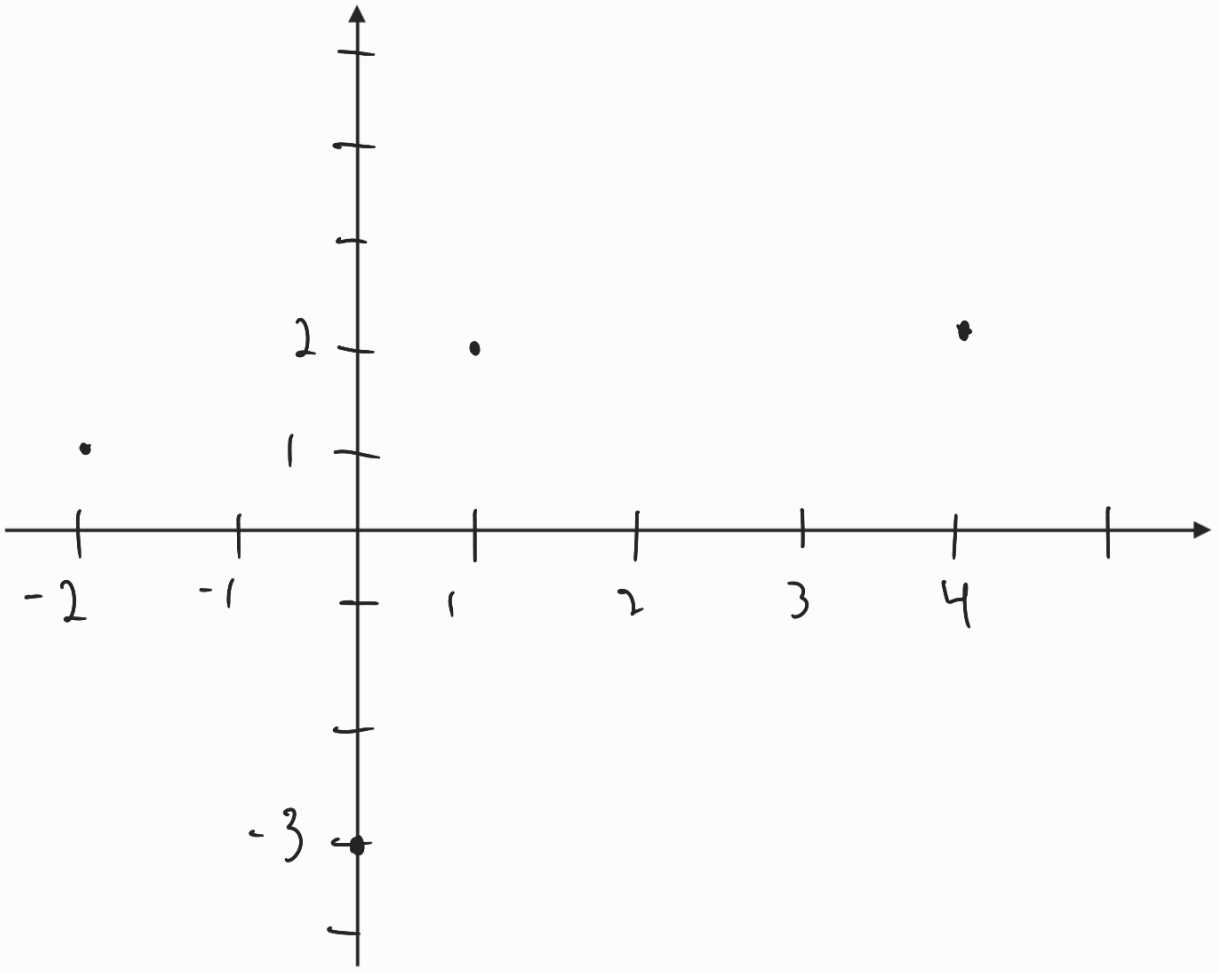
$x^{(i)}$: location of training sample

$\mu_{c^{(i)}}$: location of cluster centroid

- It may fall into local optima with bad initialization
  ↳ Try multiple times. Pick min(J(...))

- Choosing K
  ↳ Elbow method:




2
1
-2
-1
1
2
3
4
-3