ECE 449 Machine Learning (21Fa): Assignment 5

Xu Ke 3190110360

Dec. 26, 2021

1 Problem 1: K-means Cluster

For A1(2, 10), with $P(a, b) = |x2 - x1| + |y2 - y1|$ we will have:

$$\begin{cases} Cluster1 & P(A1, A1) = 0; \\ Cluster2 & P(A1, A4) = 5; \\ Cluster3 & P(A1, A7) = 9; \end{cases} \tag{1}$$

which implies that A1 belongs to Cluster 1.
Similarly, for A2(2, 5) we will have:

$$\begin{cases} Cluster1 & P(A2, A1) = 5; \\ Cluster2 & P(A2, A4) = 6; \\ Cluster3 & P(A2, A7) = 4; \end{cases} \tag{2}$$

which implies that A2 belongs to Cluster 3.
Similarly, for A3(8, 4) we will have:

$$\begin{cases} Cluster1 & P(A3, A1) = 12; \\ Cluster2 & P(A3, A4) = 7; \\ Cluster3 & P(A3, A7) = 9; \end{cases} \tag{3}$$

which implies that A3 belongs to Cluster 2.
Similarly, for A4(5, 8) we will have:

$$\begin{cases} Cluster1 & P(A4, A1) = 5; \\ Cluster2 & P(A4, A4) = 0; \\ Cluster3 & P(A4, A7) = 10; \end{cases} \tag{4}$$

which implies that A4 belongs to Cluster 2.
Similarly, for A5(7, 5) we will have:

$$\begin{cases} Cluster1 & P(A5, A1) = 10; \\ Cluster2 & P(A5, A4) = 5; \\ Cluster3 & P(A5, A7) = 9; \end{cases} \tag{5}$$

which implies that A5 belongs to Cluster 2.
Similarly, for A6(6, 4) we will have:

$$\begin{cases} Cluster1 & P(A6, A1) = 10; \\ Cluster2 & P(A6, A4) = 5; \\ Cluster3 & P(A6, A7) = 7; \end{cases} \tag{6}$$

which implies that A6 belongs to Cluster 2.
Similarly, for A7(1, 2) we will have:

$$\begin{cases} Cluster1 & P(A7, A1) = 9; \\ Cluster2 & P(A7, A4) = 10; \\ Cluster3 & P(A7, A7) = 0; \end{cases} \tag{7}$$

which implies that A7 belongs to Cluster 3.

Similarly, for A8(4, 9) we will have:

$$\begin{cases} Cluster1 & P(A8, A1) = 3; \\ Cluster2 & P(A8, A4) = 2; \\ Cluster3 & P(A8, A7) = 10; \end{cases} \tag{8}$$

which implies that A8 belongs to Cluster 2.

In short, as mentioned above, it is clear that we have A1 in Cluster1, A3,A4,A5,A6,A8 in Cluster2 and A2,A7 in Cluster 3. Thus, for Cluster1 the center point is still A1 $\boxed{(2, 10)}$ obviously. For Cluster2, the new center point will be the average of each point included, which is:

$$(\frac{8+5+7+6+4}{5}, \frac{4+8+5+4+9}{5}) = \boxed{(6, 6)} \tag{9}$$

Similarly, for Cluster3 we have the new center point $\boxed{(\frac{3}{2}, \frac{7}{2})}$.

2 Problem 2: Active Learning

2.1 definition of active learning

Active learning is a semi-supervised machine learning in which algorithm can interactively query the information source to label new data points with the desired outputs. There are situations in which unlabeled data is abundant but manual labeling is expensive, learning algorithms can actively query the information source for labels. The number of examples to learn a concept can often be much lower than the number required in normal supervised learning.

2.2 difference between uncertainty sampling and density-weighted methods

In uncertainty sampling framework, an active learner queries the instances about which it is least certain how to label, which is often straightforward for probabilistic learning models. A more general uncertainty sampling strategy uses entropy as an uncertainty measure. An alternative to entropy in more complex settings involves querying the instance whose best labeling is the least confident. **Basically, there are 3 strategies: least confident, margin sampling and entropy**

The information density framework presents a density-weighted technique, in which the main idea is that **informative instances should not only be those which are uncertain, but also those which are "representative" of the input distribution**. Reported results in density-weighted approaches are superior to methods that do not consider density or representativeness measures. Furthermore, researches show that if densities can be pre-computed efficiently and cached for later use, the time required to select the next query is essentially no different than the base informativeness measure (e.g., uncertainty sampling).

Acknowledgement: I got some idea mentioned above from Section 3.1 and Section 3.6 of a paper named *Active Learning Literature Survey* directed by Burr Settles, from University of Wisconsin–Madison in 2009.

3 Problem 3: GNN

Since the Graph Convolution Networks(GCN) is based on GCN Aggregation which is very similar to Neighborhood Aggregation, thus the number of layers we need depends on the longest path in the graph from A to G, which in this case is 4 clearly. Thus, we should use $\boxed{4}$ layers.

4 Problem 4: PCA Algorithm

In this case, we can start from m by n (in this case, 2 by 6) data matrix X, which is:

$$X = \begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 5 & 3 & 6 & 7 & 8 \end{bmatrix}, \tag{10}$$

then, we can take the mean of matrix X, namely $\hat{X}$, which is:

$$\hat{X} = \begin{bmatrix} \frac{2+3+4+5+6+7}{6} \\ \frac{1+5+3+6+7+8}{6} \end{bmatrix} = \begin{bmatrix} 4.5 \\ 5 \end{bmatrix} \tag{11}$$

Then, we can recenter, that is subtract mean from each row of X, which is:

$$X_c = X - \hat{X} = \begin{bmatrix} -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 \\ -4 & 0 & -2 & 1 & 2 & 3 \end{bmatrix} \tag{12}$$

Then, we can compute covariance matrix, $\Sigma = \frac{1}{m} X_c X_c^T$, which is:

$$\Sigma = \frac{1}{m} X_c X_c^T = \frac{1}{6} \begin{bmatrix} -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 \\ -4 & 0 & -2 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -2.5 & -4 \\ -1.5 & 0 \\ -0.5 & -2 \\ 0.5 & 1 \\ 1.5 & 2 \\ 2.5 & 3 \end{bmatrix} = \begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix} \tag{13}$$

Then, we need to find the eigenvalues of Σ, which will satisfy $|M - \lambda I| = 0$:

$$\begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2.92 - \lambda & 3.67 \\ 3.67 & 5.67 - \lambda \end{bmatrix} = 0 \tag{14}$$

Then, we will have to solve:

$$(2.92 - \lambda)(5.67 - \lambda) - (3.67 \cdot 3.67) = 0 \tag{15}$$

which is equivalent to the equation:

$$\lambda^2 - 8.59\lambda + 3.09 = 0 \tag{16}$$

where we can get two solutions, namely $\lambda_1 = 8.22$ (selected as highest eigenvalue) and $\lambda_2 = 0.38$ (ignored). Then, substitute the eigenvalue back, we will get the eigenvector, which is:

$$\begin{bmatrix} 2.92 - 8.22 & 3.67 \\ 3.67 & 5.67 - 8.22 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = 0 \tag{17}$$

Thus, we will get the eigenvector, which is:

$$\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix} \tag{18}$$

Thus, after the normalization, the principal component will be:

$$\begin{bmatrix} \hat{x_0} \\ \hat{x_1} \end{bmatrix} = \boxed{\begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix}} \tag{19}$$