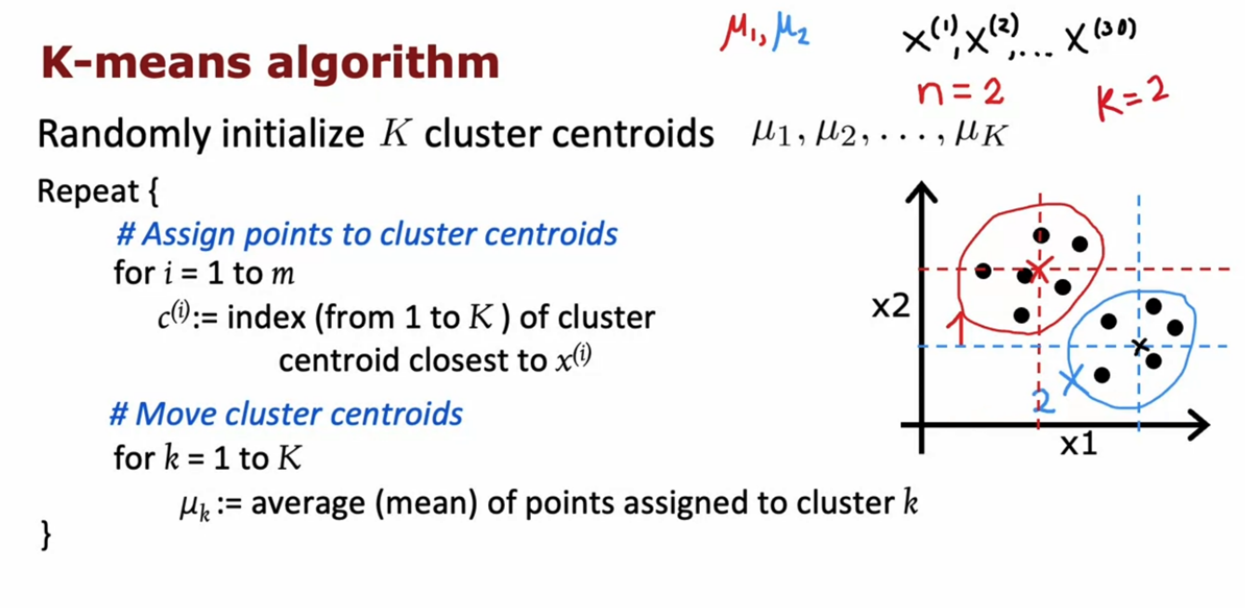
Unsupervised Learning, Recommenders, Reinforcement Learning



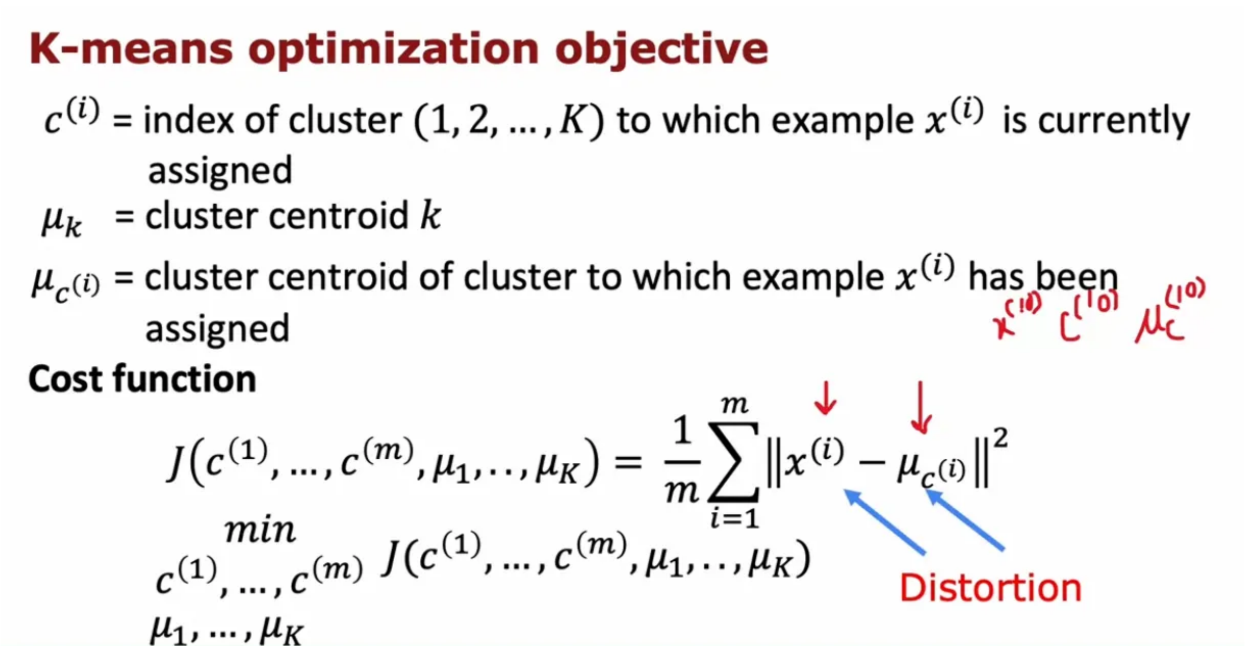
K-means algorithm

μ_1, μ_2 $x^{(1)}, x^{(2)}, \ldots x^{(30)}$

$n = 2$ $K = 2$

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K$

```
Repeat {
    # Assign points to cluster centroids
    for $i$ = 1 to $m$
        $c^{(i)}$ := index (from 1 to $K$) of cluster
                centroid closest to $x^{(i)}$
    # Move cluster centroids
    for $k$ = 1 to $K$
        $\mu_k$ := average (mean) of points assigned to cluster $k$
}
```



K-means optimization objective

$c^{(i)}$ = index of cluster $(1, 2, \ldots, K)$ to which example $x^{(i)}$ is currently assigned

μ_k = cluster centroid k

$\mu_{c^{(i)}}$ = cluster centroid of cluster to which example $x^{(i)}$ has been assigned

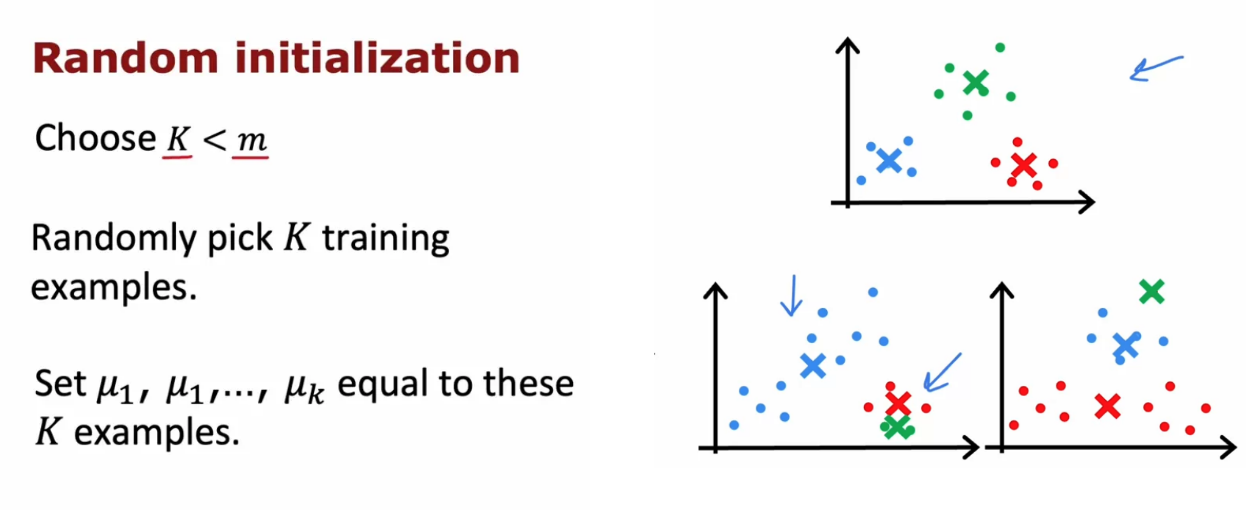
$x^{(10)}$ $c^{(10)}$ $\mu_{c^{(10)}}$

Cost function

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

$$\min_{\substack{c^{(1)}, \ldots, c^{(m)} \\ \mu_1, \ldots, \mu_K}} J\left(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K\right)$$

Distortion



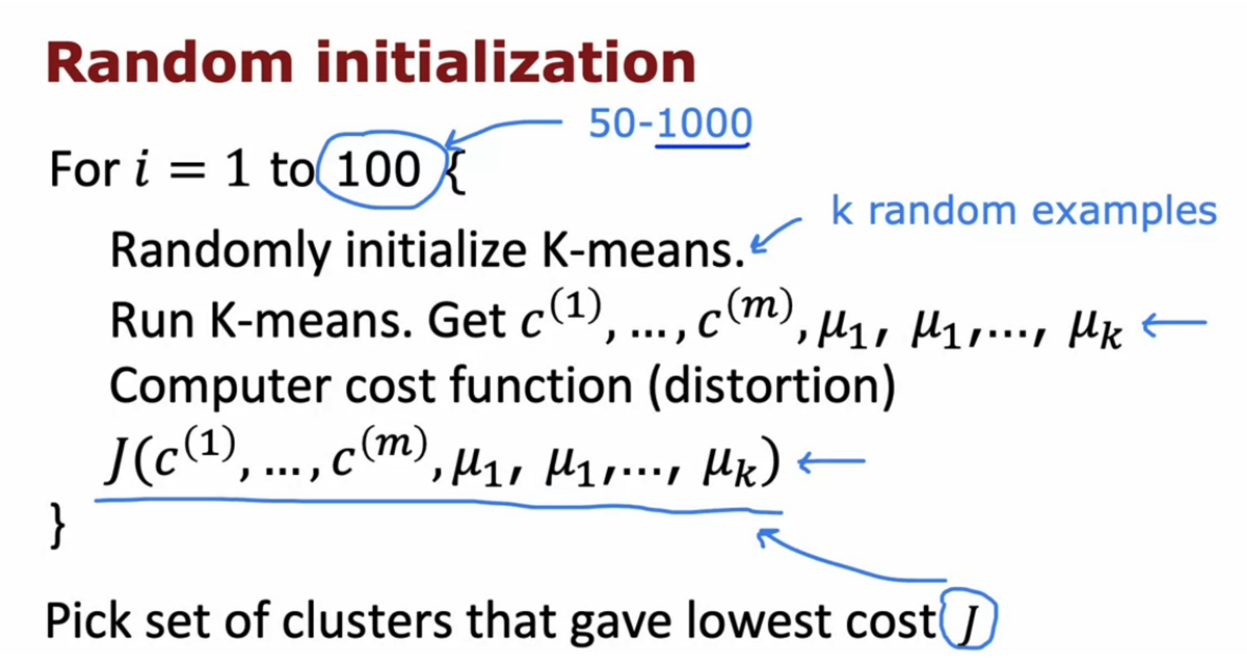
Random initialization

Choose $K < m$

Randomly pick K training examples.

Set $\mu_1, \mu_1, \ldots, \mu_k$ equal to these K examples.

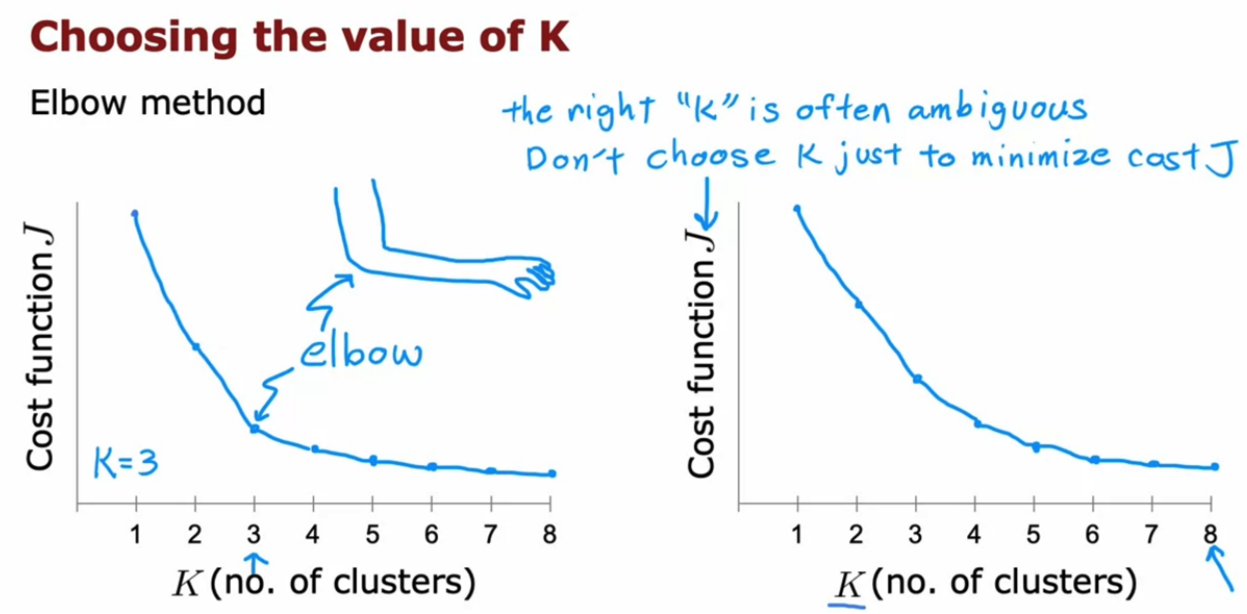
With random initialization using some of the data points themselves, depending on which points you choose, we can end up with different clusters. Local minima of the cost function may be found. We can run it with multiple random initializations and then pick the one with lowest final cost function.



Random initialization

```
For $i = 1$ to $\boxed{100}$ {    50-1000
    Randomly initialize K-means.    k random examples
    Run K-means. Get $c^{(1)}, \ldots, c^{(m)}, \mu_1, \mu_1, \ldots, \mu_k$ ←
    Computer cost function (distortion)
    $\underline{J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \mu_1, \ldots, \mu_k)}$ ←
}
```

Pick set of clusters that gave lowest cost $\boxed{J}$

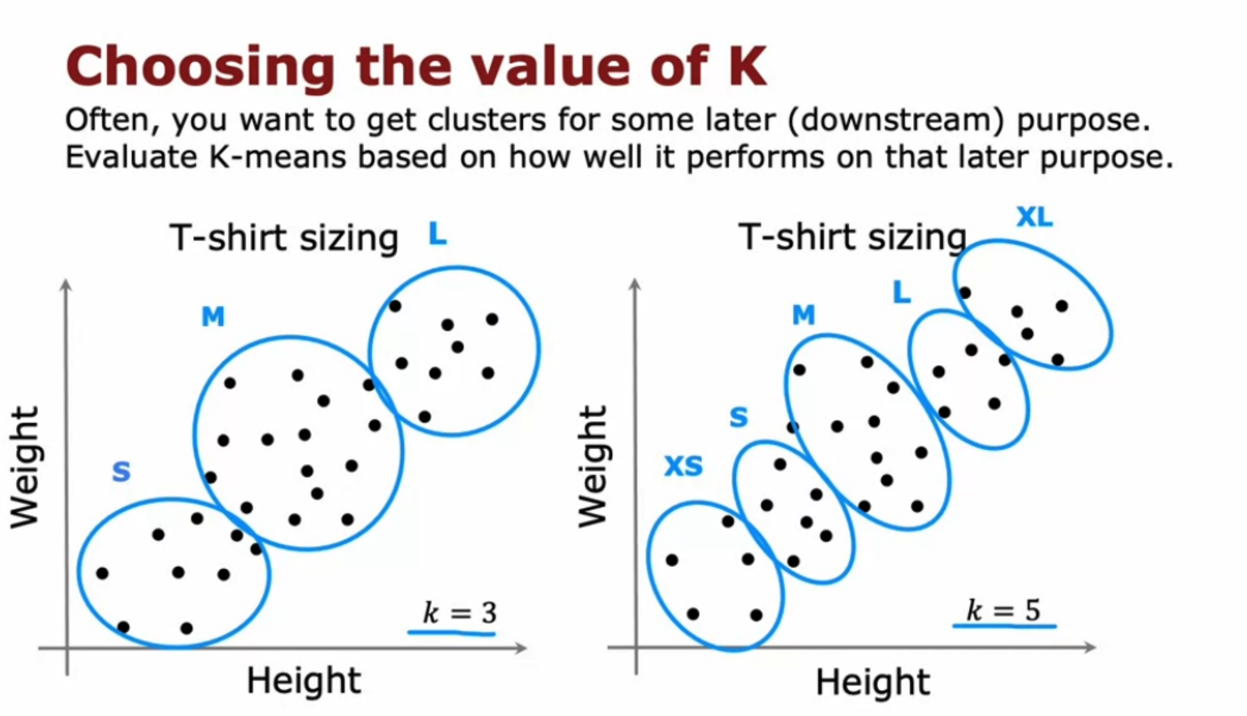


Choosing the value of K

Elbow method

the right "K" is often ambiguous

Don't choose K just to minimize cost J



Choosing the value of K

Often, you want to get clusters for some later (downstream) purpose.
Evaluate K-means based on how well it performs on that later purpose.