

Week 2 Quiz

[Help Center](#)

The **due date** for this quiz is **Sun 10 May 2015 9:55 PM PDT**.

- ☐ In accordance with the Coursera Honor Code, I (Sandipan Dey) certify that the answers here are my own work.

Question 1

Given Eps and Minpts, if a point p is density-connected to a point q, which of the following statements are correct?

- ☐ Point p is directly density-reachable from q
- ☐ Point q is density-connected to p
- ☐ Point q is directly density-reachable from p
- ☐ Point p is density-reachable from q
- ☐ Point q is density-reachable from p

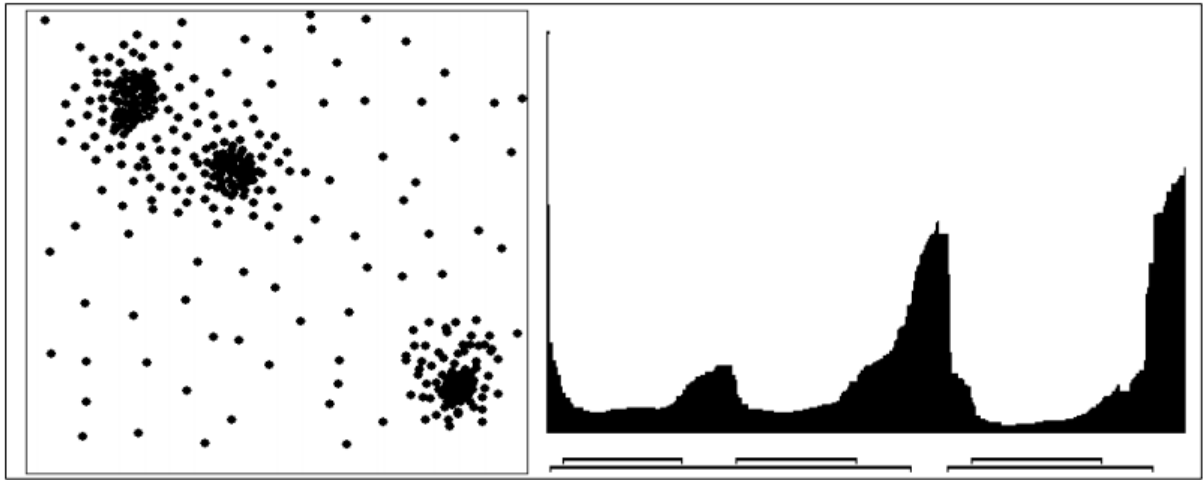
Question 2

Suppose Eps = 1cm and Minpts = 1. Randomly select two points p, q from the observed data points. We have $\text{dist}(p, q) = 4\text{cm}$. Which of the following statements are correct?

- ☐ Points p and q may be in the different clusters
- ☐ Points p and q must be in different clusters
- ☐ Points p and q may not belong to any clusters
- ☐ Points p and q must be in the same cluster

Question 3

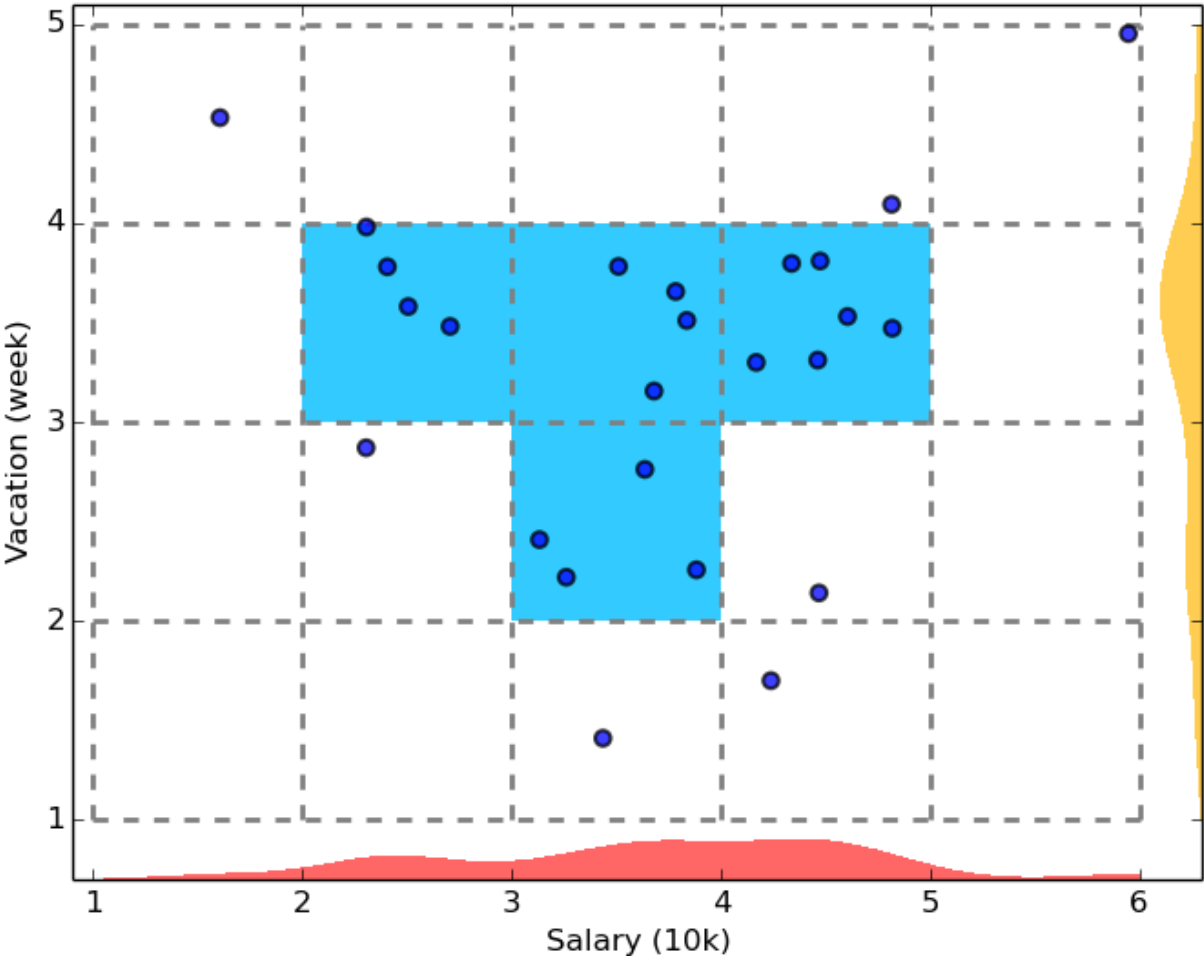
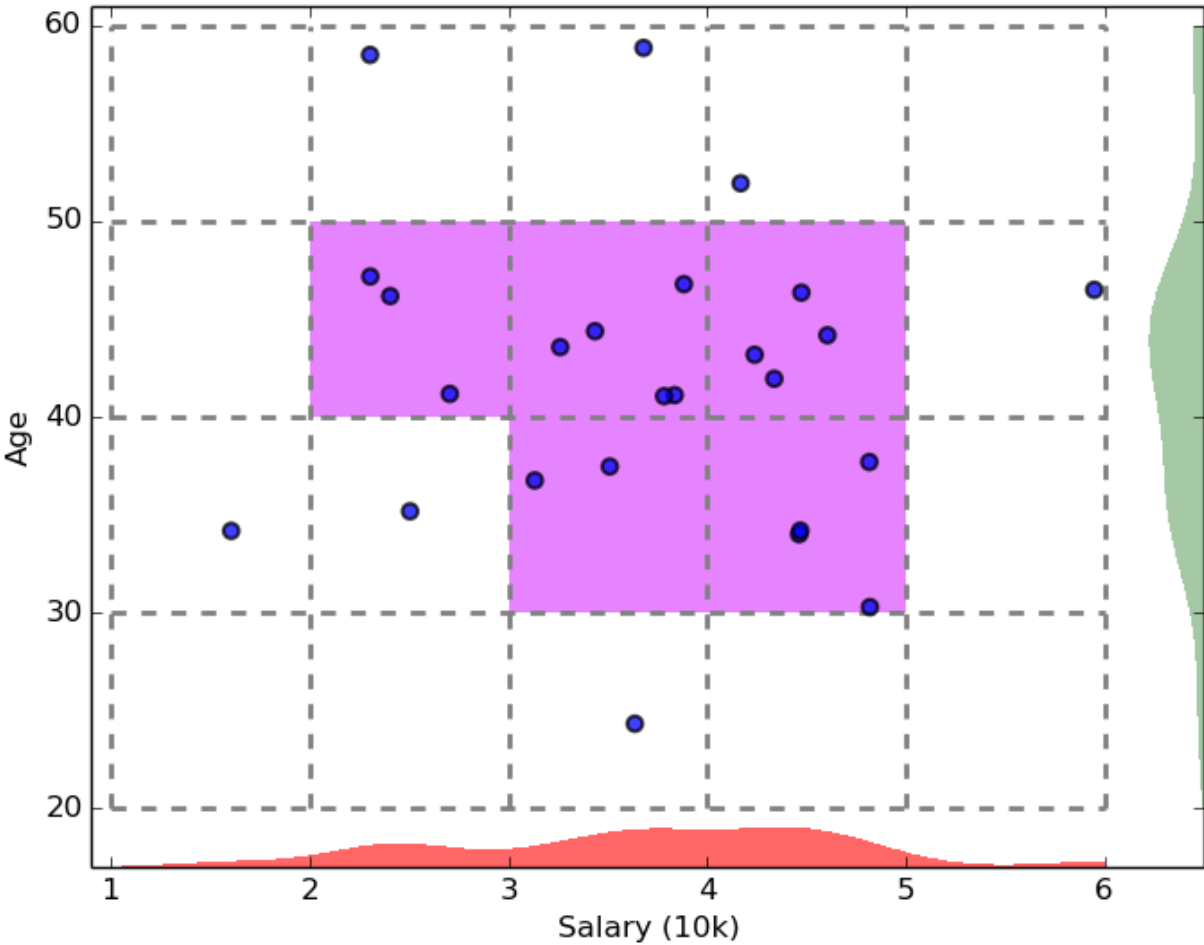
Given the following synthetic data set (left) and the reachability-plot (right), how many clusters might there be by setting different reachability thresholds?

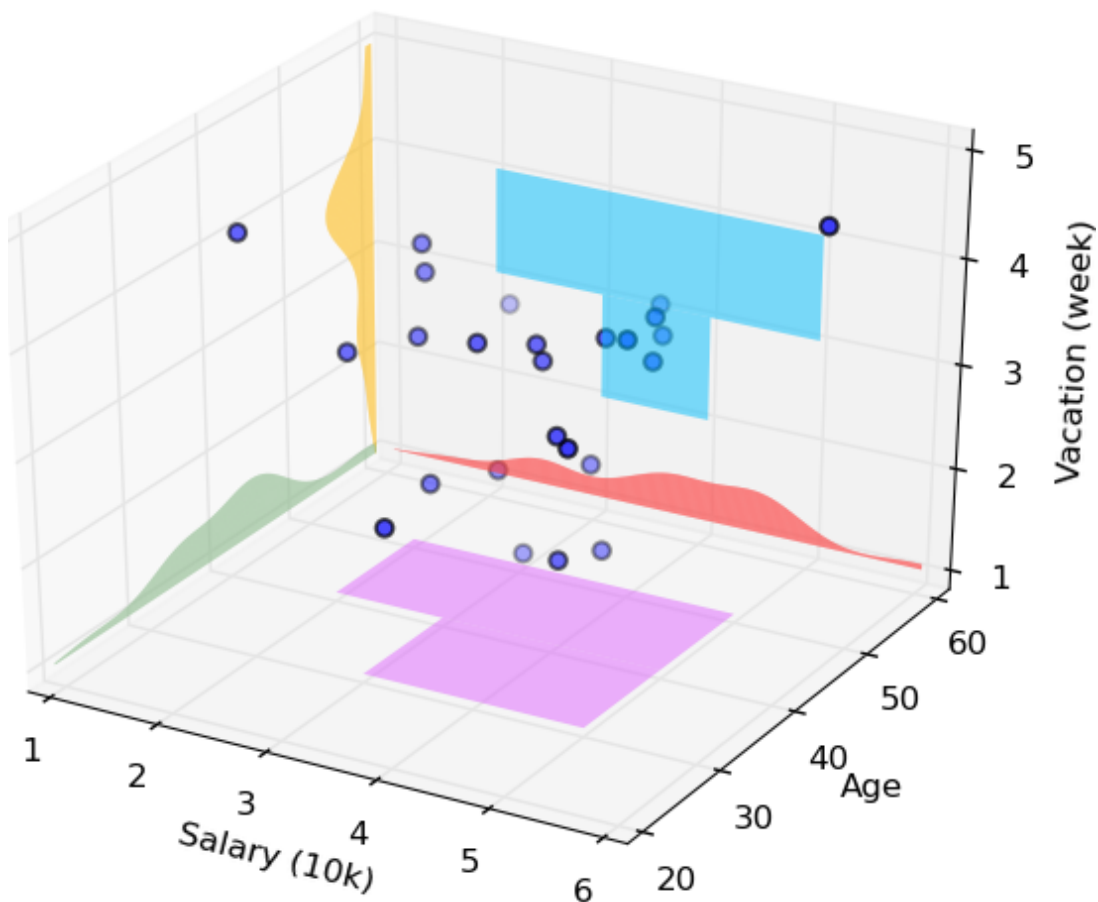


- ☐ 2
- ☐ 1
- ☐ more than 3
- ☐ 3

Question 4

Apply the CLIQUE clustering algorithm to the following depicted dataset. Suppose we have found the dense region in 3-D space. Which of the following statements are correct?





- ☐ When projecting the dense region in 3D into the 2-D space of Salary-Age, the projected area may not be a subspace of the purple region.
- ☐ When projecting the dense region in 3D into the 2-D space of Vacation-Age, the projected area must be a subspace of the purple region.
- ☐ When projecting the dense region in 3D into the 2-D space of Vacation -Age, the projected area may not be a subspace of the purple region.
- ☐ When projecting the dense region in 3D into the 2-D space of Salary-Age, the projected area must be a subspace of the purple region.

Question 5

In Gaussian Mixture Model, considering the single dimension case, for i-th cluster,

$$P(x|C_i = N(\mu_i, \sigma_i)) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}.$$

Suppose we have learned two clusters from some dataset with the parameter $\mu_1=0, \sigma_1=1, \mu_2=-2, \sigma_2=3, P(C_1$

$P(C_1) = 0.25, P(C_2) = 0.75$. For the new data point $x=1$, what is the probability that it belongs to the first cluster, i.e. $P(C_1 | x)$?

- ☐ 0.75
- ☐ 0.8
- ☐ 0.5
- ☐ 0.25

Question 6

Which of the following statements about Gaussian Mixture Models (GMM) are correct?

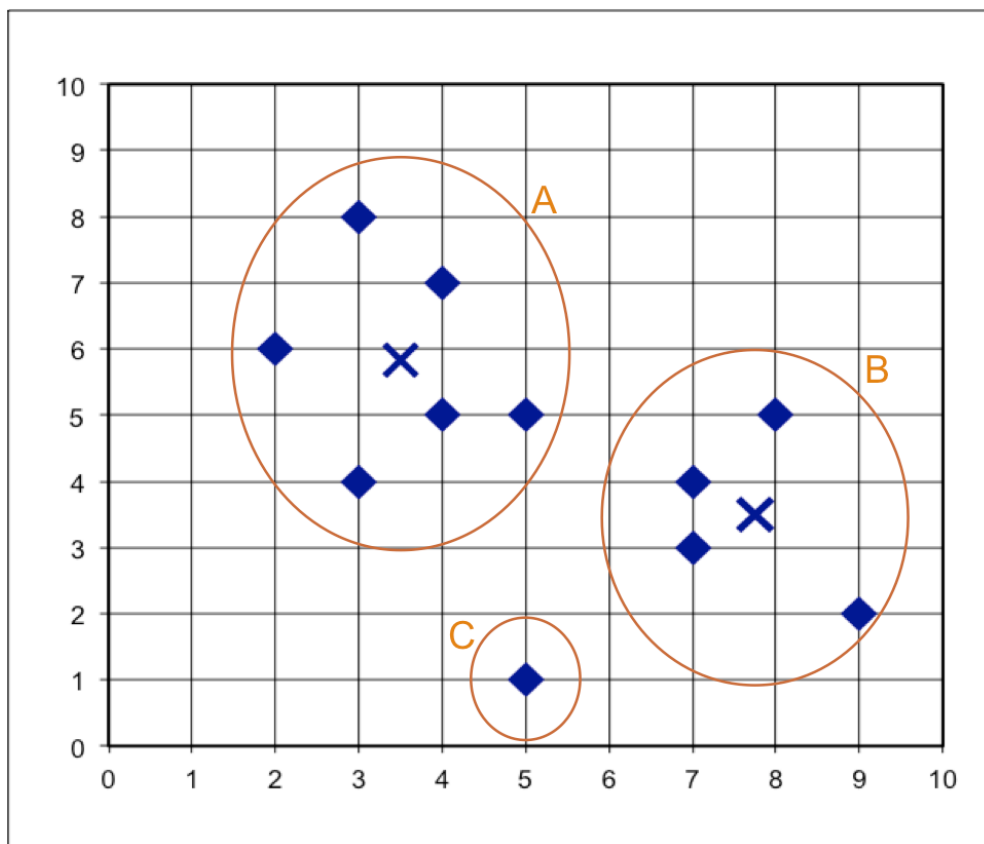
- ☐ GMM assumes the data points are generated by some Gaussian distributions.
- ☐ GMM usually converges much faster than K-means.
- ☐ In GMM, for different initializations, we will definitely have exactly same clustering result.
- ☐ In GMM, we have to specify the number of clusters.

Question 7

In which ONE of the following scenarios may Gaussian Mixture Models (GMM) not work properly?

- ☐ Build 2 clusters from 2000 10-dimensional data points randomly generated from 10 Gaussian distributions.
- ☐ Build 10 clusters from 2000 10-dimensional data points randomly generated from 10 Gaussian distributions.

Question 8



Consider the three clusters A, B, and C shown in Figure 1. Using Euclidean distance as the similarity measure, which two clusters would be merged first in agglomerative clustering using single link?

- ☐ A and B
- ☐ B and C
- ☐ A and C
- ☐ All three options above are tied

Question 9

Recall from Lecture 4-8 that the objective of learning generative models is to find the parameters that maximize the likelihood of the observed data. Suppose we have a set of points D drawn from Gaussian distribution. For $D = \{-1, 0, 1\}$, which of the following set of parameters (μ, σ) produces the maximum $L(N(\mu, \sigma^2); D)$?

- ☐ $\mu = 0, \sigma = 2$
- ☐ $\mu = 1, \sigma = 0$
- ☐ $\mu = 2, \sigma = 0$
- ☐ $\mu = 0, \sigma = 1$

Question 10

Consider the three hierarchical clustering algorithms introduced in Lecture 4, BIRCH, CURE, and CHAMELEON. Which of the following statements about these algorithms is TRUE?

(Select all that apply)

- ☐ CHAMELEON requires a graph as the input.
- ☐ All three algorithms can only work with Euclidean distance as the similarity metric.
- ☐ BIRCH and CHAMELEON both use a divisive method to partition the objects into small groups first before merging them back to form the final clusters.
- ☐ CHAMELEON and CURE are better at capturing irregular shaped clusters than BIRCH

- ☐ In accordance with the Coursera Honor Code, I (Sandipan Dey) certify that the answers here are my own work.

Submit Answers

Save Answers

You cannot submit your work until you agree to the Honor Code. Thanks!