Writeup assignment 2
CS383 Machine learning

Answers:

Theory Answers:

1) $W_j''' = ?$

$$= \frac{W''_{j+1} - W''_{j-1}}{2}$$

$$= \left[ \frac{\left( \frac{W'_{j+2} - W'_j}{2} \right) - \left( \frac{W'_j - W'_{j-2}}{2} \right)}{2} \right]$$

$$= \left[ \frac{W'_{j+2} - 2W'_j + W'_{j-2}}{2} \right] \Big/ 2$$

$$= \left[ \frac{\frac{W_{j+3} - W_{j+1}}{2} - 2\left(\frac{W_{j+1} - W_{j-1}}{2}\right) + \left(\frac{W_{j-1} - W_{j-3}}{2}\right)}{4} \right]$$

$$= \left[ \frac{W_{j+3} - W_{j+1} - 2W_{j+1} + 2W_{j-1} + W_{j-1} - W_{j-3}}{8} \right]$$

$$= \frac{W_{j+3} - 3W_{j+1} + 3W_{j-1} - W_{j-3}}{8}$$

2) $\text{Purity} = \frac{1}{N} \sum_{i=1}^{n} |C_i| \, \text{Purity}(C_i)$

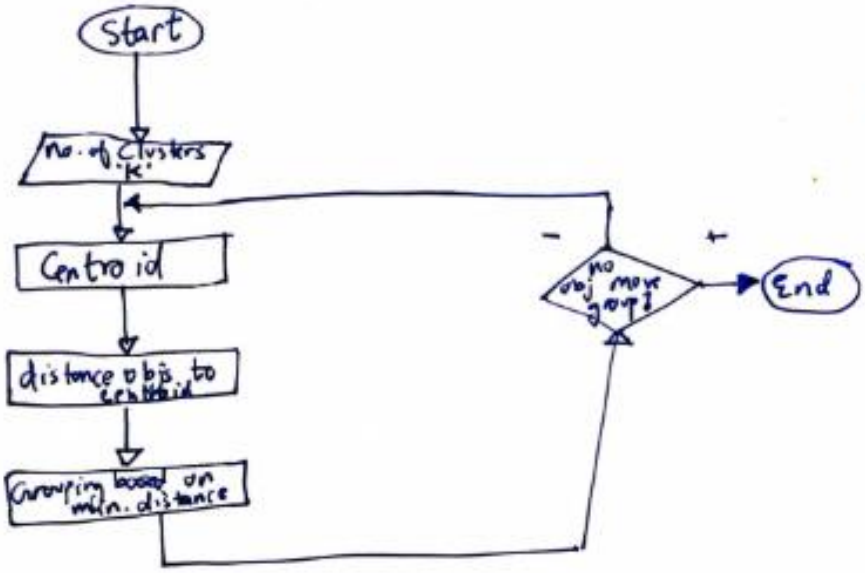
$C_1 = \{1, 2, 3, 4\}$, $C_2 = \{5, 6, 7, 8\}$

$C_1 = \{3, 4\}$ $\quad C_2 = \{1, 2, 5, 6, 7, 8\}$

$\therefore \text{Purity} = \frac{1}{8}[2 + 4] = \frac{6}{8} = \frac{3}{4} = 0.75$

Notes:

# K - means Clustering:




→ define no. of clusters k,

→ centroid = mean value, 2 → 2 centroids
3 → 3 centroids

→ find dist. of obj from centroids.

→ Based on $\hat{}$ distance, we classify to a specific cluster

→ repeat process: if group changes, we move to cluster diff

Example for K=3

Data: 2 5 7 8 13 16 18 28 30

C1 (2) assume — C2 (13) assume — C3 (30) assume

1) $c_1$ [ 2 5 7 ] M = 4.6 — [ 8 13 16 18 ] M = 13.5 — [ 28 30 ] M = 29

2) [ 2 5 7 8 ] M = 5.5 — [ 13 16 18 ] M = 15.6 — [ 28 30 ] M = 29 — run loop again — find centroids

K-means in 2 features:

using K=2.

Initiation - Randomly we chose following two Centroids for 2 clusters

I) In this case, the two centroids are:

$m_1 = (1.0, 1.0)$ & $m_2 = (5.0, 7.0)$

Data

| Individual | Var$_1$ | Var$_2$ |
|---|---|---|
| 1 | 1.0 | 1.0 |
| 2 | 1.5 | 2.0 |
| 3 | 3.0 | 4.0 |
| 4 | 5.0 | 7.0 |
| 5 | 3.5 | 5.0 |
| 6 | 4.5 | 5.0 |
| 7 | 3.5 | 4.5 |

| | Ind. | $\mu$ Value |
|---|---|---|
| Group 1 | 1 | (1.0, 1.0) |
| Group 2 | 4 | (5.0, 7.0) |

$\sqrt{(m-x_i)^2 + (m-y_i)^2}$

$m$ value which we use to calc. value wrt $C_1$ / $C_2$

eg

$d(m_1, 2) = \sqrt{|1.0-1.5|^2 + |1.0-2.0|^2} = 1.12$

$d(m_2, 2) = \sqrt{|5.0-1.5|^2 + |7.0-2.0|^2} = 6.10$

II) → we do this for all observations. (Create an updated table of sorts)

→ we thus get 2 clusters containing $\{1, 2, 3\}$ & $\{4, 5, 6, 7\}$

their new centroids are

| | Cent 1 | Cent 2 |
|---|---|---|
| 1 | 0 | 7.21 |
| 2 | 1.12 | 6.10 |
| 3 | 3.61 | 3.61 |
| 4 | 7.21 | 0 |
| 5 | 4.72 | 2.5 |
| 6 | 5.30 | 2.06 |
| 7 | 4.30 | 2.92 |

$m_1 = \left(\frac{1}{3}(1.0+1.5+3.0), \frac{1}{3}(1.0+2.0+4.0)\right) = (1.83, 2.33)$

$m_2 = \left(\frac{1}{4}(5.0+3.5+4.5+3.5), \frac{1}{4}(7.0+5.0+5.0+4.5)\right) = (4.12, 5.38)$

III) - Now using these centroids, we compute the Euclidean distance of each object as shown

- New clusters are $\{1, 2, 3\}$ & $\{4, 5, 6, 7\}$

- Next centroids are $m_1 = (1.25, 1.5)$ & $m_2 = (3.9, 5.1)$

| | Cent 1 | Cent 2 | |
|---|---|---|---|
| 1 | 1.57 | 5.38 | Result |
| 2 | 0.47 | 4.28 | |
| 3 | 2.04 | 1.78 | |
| 4 | 5.64 | 1.84 | |
| 5 | 3.15 | 0.73 | |
| 6 | 3.78 | 0.54 | |
| 7 | 2.74 | 1.08 | |

IV) we run again until observations stop changing groups. (we're calculating clusters again after each step)

References:

https://www.youtube.com/watch?v=DdlrYLqq71M
https://www.mathworks.com/help/matlab/math/basic-matrix-operations.html
https://www.mathworks.com/help/matlab/ref/double.normalize.html
https://www.mathworks.com/help/matlab/ref/colorspec.html