CSCI 567 : ASSIGNMENT- 4

NAME: NEEL SHAH

USC ID: 5417391181

Email id: neelshah@usc.edu

SOLUTION 1

The hidden layer consists of 4 neurons. Cosider the total net input to each hidden layer neuron using the logistic function as the activation function. Then repeat the process with the output layer neurons.
Let the first hidden layer neuron be h_1

$$
\begin{aligned}
k &= 1,2,3,4 \\
net_{h1} &= w_{11}\tilde{x_1} + w_{12}\tilde{x_2} + w_{13}\tilde{x_3} \\
&\quad Similarly\ we\ can\ calculate\ net_{h2},\ net_{h3}\ and\ net_{h4} \\
Z_k &= \sigma(net_{hk}) \\
&= \sigma(\sum_{i=1}^{3} w_{ki}\tilde{x_i})
\end{aligned}
$$

Now considering the hidden layer and the output layer:

$$
\begin{aligned}
for\ i = 1, ...3 \\
\hat{x_i} &= \sum_{k=1}^{4} w_{ik} z_k \\
\hat{y_j} &= \sum_{k=1}^{4} v_{jk} Z_k\ for\ j = 1,2
\end{aligned}
$$

Backpropagation updates for v_{jk}, consider v_{11}

$$
\begin{aligned}
L(y,\hat{y}) &= \frac{1}{2}((y_1 - \hat{y_1})^2 + (y_2 - \hat{y_2})^2) \\
\frac{\partial L(y,\hat{y})}{\partial v_{11}} &= \frac{\partial L(y,\hat{y})}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial v_{11}} \\
\frac{\partial L(y,\hat{y})}{\partial v_{11}} &= \frac{\partial L(y,\hat{y})}{\partial \hat{y_1}} * \frac{\partial \hat{y_1}}{\partial y_1} * \frac{\partial y_1}{\partial v_{11}} \\
&= (\hat{y_1} - y_1) * 1 * z_1 \\
&\quad Similarly\ , v_{ki} = z_i(\hat{y_k} - y_k)
\end{aligned}
$$

Backpropagation updates for w_{ki}, consider w_{11}

$$\frac{\partial E_{total}}{\partial w_{11}} = \frac{\partial E_{total}}{\partial Z_1} * \frac{\partial Z_1}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_{11}} \tag{1}$$

Now, output of hidden layer neuron contributes to the output of multiple output neurons, therefore $\frac{\partial E_{total}}{\partial Z_1}$ needs to take into consideration its effects on both the output neurons:

$$\begin{aligned}
\frac{\partial E_{total}}{\partial Z_1} &= \frac{\partial L(y,\hat{y})}{\partial Z_1} + \frac{\partial L(x,\hat{x})}{\partial Z_1} \\
&= \left[\frac{\partial L(y,\hat{y})}{\partial \hat{y_1}} * \frac{\partial \hat{y_1}}{\partial Z_1}\right] + \left[\frac{\partial L(y,\hat{y})}{\partial \hat{y_2}} * \frac{\partial \hat{y_2}}{\partial Z_1}\right] + \\
&\left[\frac{\partial L(x,\hat{x})}{\partial \hat{x_1}} * \frac{\partial \hat{x_1}}{\partial Z_1}\right] + \left[\frac{\partial L(x,\hat{x})}{\partial \hat{x_2}} * \frac{\partial \hat{x_2}}{\partial Z_1}\right] + \left[\frac{\partial L(x,\hat{x})}{\partial \hat{x_3}} * \frac{\partial \hat{x_3}}{\partial Z_1}\right] \\
&= (\hat{y_1} - y_1)v_{11} + (\hat{y_2} - y_2)v_{21} + (\hat{x_1} - x_1)w_{11} + (\hat{x_2} - x_2)w_{21} + (\hat{x_3} - x_3)w_{31}
\end{aligned}$$

Using the above result in equation 1, we get:

$$\frac{\partial E_{total}}{\partial w_{11}} = [(\hat{y_1} - y_1)v_{11} + (\hat{y_2} - y_2)v_{21} + (\hat{x_1} - x_1)w_{11} + (\hat{x_2} - x_2)w_{21} + (\hat{x_3} - x_3)w_{31}] *$$
$$Z_1(1 - Z_1)\tilde{x_1}$$

Similarly,

$$w_{ki} = [(\hat{y_1} - y_1)v_{1k} + (\hat{y_2} - y_2)v_{2k} + (\hat{x_1} - x_1)w_{k1} + (\hat{x_2} - x_2)w_{k2} + (\hat{x_3} - x_3)w_{k3}] *$$
$$Z_k(1 - Z_k)\tilde{x_i}$$

SOLUTION 2

Part 1.a Density function is given as:

$$f_X(x,\lambda) = \lambda e^{-\lambda x} \ , x >= 0$$

Maximum log-likelihood of $f_X(x,\lambda)$:

$$\begin{aligned} l(f_X(x,\lambda)) &= log\pi_{i=1}^{n}\left[\lambda e^{-\lambda[\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n} x_j]}\right] \\ &= \sum_{i=1}^{n} log\left[\lambda e^{-\lambda[\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n} x_j]}\right] \\ &= nlog\lambda - \lambda[\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n} x_j] \end{aligned}$$

Part 1.b EXPECTATION STEP: Let the unobserved variables $\sum_{j=r=1}^{n} x_j$ be represented by Z

$$\begin{aligned} Q(\lambda,\lambda^{old}) &= \sum_{i=1}^{n} log(l(\lambda|X,y))p(y|X,\lambda^{old}) \\ &= \sum_{y}^{n}[nlog\lambda - \lambda[\sum_{k=1}^{r} x_k + \sum_{j=r+1}^{n} x_j]]p(y_i|x_i,\lambda^{old}) \\ & for\ i = r+1, ...n\ y_i = c_i \end{aligned}$$

Part 1.c MAXIMIZATION STEP:
Differentiate with respect to λ and equate it to zero to find new value of λ :

$$
\begin{aligned}
\frac{\partial Q(\lambda,\lambda^{old})}{\partial \lambda} &= \frac{\partial}{\partial \lambda}\sum_{i=1}^{n}[nlog\lambda - \lambda[\sum_{k=1}^{r} x_k + \sum_{j=r+1}^{n} x_j]]p(y_i|x_i,\lambda^{old}) \\
0 &= \sum_{i=1}^{n}\left[\frac{n}{\lambda} - [\sum_{k=1}^{r} x_k + \sum_{j=r+1}^{n} x_j]\right]p(y_i|x_i,\lambda^{old})
\end{aligned}
$$

$After\ solving\ for\ \lambda$

$$
\lambda^{new} = \frac{n\sum_{i=1}^{n} p(y_i|x_i,\lambda^{old})}{\sum_{i=1}^{n}[\sum_{k=1}^{r} x_k + \sum_{j=r+1}^{n} x_j]p(y_i|x_i,\lambda^{old})}
$$

SOLUTION 3

Part 3.a:
Given the following expression:

$$\begin{aligned}
LHS &= \sum_{x \varepsilon X} ||x - s||^2 - \sum_{x \varepsilon X} ||x - \bar{x}||^2 \\
&= \sum_{i=1}^{|X|} (x_i - s_i)^2 - \sum_{i=1}^{|X|} (x_i - \bar{x}_i)^2 \\
&= \sum_{i=1}^{|X|} (x_i - s_i + x_i - \bar{x}_i)(x_i - s_i - x_i + \bar{x}_i) \\
&= \sum_{i=1}^{|X|} (\bar{x}_i - s_i)(\bar{x}_i - s_i - 2(\bar{x}_i - x_i)) \\
&= \sum_{i=1}^{|X|} (\bar{x}_i - s_i)^2 - 2\sum_{i=1}^{|X|} (\bar{x}_i - s_i)(\bar{x}_i - x_i)
\end{aligned}$$

Now inside the summation sign,

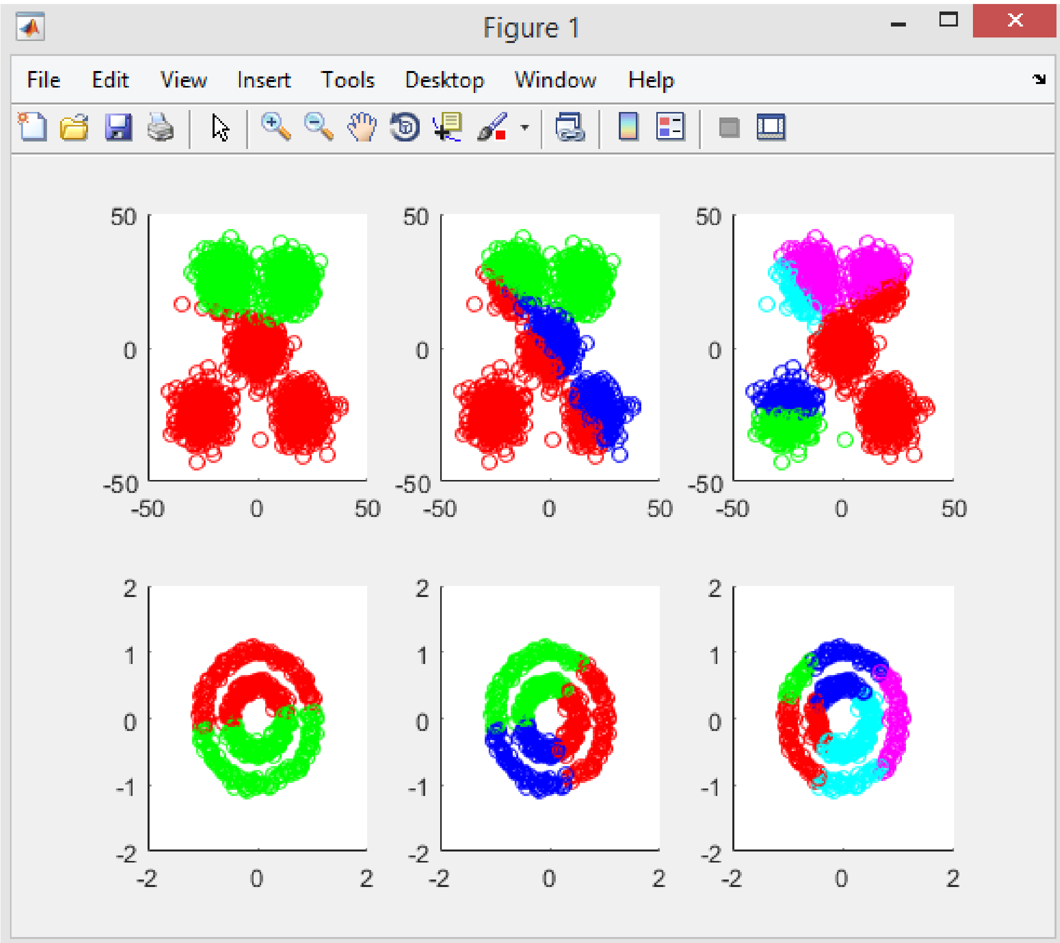
$$\bar{x}_i = x_i$$

$$So\ the\ second\ summation\ term\ becomes\ zero$$

$$LHS = |X|.||\bar{x} - s||^2$$

Part 3.b:
Efficiency: k-medoids is computaionally harder than K-means as computing the medoid is harder than cmputing the average.
Sensitivity to Outliers: K-Medoids method is more robust than k-means in the presence of noise and outliers becuase k-medoid is roughly comparable to the median and the median is more robust to outliers than the aritmetic mean. Also it minimizes a sum of pairwise dissimilarities instead of a sum of Euclidean distances.
Objective Function: K-medoid picks actual data points to represent clusters instead of mean values. The objective function minimizes the sum of the dissimilarities between each object and its corresponding reference point.

Q.4.1



It can be observed from the above plot that k-means algorithm does not separate the two circles. Since both the concentric circles have same mean and the k-means algorithm is based on the principal that different classes have different means, thus k-means fails to separate two circles. However, if the circles are mapped o higher dimensions, then they can be separated.

Q.4.2

C.

Best log likelihood values on heldout data:

K=3; **-4.6437+e03**

K=5; **-4.5858+e03**

k=7; **-4.6069+e03**

k=9; **-4.6094+e03**

k=11; **-4.6345+e03**

Best log likelihood values on training data:

K=3; **-1.7381e+03**

K=5; **-1.6734e+03**

k=7; **-1.6621e+03**

k=9; **-1.6504e+03**

k=11; **-1.6406e+03**

Iterations:

K=3; **669**

K=5; **318**

K=7; **2098**

K=9; **1512**

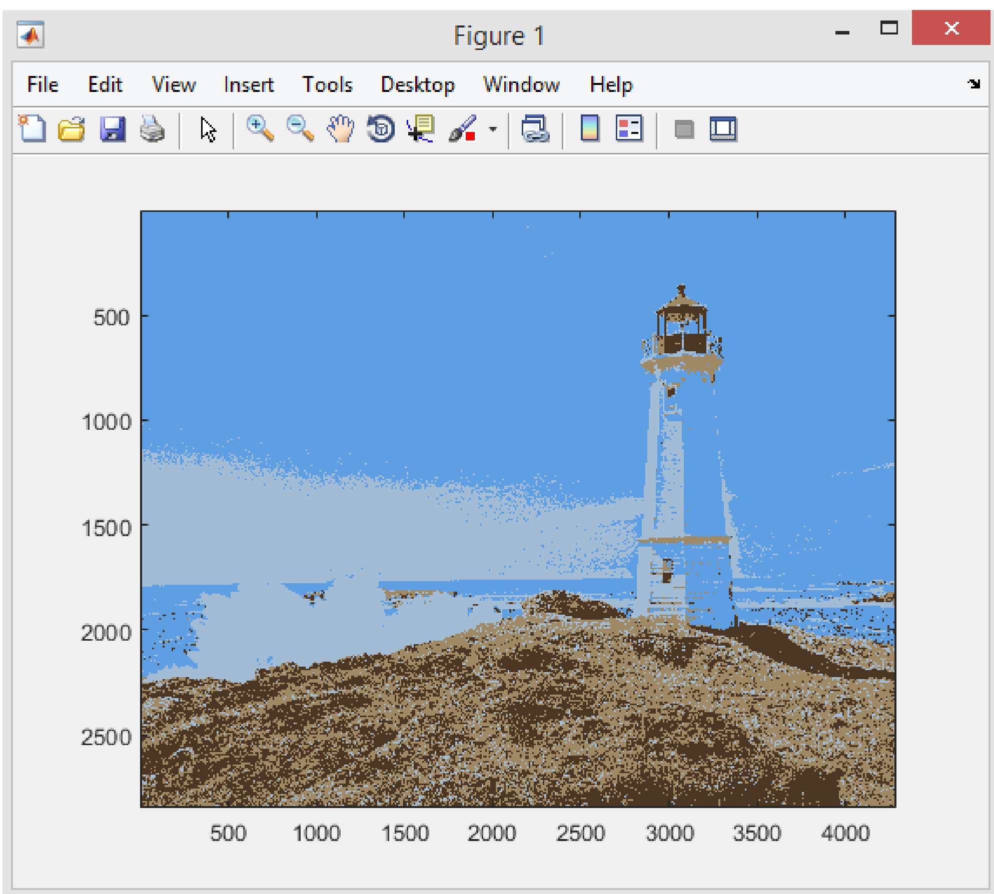
K=11; **1834**

D.

The maximum log likelihood among all the values of K is **-4.6437+e03 for k=3.** Thus, component k=3 should be selected. The value maybe different for other iterations as it depends on k-means algorithm.

Q.4.3

K=4



K=8

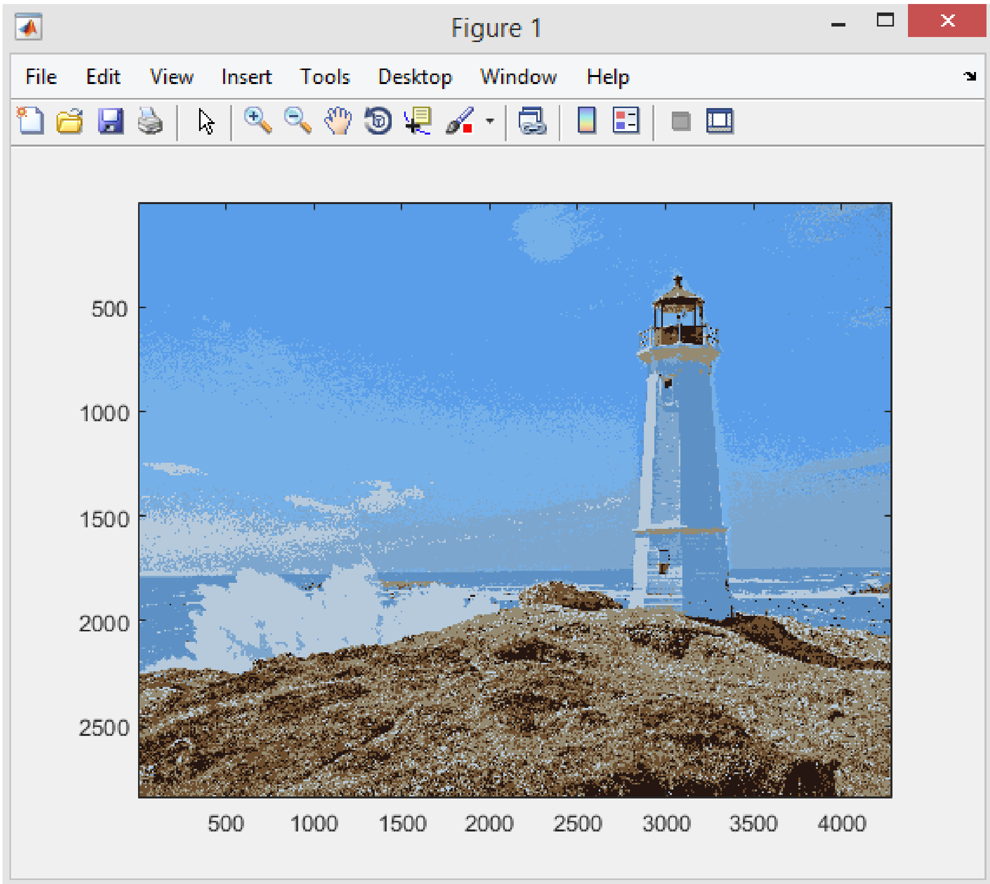


Figure 1
File Edit View Insert Tools Desktop Window Help
500
1000
1500
2000
2500
500 1000 1500 2000 2500 3000 3500 4000

K=24

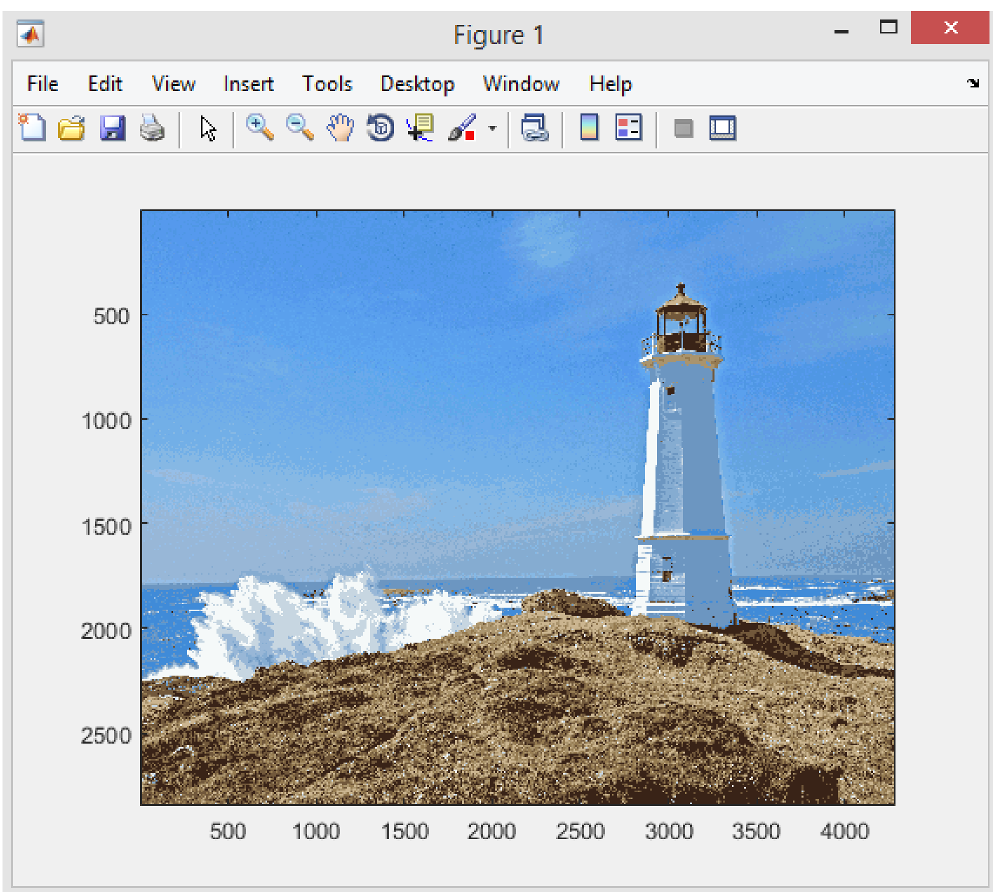


Figure 1
File Edit View Insert Tools Desktop Window Help
500
1000
1500
2000
2500
500 1000 1500 2000 2500 3000 3500 4000