Video 4

Monday, March 30, 2020 1:28 PM

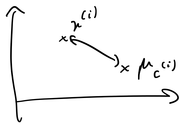
Optimization objective (Distortion of k-means algorithm)

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

$c^{(1)}, \dots, c^{(m)}$: centroid assignment

$\mu_1, \dots, \mu_K$: centroids

$\mu_{c^{(i)}}$: centroid to which $x^{(i)}$ was assigned



2 steps of k-means algorithm:

1. Cluster assignment step

$\min_{c^{(1)}, \dots, c^{(m)}} J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$

wrt $c^{(1)}, \dots, c^{(m)}$

while fixing $\mu_1, \dots, \mu_K$

2. Move centroids step.

$\min_{\mu_1, \dots, \mu_K} J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$

wrt. $\mu_1, \dots, \mu_K$

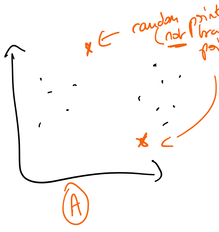
while fixing $c^{(1)}, \dots, c^{(m)}$

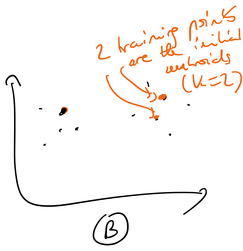
Initialization

Pick the value of K

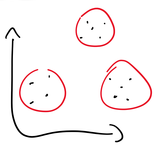
Strategy: (A) pick random points as the K initial cluster centroids. (random points in the input domain)

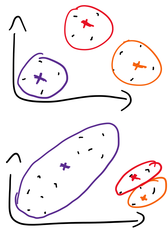
(B) pick randomly K training points as the K initial cluster centroids.





Local Optima





Distortion $J(c^{(1)}, \dots c^{(m)}, \mu_1, \dots, \mu_K)$

For some number of trials:

Randomly initialize K centroids $\mu_1, \dots, \mu_K$

Run k-means to obtain $c^{(1)}, \dots, c^{(m)}$ (assignments) and $\mu_1, \dots, \mu_K$ (centroids)

Compute distortion/cost $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$

Pick random initialization that yields min J

↳ $\mu_1, \dots, \mu_K$, $c^{(1)}, \dots, c^{(m)}$ } keeps the outputs of that trial of K-means.

Heuristic works well when K is small.

Choosing the nb of centroids/clusters: K