Homework #3 solution key

Prayag
Neural networks and learning systems-I

April 23, 2019

Problem 5.4.

Solution. Cost function $J(\mu)$ is given by

$$J(\mu) = \sum_{j=1}^{K} \sum_{i=1}^{N} w_{ij} \|\mathbf{x_i} - \mu_\mathbf{j}\|^2 \tag{1}$$

where the weighting factor w_{ij} is defined as

$$w_{ij} = \begin{cases} 1 & \text{if the data point } \mathbf{x_i} \text{ lies in cluster } j \\ 0 & \text{otherwise} \end{cases}$$

We differentiate Equation (1) with respect to $\mu_\mathbf{j}$ to get

$$2 \sum_{i=1}^{N} w_{ij}(\mathbf{x_i} - \mu_\mathbf{j}) = 0$$

$$\sum_{i=1}^{N} w_{ij}\mu_\mathbf{j} = \sum_{i=1}^{N} w_{ij}\mathbf{x_i}$$

$$\mu_\mathbf{j} = \frac{\sum_{i=1}^{N} w_{ij}\mathbf{x_i}}{\sum_{i=1}^{N} w_{ij}}, \quad j = 1, 2, \ldots, K$$

The numerator is the sum of all data points present in cluster k. The denominator in the above expression gives the number of data points assigned to the cluster k. Therefore, it can be interpreted that for each cluster j, the $\mu_\mathbf{j}$ is set to the mean of all the data points $\mathbf{x_i}$ assigned to cluster j. ■

Problem 5.10.

Solution. K = 6 clusters
Training sample = 1000 data points

(a) The 6 clusters obtained using K-means clustering for varying D values are as given in Figure 1. It is observed that the position of the cluster centers depends on the initial conditions. The cluster mean and variance obtained for each value of separation is given in Table 1 and 2.

Table 1: Cluster means for various values of separation d.

	d=1		**d=0**		**d=-1**		**d=-2**		**d=-3**		**d=-4**		**d=-5**		**d=-6**	
	x	**y**	**x**	**y**	**x**	**y**	**x**	**y**	**x**	**y**	**x**	**y**	**x**	**y**	**x**	**y**
Cluster 1	1.2	-5.6	10.1	-10.1	-8.8	3.3	18.6	-2.5	1.2	-1.8	-7.8	5.1	-8.8	4.5	11.2	-3.3
Cluster 2	8.2	5.3	8.8	4.7	4.1	9	0.2	9.9	9.4	-6.8	19.5	0.7	10	-4.6	18.8	1.5
Cluster 3	9.5	-10.8	1.7	-5.3	9.6	3.4	-8.5	4.8	-1.2	9.9	3.1	-2.7	2	-0.6	-0.2	9.2
Cluster 4	-8.4	5.5	-8.9	4.7	3.3	-5.6	1.1	-2.6	18	-2.2	0.8	9.7	18.7	0.2	-8.4	5.2
Cluster 5	18.4	-6.2	18.1	-5.3	-3.4	9.5	9.6	-7.9	-8.9	4.5	8.2	5	8.4	5.3	2.7	-0.8
Cluster 6	-0.6	9.8	-0.1	9.6	16.5	-5.7	8.7	4.6	8	5.4	12.9	-5.5	-0.6	8.7	8.4	5.4

(b) It is observed that as the separation d reduces, the clusters formed are obtained have data points from both the moons. The cluster means come closer as d reduces.

(c) Figure 2 shows the variance of each cluster for different values of the separation d.

(d) Table 2 shows the comparison between the cluster variance and variance obtained from calculation. It is observed that the common variance computed decreases as the separating distance reduces.

■

Problem 3.

Solution. (a) Performed (Kmeans, LMS) and (Kmeans, RLS) algorithm for the data for d=-5. Generated 1000 data points for training the network. Tested the algorithm with 2000 data points. Performed Kmeans with $K = 20$. Obtained the means and calculated the variance. Used the mean values to calculate the kernel output. The weights were

Table 2: Cluster variance for various values of separation d.

	d=1	**d=0**	**d=-1**	**d=-2**	**d=-3**	**d=-4**	**d=-5**	**d=-6**
Cluster 1	10.3	8.79	9.32	4.74	21.29	9.42	15.29	7.64
Cluster 2	9.9	9.82	11.51	7.04	16.74	11.02	4.82	7.08
Cluster 3	7.55	10.98	8.92	4.17	5.19	8.59	7.66	8.25
Cluster 4	8.32	12.29	8.81	22.45	5.16	9.34	7.42	7.8
Cluster 5	11.56	9.33	10.26	6.94	5.38	8.48	9.51	8.36
Cluster 6	8.05	8.02	7.98	18.87	6.9	8.66	15.16	7.17
D_max	29.06	29.36	28.32	26.98	26.84	26.89	26.95	27.69
σ^2	8.39	8.48	8.18	7.79	7.75	7.76	7.78	7.99

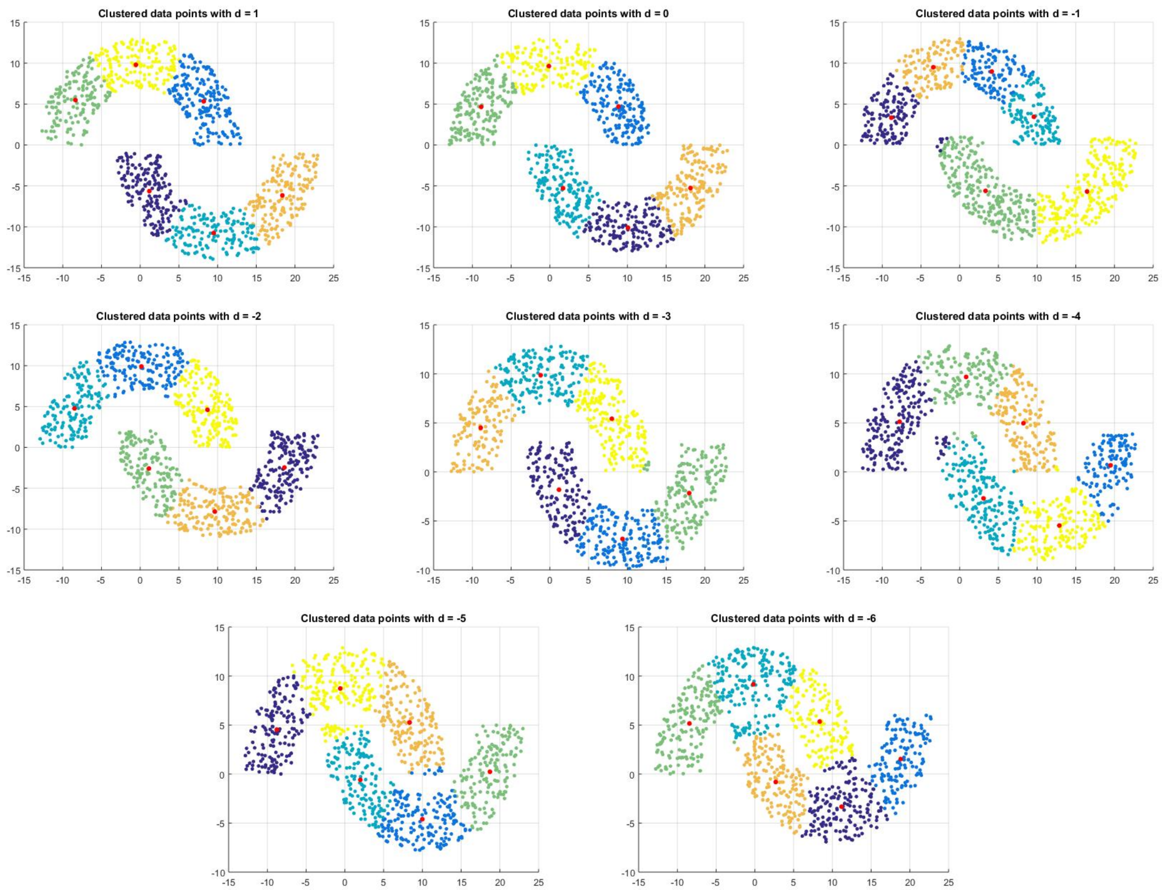
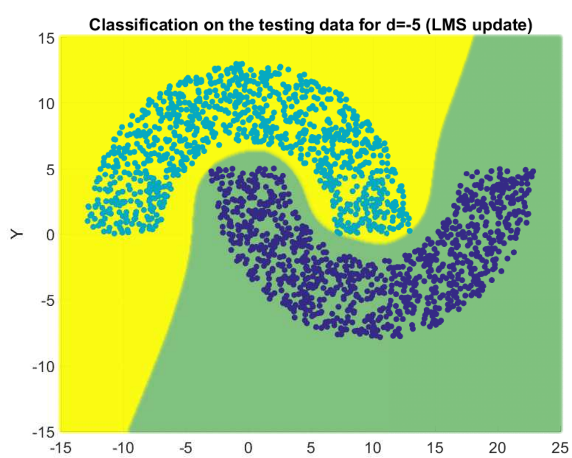
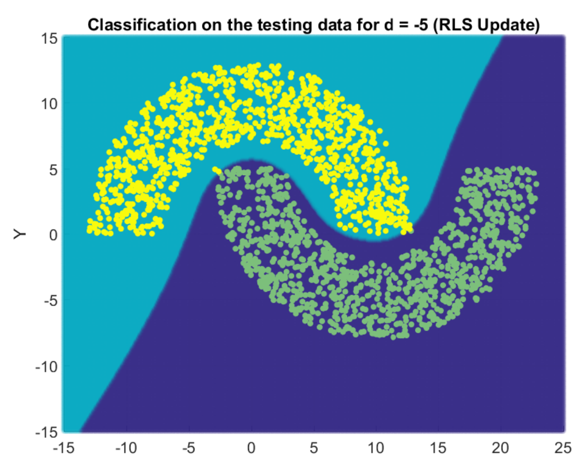


Figure 1: Clustering of the data points for D = 1 to -6.

trained using both the algorithms and the testing results with the decision boundary are as given below.





(b) Performed (Kmeans, LMS) and (Kmeans, RLS) algorithm for the data for d=-6. Generated 1000 data points for training the network. Tested the algorithm with 2000 data

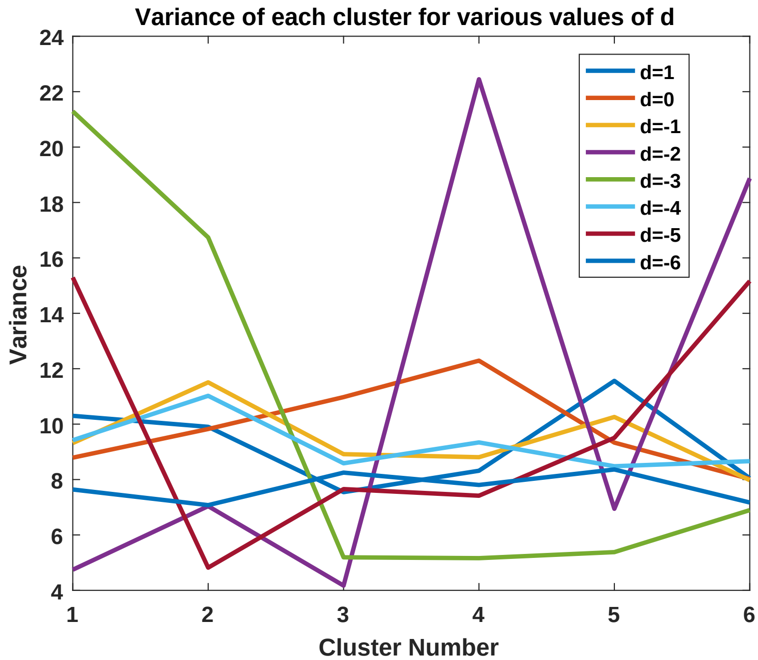
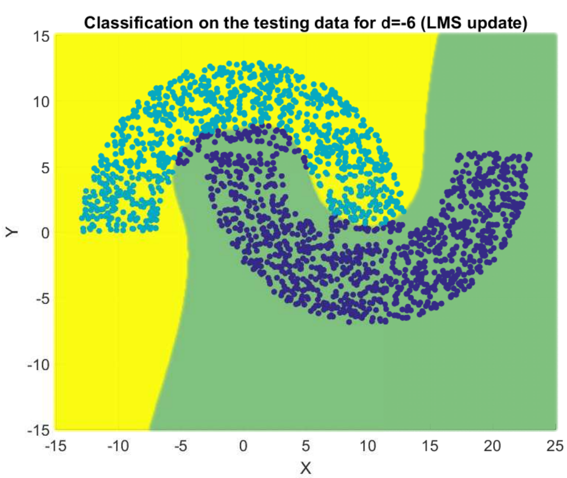
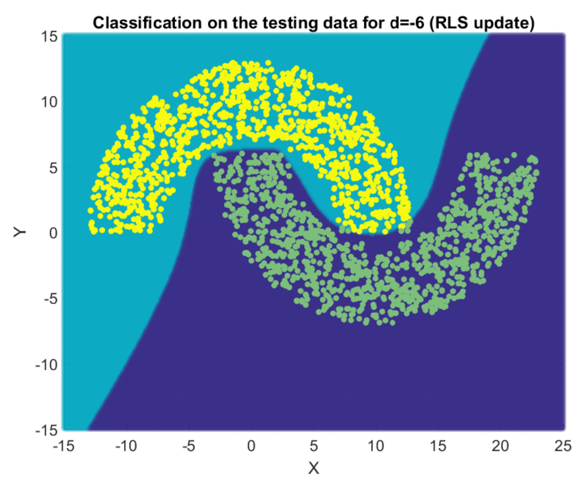


Figure 2: Plot of variance of each cluster for various values of d

points. Performed Kmeans with $K = 20$. Obtained the means and calculated the variance. Used the mean values to calculate the kernel output. The weights were trained using both the algorithms and the testing results with the decision boundary are as given below.





(c) The comparison results of both the algorithms is given in Table below.

It is observed that as the separation between the moons increases, the algorithms are able to classify the data better.

■

Table 3: Classification accuracy for both the methods for both the separation values

	$d = -5$	$d = -6$
LMS Update	100%	94.90%
RLS Update	99.80%	99.50%