ECE 403/503 Solutions to Assignment 2

Problems for this assignment: 1.5 and 1.6.

Problem 1.5 (6 points) With $K = 2$, $\mu_1^{(0)} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$, and $\mu_2^{(0)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, apply the K-means algorithm to cluster dataset

$$\mathcal{D} = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$$

(a) The first iteration starts with the initial centres $\{\mu_1^{(0)}, \mu_2^{(0)}\}$ and use (1.20) to perform Step 2 of the algorithm, namely, to compute $\{r_{n,k}\}$ for $n = 1, 2, \ldots, 6$ and $k = 1$ and 2. It is a good idea to prepare $\{r_{n,k}\}$ as a matrix of size 6×2, and complete the matrix row by row using Eq. (1.20). The matrix $\{r_{n,k}\}$ so completed can then be used together with Eq. (1.21) to identify which points are associated to which centre so as to specify two clusters C_1 and C_2. And this completes Step 3. Next, use (1.22) to update the centres to $\{\mu_1^{(1)}, \mu_2^{(1)}\}$, and this completes the first iteration. Report the updated clusters C_k for $k = 1$ and 2, and your numerical results of $\{\mu_1^{(1)}, \mu_2^{(1)}\}$.

Solution

For initial centres $\{\mu_1^{(0)}, \mu_2^{(0)}\}$, matrix $\{r_{n,k}\}$ is obtained as follows. There were two occasions, namely points x_2 and x_4, where it was a tie for the distances between a data point and the two centres. In a case like this, a centre was selected at random. For points x_2, and x_4, $\mu_1^{(0)}$ $\mu_2^{(0)}$ were selected respectively.

	$\mu_1^{(0)} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$	$\mu_2^{(0)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$
$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	0	1
$x_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$	1	0
$x_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$	0	1
$x_4 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$	0	1
$x_5 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$	1	0
$x_6 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$	0	1

Above table defines two clusters $C_1 = \{x_2, x_5\}$ and $C_2 = \{x_1, x_3, x_4, x_6\}$ which were used to update the two centres to

$$\boldsymbol{\mu}_1^{(1)} = \frac{1}{2}(\boldsymbol{x}_2 + \boldsymbol{x}_5) = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix}, \quad \boldsymbol{\mu}_2^{(1)} = \frac{1}{4}(\boldsymbol{x}_1 + \boldsymbol{x}_3 + \boldsymbol{x}_4 + \boldsymbol{x}_6) = \begin{bmatrix} 1 \\ 0.75 \end{bmatrix}$$

■

(b) Repeat the process in part (a) and report the updated clusters C_k for $k = 1$ and 2, and your numerical results of $\{\boldsymbol{\mu}_1^{(2)}, \boldsymbol{\mu}_2^{(2)}\}$.

Solution

With centres $\{\boldsymbol{\mu}_1^{(1)}, \boldsymbol{\mu}_2^{(1)}\}$, matrix $\{r_{n,k}\}$ is updated to:

	$\boldsymbol{\mu}_1^{(1)} = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix}$	$\boldsymbol{\mu}_2^{(1)} = \begin{bmatrix} 1 \\ 0.75 \end{bmatrix}$
$\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	0	1
$\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$	0	1
$\boldsymbol{x}_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$	1	0
$\boldsymbol{x}_4 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$	0	1
$\boldsymbol{x}_5 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$	1	0
$\boldsymbol{x}_6 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$	0	1

Above table defines two clusters $C_1 = \{\boldsymbol{x}_3, \boldsymbol{x}_5\}$ and $C_2 = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_6\}$ which were used to update the two centres to

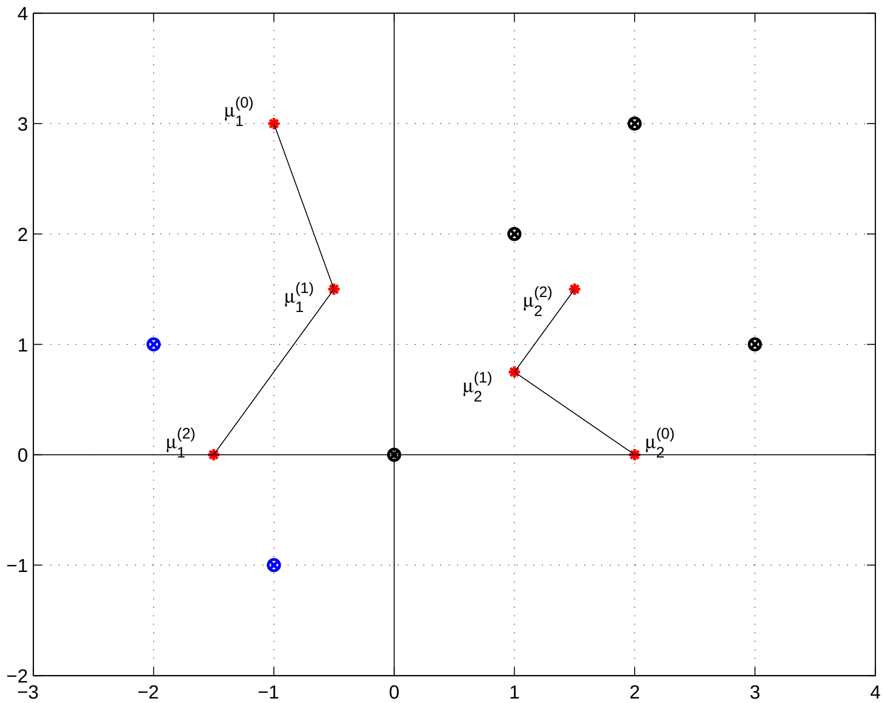
$$\boldsymbol{\mu}_1^{(2)} = \frac{1}{2}(\boldsymbol{x}_3 + \boldsymbol{x}_5) = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2^{(2)} = \frac{1}{4}(\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_4 + \boldsymbol{x}_6) = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$$

■

(c) Display in a figure the locations of the six data points as well as the two centres $\{\boldsymbol{\mu}_1^{(2)}, \boldsymbol{\mu}_2^{(2)}\}$ with two colors (one for each cluster and its centre). Compare the initial centres and clusters with those after two iterations and comment on the results you obtained.

Solution

The plot below depicts the two clusters C_1 (dots in blue) and C_2 (dots in black) obtained in part (b). Also shown are the trajectories of the two centres (stars in red) moving from their initial positions to $\{\boldsymbol{\mu}_1^{(2)}, \boldsymbol{\mu}_2^{(2)}\}$.



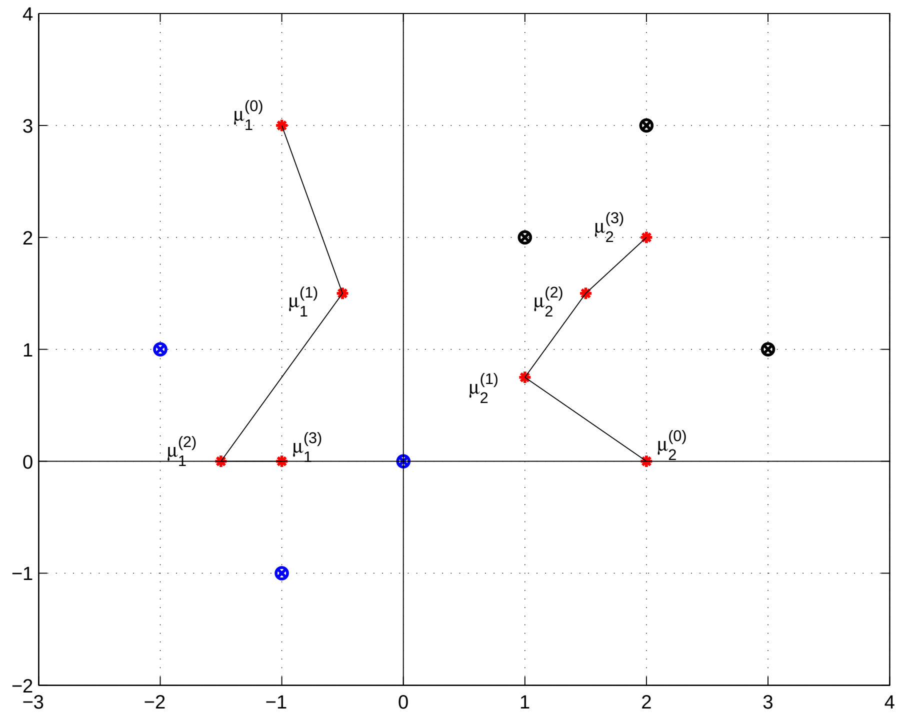
As expected the clustering process works as the sum of squared distances from data points to their respective centres was reduced from the initial value 35 to 22.25 after one iteration, and to 10.25 after two iterations. Furthermore, centres $\left\{\boldsymbol{\mu}_1^{(2)}, \boldsymbol{\mu}_2^{(2)}\right\}$ can be used to revise matrix $\{r_{n,k}\}$ as

	$\boldsymbol{\mu}_1^{(2)} = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix}$	$\boldsymbol{\mu}_2^{(2)} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$
$\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	1	0
$\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$	0	1
$\boldsymbol{x}_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$	1	0
$\boldsymbol{x}_4 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$	0	1
$\boldsymbol{x}_5 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$	1	0
$\boldsymbol{x}_6 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$	0	1

Above table defines two clusters $C_1 = \{\boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{x}_5\}$ and $C_2 = \{\boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_6\}$ which update the two centres to

$$\boldsymbol{\mu}_1^{(3)} = \frac{1}{3}(\boldsymbol{x}_1 + \boldsymbol{x}_3 + \boldsymbol{x}_5) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \ \boldsymbol{\mu}_2^{(3)} = \frac{1}{3}(\boldsymbol{x}_2 + \boldsymbol{x}_4 + \boldsymbol{x}_6) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

With this the sum of squared distances was further reduced to 8. The figure below indicates that the K-mean algorithm has converged as additional iterations will not reduce the objective function further.



We remark that in the first iteration of the algorithm there are four possible random selections of the centres for points x_2 and x_4, and what we did above was one of them. We tried the other three selections and found the following:

(1) Both x_2 and x_4 selected $\boldsymbol{\mu}_2^{(0)}$ as centre, it then took the algorithm three iterations to converge to the same global solution as the one presented above.

(2) Both x_2 and x_4 selected $\boldsymbol{\mu}_1^{(0)}$ as centre, then the algorithm converged to a *suboptimal* solution with $C_1 = \{x_2, x_4, x_5\}$ and $C_1 = \{x_1, x_3, x_6\}$ after one iteration.

(3) Points x_2 and x_4 selected $\boldsymbol{\mu}_2^{(0)}$ and $\boldsymbol{\mu}_1^{(0)}$, respectively, as their centres. Then the algorithm converged to the same suboptimal solution as in case (2) after two iterations. ■

Problem 1.6 (4 points) Repeat Example 1.10, but this time the linear regression method is applied to the classification of the second and third data classes, namely Iris Versicolor and Iris Virginica. As in Example 1.10, in each class 40 samples are selected at random for training, leaving the rest 10 examples for testing. To complete the problem, it is required to compute optimal parameters $\boldsymbol{w}^*$ and b^* for the regression model (1.35) and use the 20 test samples to evaluate the classification performance in terms of a confusion matrix. Report the numerical values of $\boldsymbol{w}^*$, b^*, and the confusion matrix.

To ensure consistent results, use the steps below to prepare the dataset:

1. Download matrix **`D_iris.mat`** from the course web site.
2. In MATLAB, use **`D = D_iris(1:4,51:150)`** to keep only the second (Versicolor) and third (Virginica) data classes.
3. Use the code below to prepare 80 samples (40 samples from each class) for training and 20 samples (10 samples from each class) for testing:

```
D1 = D(:,1:50);
D2 = D(:,51:100);
rand('state',15)
r1 = randperm(50);
D1train = D1(:,r1(1:40));
D1test = D1(:,r1(41:50));
rand('state',16)
r2 = randperm(50);
D2train = D2(:,r2(1:40));
D2test = D2(:,r2(41:50));
Dtrain = [D1train D2train];
Dtest = [D1test D2test];
```

NOTE: Here **`Dtrain`** includes 80 samples for training, with the first 40 samples from iris versicolor and the rest 40 samples from iris virginica; **`Dtest`** includes 20 samples for testing, with the first 10 samples from iris versicolor and the rest 10 samples from iris virginica.

Solution

Running the above code, data sets **`Dtrain`** and **`Dtest`** were constructed. Matrix $\hat{\boldsymbol{X}}$ and vector $\boldsymbol{y}$ required by (1.34) are produced as

```
Xh = [Dtrain' ones(80,1)];
y = [ones(40,1); -ones(40,1)];
```

and optimal model parameters $\boldsymbol{w}^*$ and b^* are computed as

```
wh = (Xh'*Xh)\(Xh'*y);
ws = wh(1:4);
bs = wh(5);
```

which yield

$$\boldsymbol{w}^* = \begin{bmatrix} 0.278912236023776 \\ 0.744705475210416 \\ -0.679989327564937 \\ -1.414391405604721 \end{bmatrix} \text{ and } b^* = 1.828548464919417$$

Binary classifier (1.37) was applied to classify the 20 test samples. All except the 11$^{\text{th}}$ sample (an iris virginica misclassified to class iris versicolor) was classified correctly. Hence the confusion matrix is given by

$$\boldsymbol{C} = \begin{bmatrix} 10 & 1 \\ 0 & 9 \end{bmatrix}$$

The figure below depicts the inner products $w^{*T} x + b^{*}$ for the 20 test samples, where the positive valued 11th inner product was highlighted in red as it was supposed to be negative valued. ■