CS 445/545 Machine Learning

Winter Term, 2014

Quiz 6

Name Solutions

This quiz is closed book/notes. You may use a calculator. Please write all answers on these sheets.

1. Suppose you have the following data set:

Instance	Class
$\mathbf{x}_1$: (2,1)	1
$\mathbf{x}_2$: (0,1)	1
$\mathbf{x}_3$: (2,2)	−1
$\mathbf{x}_4$: (4,0)	−1
$\mathbf{x}_5$: (6,1)	1

(a) Given initial cluster centers $\mathbf{m}_1 = (0, 0)$ and $\mathbf{m}_2 = (4, 3)$, give the cluster membership of each cluster and the new cluster centers after one iteration of K-means clustering. **(2 points)**

$d(x, (0,0))$

$x_1 : \sqrt{5}$

$x_2 : 1$

$x_3 : \sqrt{8}$

$x_4 : 4$

$x_5 : \sqrt{37}$

$d(x, (4,3))$

$x_1 : \sqrt{8}$

$x_2 : \sqrt{20}$

$x_3 : \sqrt{5}$

$x_4 : 3$

$x_5 : \sqrt{8}$

Cluster memberships:

$m_1 : x_1, x_2$

$m_2 : x_3, x_4, x_5$

New cluster centers

$m_1 = \text{Average}[x_1, x_2] = (1, 1)$

$m_2 = \text{Average}[x_3, x_4, x_5) = (4, 1)$

(b) Give the sum-squared error (SSE) of the clustering you obtained in part (a), where

$$SSE = \sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} d(\mathbf{x}, \mathbf{m}_i)^2$$

(1 point)

$d(x_1, m_1)^2 + d(x_2, m_1)^2 + d(x_3, m_2)^2 + d(x_4, m_2)^2 + d(x_5, m_2)^2$

$= 1 + 1 + 5 + 1 + 4$

$= 12$

(c) Give the sum-squared separation of the clustering you obtained in part (a), where

$$Sum\ Squared\ Separation\ (clustering) = \sum_{\text{all distinct pairs of clusters } i, j\ (i \neq j)} \mathbf{d}(m_i, m_j)^2$$

(1 point)

$d(m_1, m_2)^2 = \left(\sqrt{3^2}\right)^2 = 9$

(d) Using the class assignments given above, give the mean entropy of the clustering you obtained in part (a), where

$$mean\ entropy(Clustering) = \sum_{1}^{K} \frac{m_i}{m}\ entropy(C_i)$$

where m_i is the number of instances in cluster i

and m is the total number of instances in the clustering.

and

$$entropy(C_i) = -\sum_{j=1}^{|Classes|} p_{i,j} \log_2 p_{i,j}$$

where

$p_{i,j}$ = probability that a member of cluster i belongs to class j

$= \frac{m_{i,j}}{m_i}$, where $m_{i,j}$ is the number of instances in cluster i with class j

and m_i is the number of instances in cluster i

(2 points)

cluster m_1 : [1, 1] Entropy = 0
classes

cluster m_2 : [-1, -1, 1] Entropy = $-\left[\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right]$
classes

$= .918$

Mean entropy = $\frac{2}{5}[0] + \frac{3}{5}[.918] = .551$

(e) Given a new instance, $\mathbf{x}_6 = (2, 3)$, what class would be assigned to this instance by the clustering given in part a? **(1 point)**

$d((2,3), m_1) = \sqrt{1^2+2^2} = \sqrt{5}$

$d((2,3), m_2) = \sqrt{2^2+2^2} = \sqrt{8}$

(2,3) is assigned to cluster m_1.

class = 1

2. List two weaknesses or limitations of K-means clustering, and for each of these, explain a method for dealing with this weakness or limitation. **(2 points)**