THE CHINESE UNIVERSITY OF HONG KONG, SHENZHEN

DDA 3020

MACHINE LEARNING

Assignment4 Report

Author:
Zhao Rui

Student Number:
121090820

May 14, 2023

Contents

1 Written Questions

1.1 Question 1

Question 1:

(1) $p(x_i|\mu_k) = \prod_{j=1}^{D} \mu_{kj}^{x_{ij}} (1-\mu_{kj})^{1-x_{ij}}$

$$\log p(x_i|\mu_k) = \sum_{j=1}^{D} \left[x_{ij} \log \mu_{kj} + (1-x_{ij}) \log(1-\mu_{kj})\right]$$

$$L(q;\mu) = \sum_i r_{ik} \sum_{j=1}^{D} \left[x_{ij} \log \mu_{kj} + (1-x_{ij}) \log(1-\mu_{kj})\right]$$

$$\frac{\partial L(q;\mu)}{\partial \mu_{kj}} = \sum_i \left[r_{ik}\left(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{kj}}\right)\right] = 0$$

$$\mu_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}}$$

(2) $\beta(\alpha, \beta, \mu_{kj}) = A \mu_{kj}^{\alpha-1} (1-\mu_{kj})^{\beta-1}$

$$\log \beta(\alpha, \beta, \mu_{kj}) = \log A + (\alpha-1) \log \mu_{kj} + (\beta-1) \log(1-\mu_{kj})$$

$$\frac{\partial \log \beta(\alpha, \beta, \mu_{kj})}{\partial \mu_{kj}} + \frac{\partial L(q;\mu)}{\partial \mu_{kj}} = 0$$

$$\therefore \frac{\alpha-1}{\mu_{kj}} - \frac{\beta-1}{1-\mu_{kj}} + \sum_i \left[r_{ik}\left(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{kj}}\right)\right] = 0$$

$$\therefore \mu_{kj} = \frac{(\sum_i r_{ik} x_{ij}) + \alpha - 1}{(\sum_i r_{ik}) + \alpha + \beta - 2}$$

1.2 Question 2

(1) choose A_1, A_4, A_7 as initialized cluster center

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
A_1	0.0 ✓	5.0	8.5	3.6	7.1	7.2	8.1	2.2
A_4	3.6	4.2	5.0 ✓	0 ✓	3.6 ✓	4.1 ✓	7.2	1.4 ✓
A_7	8.1	3.2 ✓	7.3	7.2	6.7	5.4	0 ✓	7.6

$\{A_1\}$ $\{A_3, A_4, A_5, A_6, A_8\}$ $\{A_2, A_7\}$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
$(2,10)$	0.0 ✓	5.0	8.5	3.6	7.1	7.2	8.1	2.2 ✓
$(6,6)$	5.7	4.1	2.8 ✓	2.2 ✓	1.4 ✓	2.0 ✓	6.4	3.6
$(\frac{3}{2},\frac{7}{2})$	6.5	1.6 ✓	6.5	5.7	5.7	4.5	1.6 ✓	6.0

$\{A_1, A_8\}$ $\{A_3, A_4, A_5, A_6\}$ $\{A_2, A_7\}$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
$(3, \frac{19}{2})$	1.1 ✓	4.6	7.4	2.5 ✓	6.0	6.3	7.8	1.1 ✓
$(\frac{13}{2}, \frac{21}{4})$	6.5	4.5	2.0 ✓	3.1	0.6 ✓	1.4 ✓	6.4	4.5
$(\frac{3}{2}, \frac{7}{2})$	6.5	1.6 ✓	6.5	5.7	5.7	4.5	1.6 ✓	6.0

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

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
$(\frac{11}{3}, 9)$	1.9	4.3	6.6	1.7	5.2	5.5	7.5	0.3
$(7, \frac{13}{3})$	7.6	5.0	1.1	4.2	0.7	1.1	6.4	5.6
$(\frac{3}{2}, \frac{7}{2})$	6.5	1.6	6.5	5.7	5.7	4.5	1.6	6.0

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

clustering do not change so

$\{A_1, A_4, A_8\}$ $\{A_3, A_5, A_6\}$ $\{A_2, A_7\}$

(2)

	1	2	3
A_1	7.6	6.5	1.9 ✓
A_2	5.0	1.6 ✓	4.3
A_3	1.0 ✓	6.5	6.6
A_4	4.2	5.7	1.7 ✓
A_5	0.7 ✓	5.7	5.2
A_6	1.1 ✓	4.5	5.5
A_7	6.4	1.6 ✓	7.5
A_8	5.6	6.0	0.3 ✓

center 1: $(7, \frac{13}{3})$ $\{3, 5, 6\}$
center 2: $(\frac{3}{2}, \frac{7}{2})$ $\{2, 7\}$ do not change.
center 3: $(\frac{11}{3}, 9)$ $\{1, 4, 8\}$

violate A_2 and A_6 must link

then change into

center 1: $(\frac{15}{2}, \frac{9}{2})$ $\{3, 5\}$
center 2: $(3, \frac{11}{3})$ $\{2, 6, 7\}$
center 3: $(\frac{11}{3}, 9)$ $\{1, 4, 8\}$

	1	2	3
A_1	7.8	6.4	1.9 ✓
A_2	5.5	1.7 ✓	4.3
A_3	0.7 ✓	5.0	6.6
A_4	4.3	4.8	1.7 ✓
A_5	0.7 ✓	4.2	5.2
A_6	1.6	3.0 ✓	5.5
A_7	7.0	2.6 ✓	7.5
A_8	5.7	5.4	0.3 ✓

the class do not change, so we can classify as

$\{A_3, A_5\}$ $\{A_2, A_6, A_7\}$ $\{A_1, A_4, A_8\}$

1.3 Question 3

Question 3:

$$X = \begin{bmatrix} 7 & 4 & 6 & 8 & 8 & 7 & 5 & 9 & 7 & 8 \\ 4 & 1 & 3 & 6 & 5 & 2 & 3 & 5 & 4 & 2 \\ 3 & 8 & 5 & 1 & 7 & 9 & 3 & 8 & 5 & 2 \end{bmatrix}$$

mean column vector $\mu = \begin{bmatrix} 6.9 \\ 3.5 \\ 5.1 \end{bmatrix}$

$$\Sigma = \frac{1}{10}(X-\mu)(X-\mu)^{\top} = \begin{bmatrix} 20.9 & 14.5 & -3.9 \\ 14.5 & 22.5 & -11.5 \\ -3.9 & -11.5 & 70.9 \end{bmatrix}$$

$$q_1 = \begin{bmatrix} -0.70 \\ 0.71 \\ 0.08 \end{bmatrix} \quad q_2 = \begin{bmatrix} 0.70 \\ 0.66 \\ 0.27 \end{bmatrix} \quad q_3 = \begin{bmatrix} -0.14 \\ -0.25 \\ 0.96 \end{bmatrix}$$

$\lambda_1 = 6.75 \quad \lambda_2 = 33.09 \quad \lambda_3 = 74.47$

new representation $U = \begin{bmatrix} -0.14 & 0.70 \\ -0.25 & 0.66 \\ 0.96 & 0.27 \end{bmatrix}$

$$U^{\top}(X-\mu) = \begin{bmatrix} -2.15 & 3.80 & 0.15 & -4.71 & 1.29 & 4.10 & -1.63 & 2.11 & -0.23 & -2.75 \\ -0.17 & -2.89 & -0.99 & 1.30 & 2.28 & 0.14 & -2.23 & 3.25 & 0.37 & -1.07 \end{bmatrix}$$

2 Programming Question

In this programming question, first we use the PCA to project the data points into two-dimensional subsapce. So the step of PCA shows as follow:

Step 1: Calculate the empirical covariance matrix $\mathbf{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(x^{(n)} - \mu)(x^{(n)} - \mu)^{\top}$

Step 2: Do SVD decomposition of $\mathbf{\Sigma}$ to obtain its D eigenvalues and eigenvectors, and rank them from large to small according to the eigenvalues.

Step 3: Pick the top-K eigenvectors to form the matrix $\mathbf{U} = [\mathbf{q_1}, ..., \mathbf{q_K}] \in \mathbb{R}^{D \times K}$

Step 4: The new representation of $x^{(n)}$ is $\mathbf{U}^{\top}(x^{(n)} - \mu)$

Then after the PCA, we can get all 210 data which have two dimensions.

Then we use K-means to classify the data with the following steps:

First, choose K — the number of clusters. Then randomly put K feature vectors, called centroids, into the feature space, here I choose the first three vectors.

Next, compute the distance from each example x to each centroid c using the Euclidean distance. The distance should be $\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$Then we assign the closest centroid to each example (like if we labeled each example with a centroid id as the label).

Besides, for each centroid, we calculate the average feature vector of the examples labeled with it. These average feature vectors become the new locations of the centroids.

In addition, we recompute the distance from each example to each centroid, modify the assignment and repeat the procedure until the assignments don't change after the centroid locations are recomputed.

Finally we conclude the clustering with a list of assignments of centroids IDs to the examples.

After classifying, we need to calculate the silhouette coefficient. For each data point, we calculte the s_i as follow:

$s_i = \frac{b-a}{max(a,b)}$

here a is the mean distance between a point and all other points in the same cluster and b is the mean distance between a point and all other points in the next nearest cluster.

Then the silhouette coefficient should be the mean of s_i, so the silhouette coefficient is $s = \frac{1}{210}\sum_{i=1}^{2} 10s_i = 0.49514929663075496$

Another way to evaluate the performance of the classification is the rand index, given a set of n samples S, we calculate the rand index as follows:

We have the K-means classified result noted as X and the result of the dataset noted as Y. The rand index equal to $\frac{a+b}{a+b+c+d}$ where a is the number of pairs of elements in S that are in the same subset in X and in the same subset in Y. b is the number of pairs of elements in S that are in the different subset in X and in the different subset in Y. c is the number of pairs of elements in S that are in the same subset in X and in the different subset in Y d is the number of pairs of elements in S that are in the different subset in X and in the same subset in Y.

And finally we calculate the rand index equal to 0.8713602187286398.