1. (True/False) While the EM algorithm maintains uncertainty about the cluster assignment for each observation via soft assignments, the model assumes that every observation comes from only one cluster. 1 point

- (●) True
- ( ) False

2. (True/False) In high dimensions, the EM algorithm runs the risk of setting cluster variances to zero. 1 point

- (●) True
- ( ) False

3. In the EM algorithm, what do the E step and M step represent, respectively? 1 point

- (●) Estimate cluster responsibilities, Maximize likelihood over parameters
- ( ) Estimate likelihood over parameters, Maximize cluster responsibilities
- ( ) Estimate number of parameters, Maximize likelihood over parameters
- ( ) Estimate likelihood over parameters, Maximize number of parameters

4. 1 point

Suppose we have data that come from a mixture of 6 Gaussians (i.e., that is the true data structure). Which model would we expect to have the highest log-likelihood after fitting via the EM algorithm?

- ( ) A mixture of Gaussians with 2 component clusters
- ( ) A mixture of Gaussians with 4 component clusters
- ( ) A mixture of Gaussians with 6 component clusters
- ( ) A mixture of Gaussians with 7 component clusters
- (●) A mixture of Gaussians with 10 component clusters

5. 1 point

Which of the following correctlydescribes the differences between EM for mixtures of Gaussians and k-means? Choose all that apply.

- [ ] k-means often gets stuck in a local minimum, while EM tends not to
- [x] EM is better at capturing clusters of different sizes and orientations
- [ ] EM is better at capturing clusters with overlaps
- [ ] EM is less prone to overfitting than k-means

- [x] k-means is equivalent to running EM with infinitesimally small diagonal covariances.

6. Suppose we are running the EM algorithm. After an E-step, we obtain the following responsibility matrix: 1 point

Cluster responsibilities	Cluster A	Cluster B	Cluster C
Data point 1	0.20	0.40	0.40
Data point 2	0.50	0.10	0.40
Data point 3	0.70	0.20	0.10

Which is the most probable cluster for data point 3?

- (•) Cluster A
- ( ) Cluster B
- ( ) Cluster C

7. Suppose we are running the EM algorithm. After an E-step, we obtain the following responsibility matrix: 1 point

Cluster responsibilities	Cluster A	Cluster B	Cluster C
Data point 1	0.20	0.40	0.40
Data point 2	0.50	0.10	0.40
Data point 3	0.70	0.20	0.10

Suppose also that the data points are as follows:

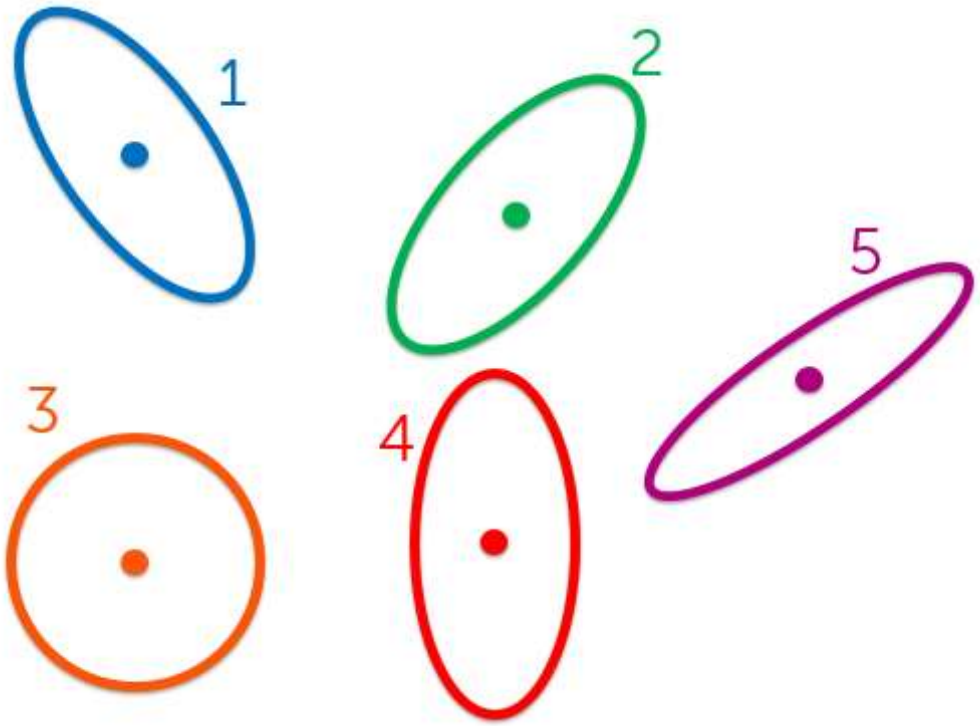
Dataset	X	Y	Z
Data point 1	3	1	2
Data point 2	0	0	3
Data point 3	1	3	7

Let us compute the new mean for Cluster A. What is the Z coordinate of the new mean? Round your answer to 3 decimal places.

4.86

8. 1 point

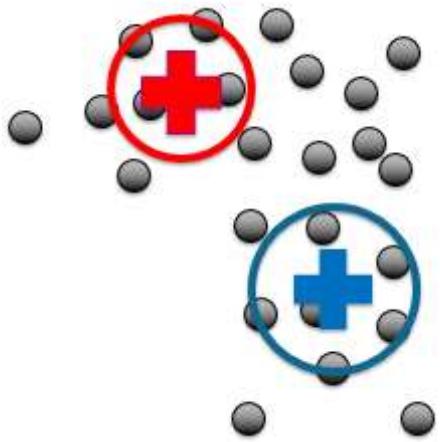
Which of the following contour plots describes a Gaussian distribution with diagonal covariance? Choose all that apply.



- [ ] (1)
- [ ] (2)
- [x] (3)
- [x] (4)
- [ ] (5)

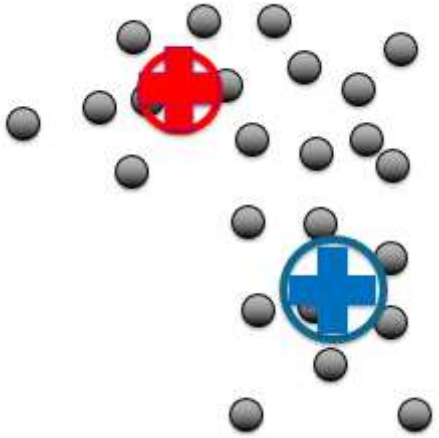
9. 2 points

Suppose we initialize EM for mixtures of Gaussians (using full covariance matrices) with the following clusters:

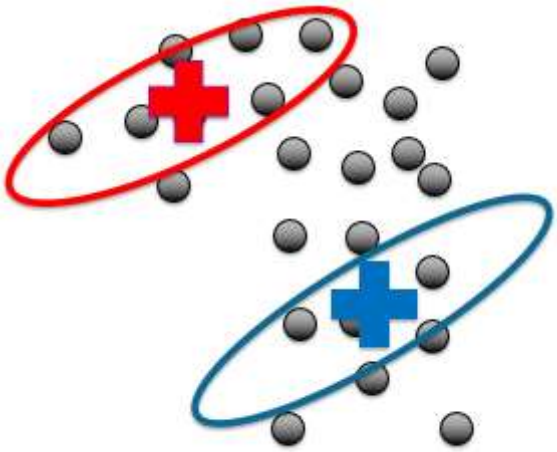


Which of the following best describes the updated clusters after the first iteration of EM?

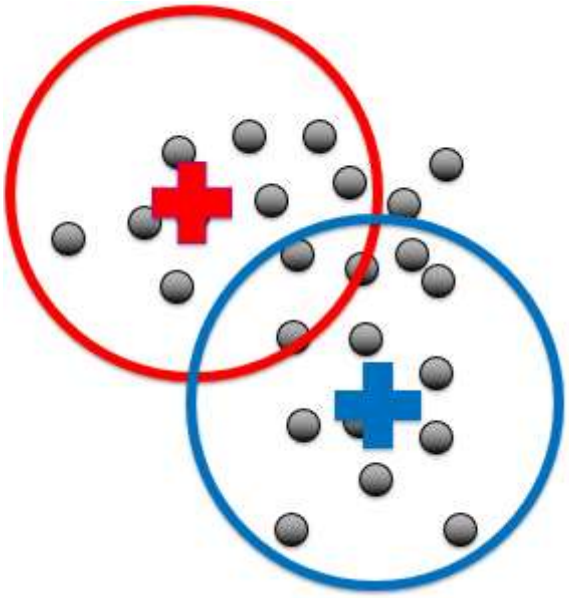
○



○



○



◉