Clustering

Damien Benveniste

June 24, 2017

1 k-means

The algorithm

Data: X
Initialize at centroid at random: $\{\mu_1, \mu_2, \ldots, \mu_k\}$
while *not convergence* **do**
  for i *in* $\{0, 1, \ldots, N\}$ **do**
    We attribute each sample to a cluster:
    $c^{(i)} = \arg\min_j \|\mathbf{X}^{(i)} - \mu_j\|$
    for j *in* $\{0, 1, \ldots, k\}$ **do**
      We recompute the position of each centroid:
      $\mu_j = \frac{\sum_{i=1}^{N} \delta_{c^{(i)},j} \mathbf{X}^{(i)}}{\sum_{i=1}^{N} \delta_{c^{(i)},j}}$
    end
  end
end