Unsupervised Continuous Clustering

Machine Learning: Jordan Boyd-Graber
University of Colorado Boulder
LECTURE 16

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

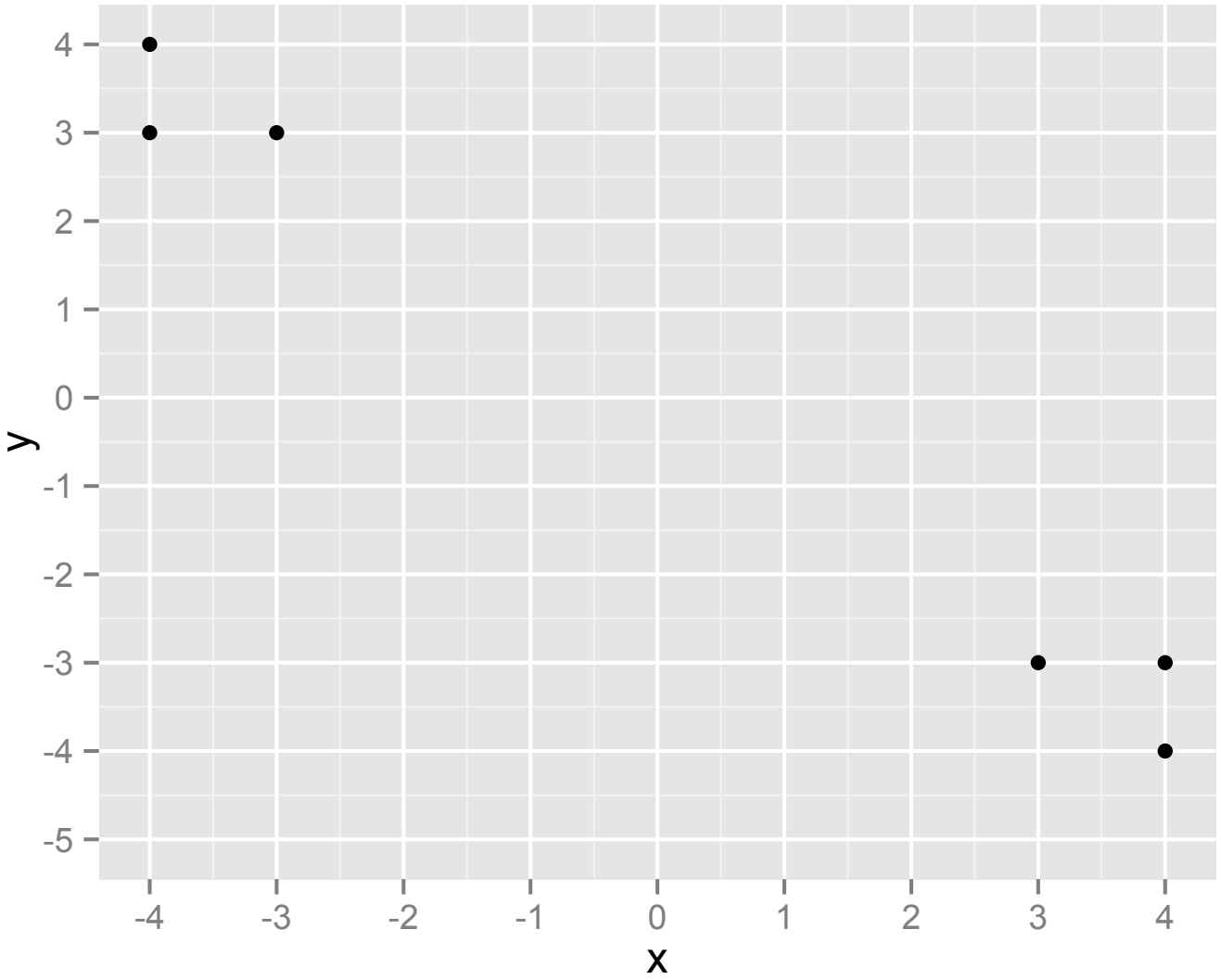
Content Questions

Content Questions

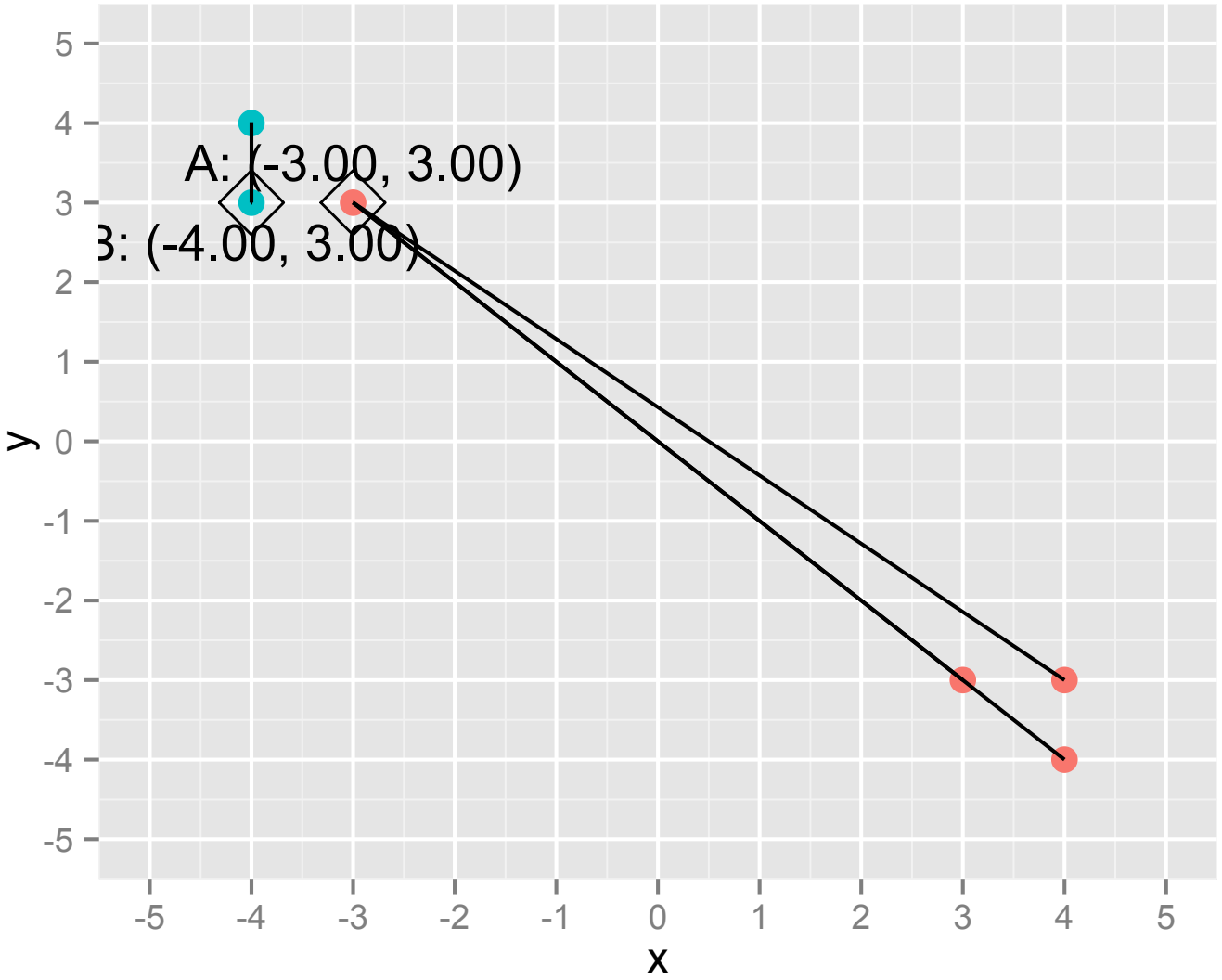
Administrivia

- Midterm grades: worry below 150
- Project proposals: watch on Piazza
- One more homework left

Two Points



Two Points



Two Points

$$\mu_A = \frac{1}{4}\left((-3,3) + (3,-3) + (4,-3) + (4,-4)\right)$$

$$=$$

$$\mu_B = \frac{(-4,3) + (-4,4)}{2}$$

$$=$$

Two Points

$$\mu_A = \frac{1}{4}\left((-3,3) + (3,-3) + (4,-3) + (4,-4)\right)$$

$$= (2, -1.75)$$

$$\mu_B = \frac{(-4,3) + (-4,4)}{2}$$

$$=$$

Two Points

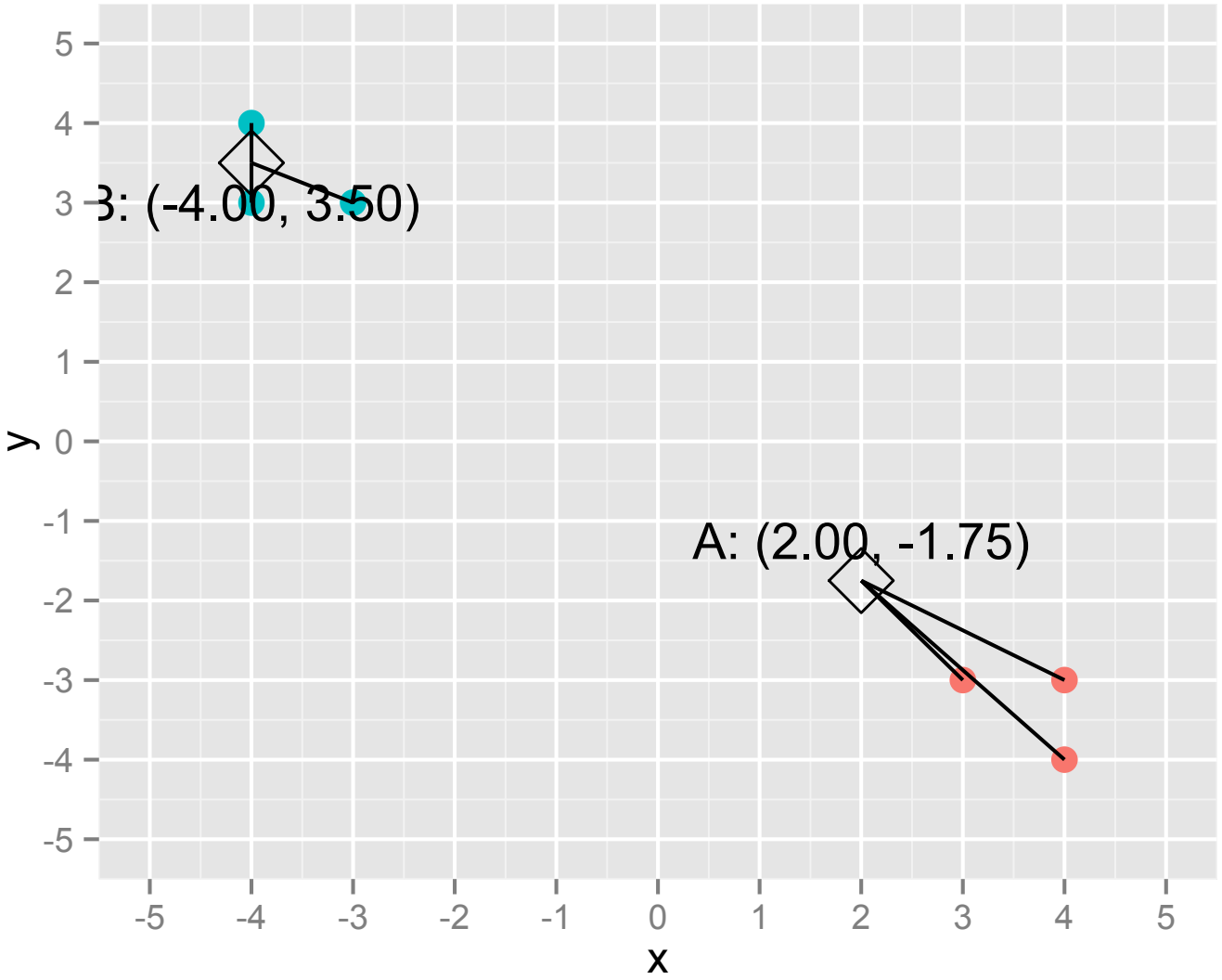
$$\mu_A = \frac{1}{4}\left((-3,3) + (3,-3) + (4,-3) + (4,-4)\right)$$

$$= (2, -1.75)$$

$$\mu_B = \frac{(-4,3) + (-4,4)}{2}$$

$$= (-4, 3.5)$$

Two Points



Two Points

$$\mu_A = \frac{(3, -3) + (4, -3) + (4, -4)}{3}$$

$$=$$

$$\mu_B = \frac{(-4, 3) + (-4, 4) + (-3, 3)}{3}$$

$$=$$

Two Points

$$\mu_A = \frac{(3, -3) + (4, -3) + (4, -4)}{3}$$

$$= (3.67, -3.33)$$

$$\mu_B = \frac{(-4, 3) + (-4, 4) + (-3, 3)}{3}$$

$$=$$

Two Points

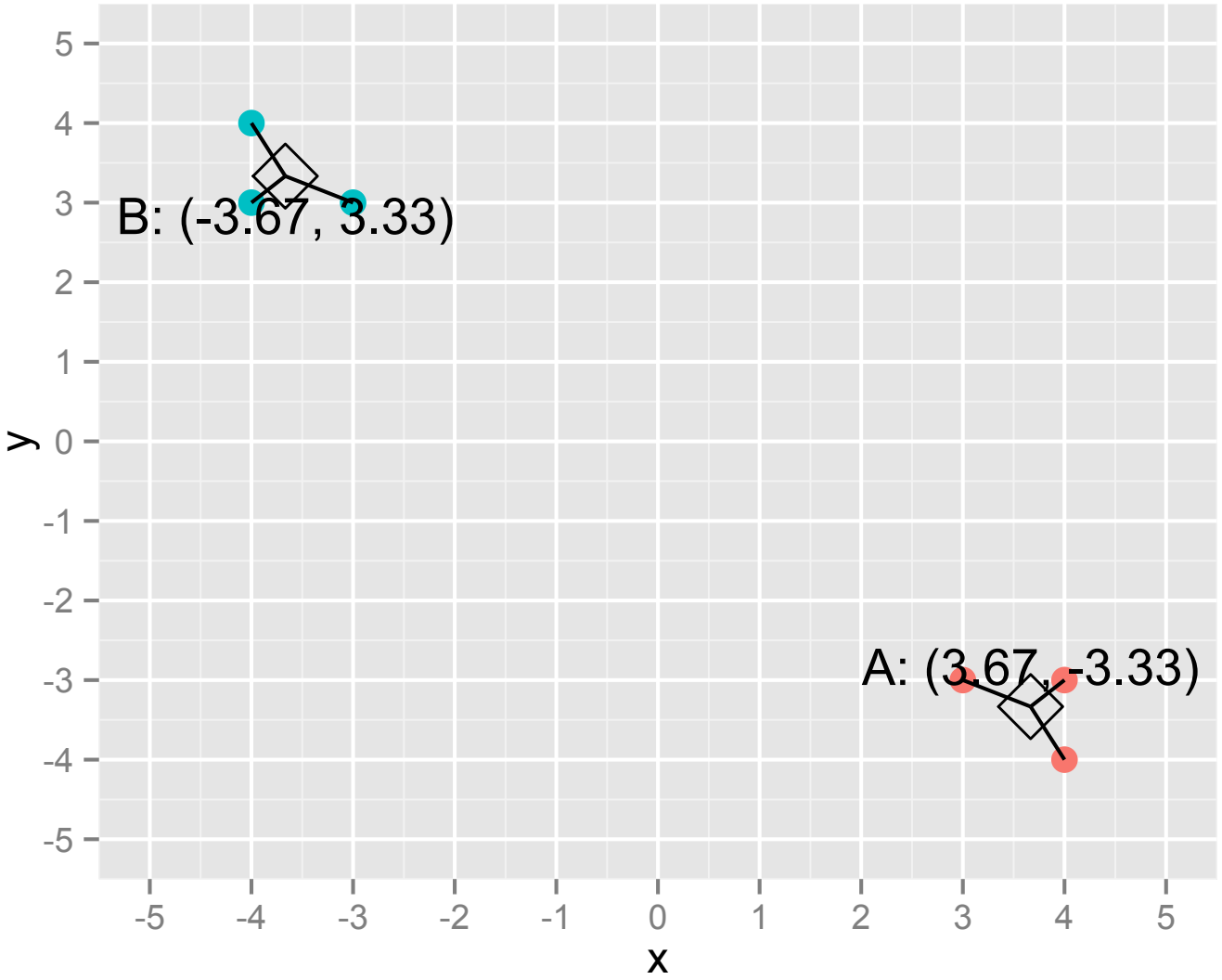
$$\mu_A = \frac{(3, -3) + (4, -3) + (4, -4)}{3}$$

$$= (3.67, -3.33)$$

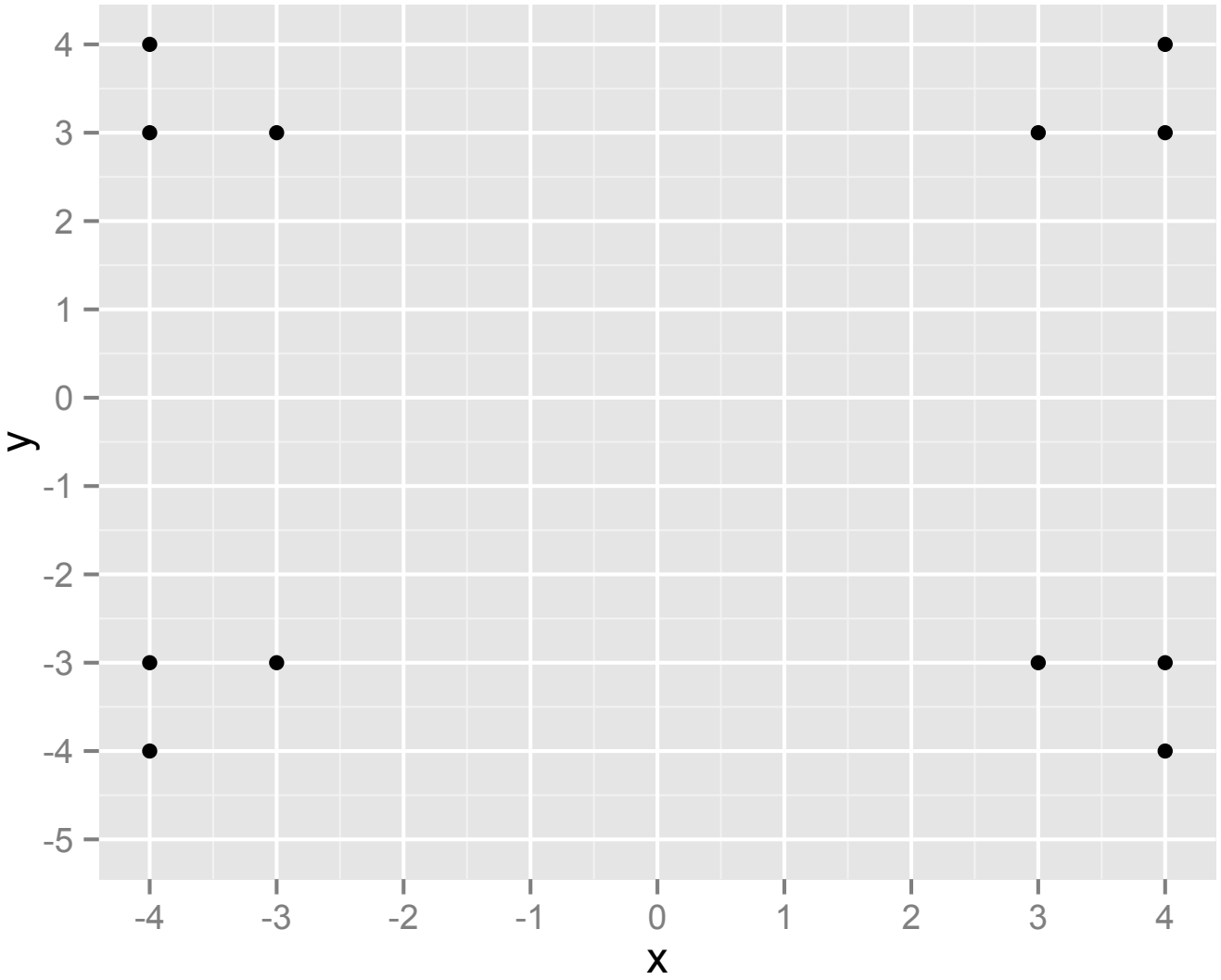
$$\mu_B = \frac{(-4, 3) + (-4, 4) + (-3, 3)}{3}$$

$$= (-3.67, 3.33)$$

Two Points



Four Points



Four Points

The observation at $(3, 3)$ is the same distance from μ_A and μ_C. If you look at Line 10 in the algorithm, the **first** mean with the smallest distance gets the assignment. So $(3, 3)$ gets assigned to cluster A.

$$\mu_A =$$

$$\mu_B =$$

$$\mu_C =$$

$$\mu_D =$$

Four Points

The observation at $(3, 3)$ is the same distance from μ_A and μ_C. If you look at Line 10 in the algorithm, the **first** mean with the smallest distance gets the assignment. So $(3, 3)$ gets assigned to cluster A.

$$\mu_A = (-1, 1)$$
$$\mu_B =$$
$$\mu_C =$$
$$\mu_D =$$

Four Points

The observation at $(3, 3)$ is the same distance from μ_A and μ_C. If you look at Line 10 in the algorithm, the **first** mean with the smallest distance gets the assignment. So $(3, 3)$ gets assigned to cluster A.

$$\mu_A = (-1, 1)$$
$$\mu_B = (-4, 0)$$
$$\mu_C =$$
$$\mu_D =$$

Four Points

The observation at $(3, 3)$ is the same distance from μ_A and μ_C. If you look at Line 10 in the algorithm, the **first** mean with the smallest distance gets the assignment. So $(3, 3)$ gets assigned to cluster A.

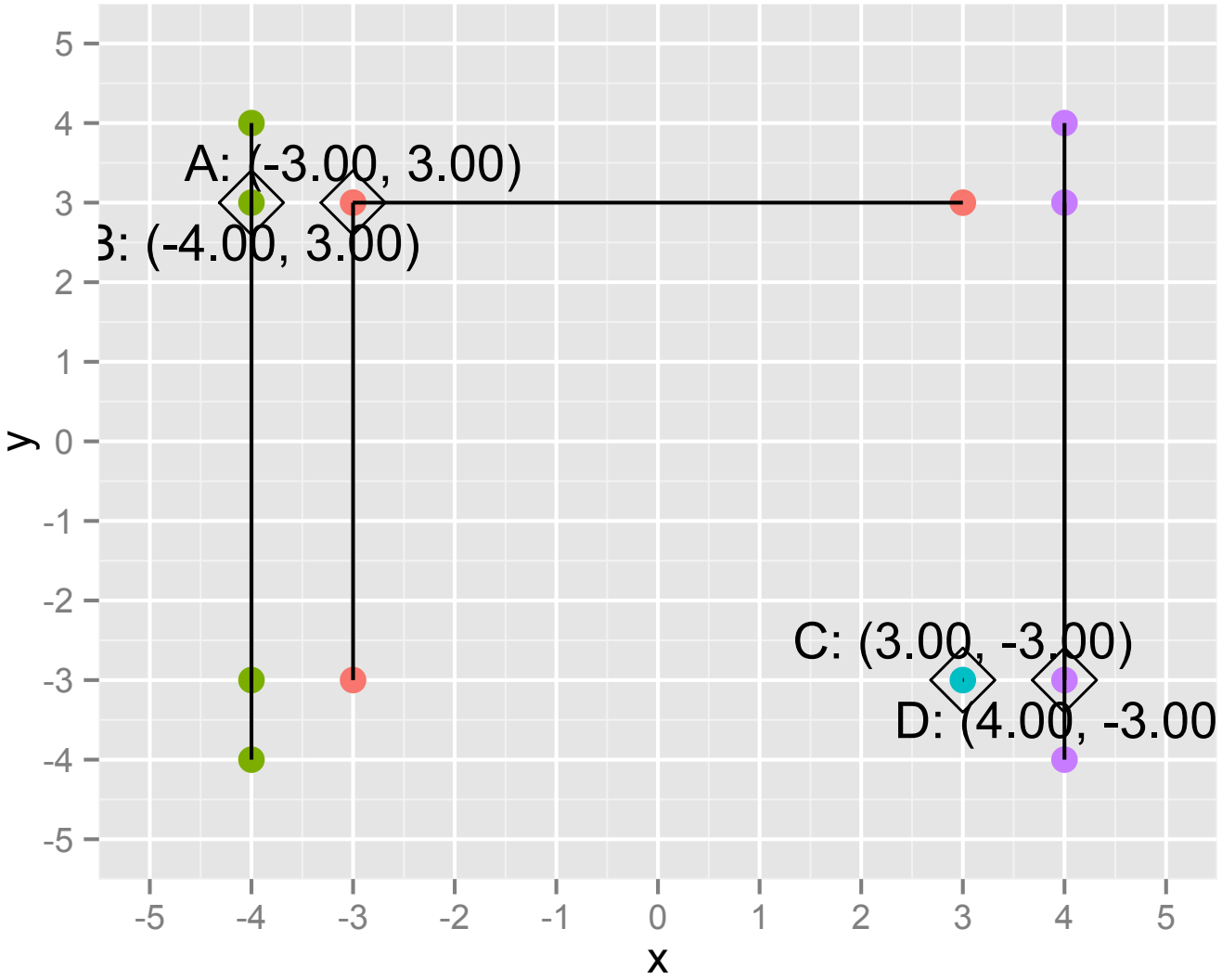
$$\mu_A = (-1, 1)$$
$$\mu_B = (-4, 0)$$
$$\mu_C = (3, -3)$$
$$\mu_D =$$

Four Points

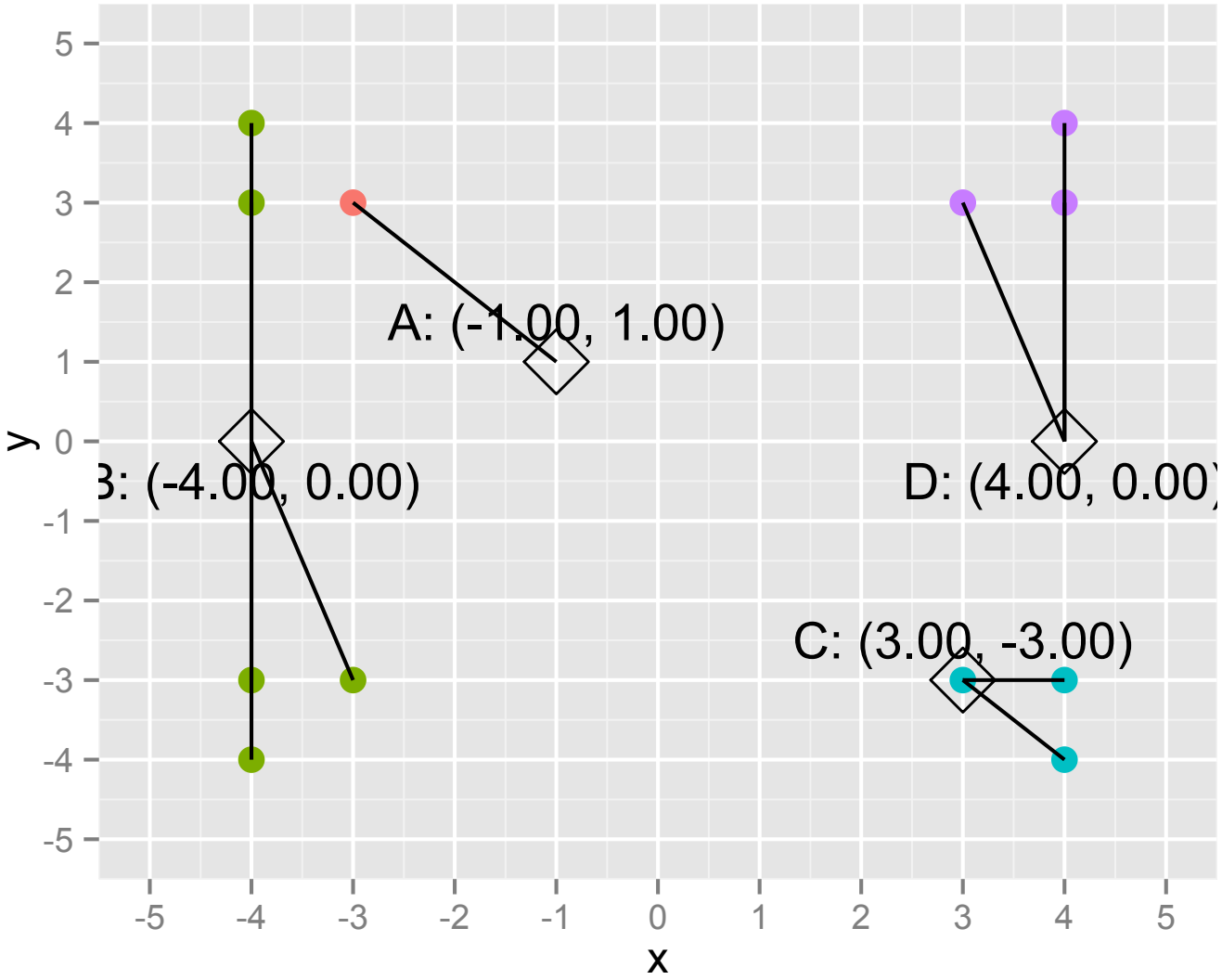
The observation at $(3, 3)$ is the same distance from μ_A and μ_C. If you look at Line 10 in the algorithm, the **first** mean with the smallest distance gets the assignment. So $(3, 3)$ gets assigned to cluster A.

$$\mu_A = (-1, 1)$$
$$\mu_B = (-4, 0)$$
$$\mu_C = (3, -3)$$
$$\mu_D = (4, 0)$$

Four Points


A: (-3.00, 3.00)
B: (-4.00, 3.00)
C: (3.00, -3.00)
D: (4.00, -3.00
x
y
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

Four Points


A: (-1.00, 1.00)
3: (-4.00, 0.00)
D: (4.00, 0.00)
C: (3.00, -3.00)
y
x
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

Four Points

$$\mu_A =$$

$$\mu_B =$$

$$\mu_C =$$

$$\mu_D =$$

Four Points

$$\mu_A = (-3, 3)$$

$$\mu_B =$$

$$\mu_C =$$

$$\mu_D =$$

Four Points

$$\mu_A = (-3, 3)$$
$$\mu_B = (-3.8, -0.6)$$
$$\mu_C =$$
$$\mu_D =$$

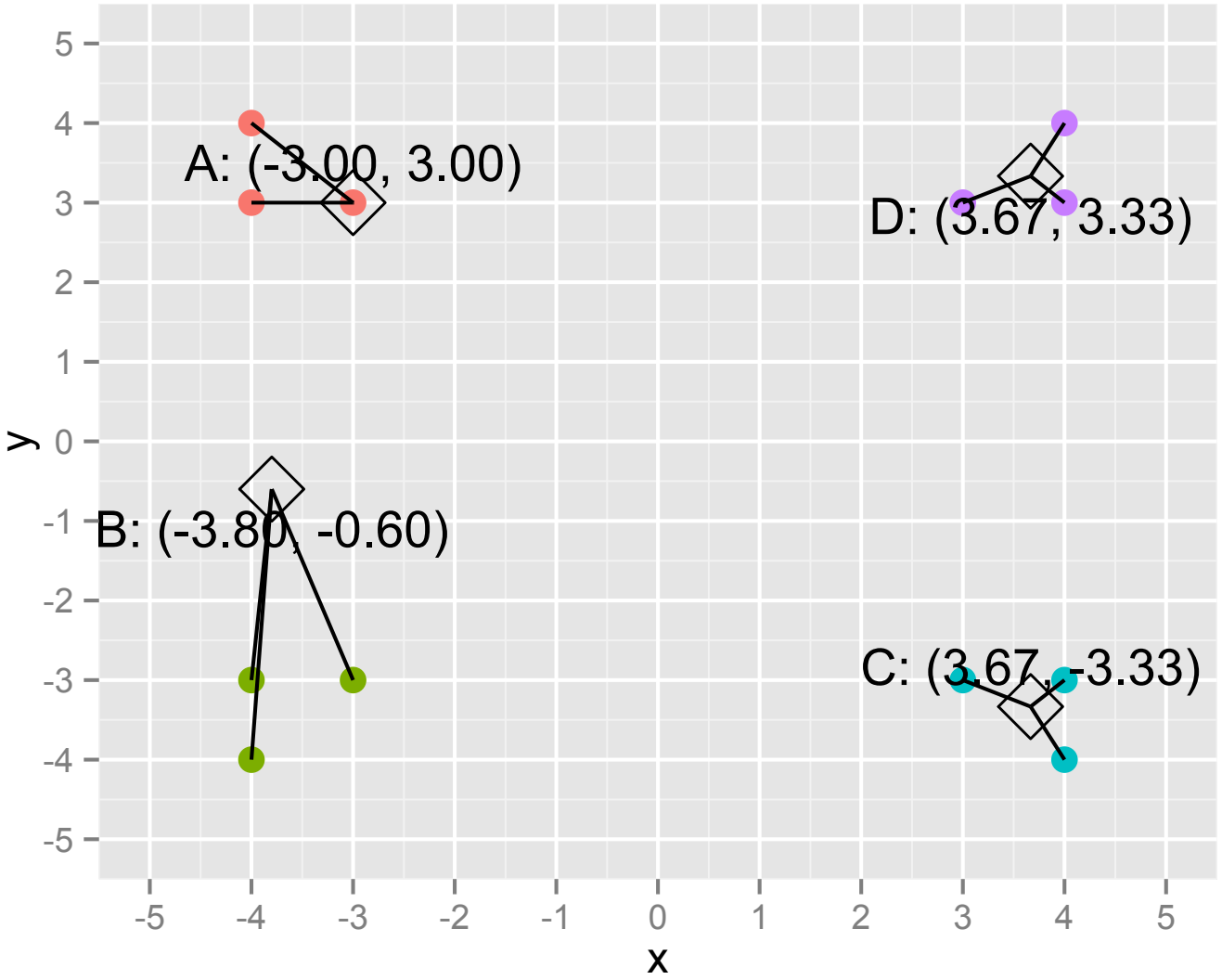
Four Points

$$\mu_A = (-3, 3)$$
$$\mu_B = (-3.8, -0.6)$$
$$\mu_C = (3.67, -3.33)$$
$$\mu_D =$$

Four Points

$$\mu_A = (-3, 3)$$
$$\mu_B = (-3.8, -0.6)$$
$$\mu_C = (3.67, -3.33)$$
$$\mu_D = (3.67, 3.33)$$

Four Points


A: (-3.00, 3.00)
B: (-3.80, -0.60)
C: (3.67, -3.33)
D: (3.67, 3.33)
y
x
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

Four Points

$$\mu_A =$$

$$\mu_B =$$

$$\mu_C =$$

$$\mu_D =$$

Four Points

$$\mu_A = (-3.67, 3.33)$$

$$\mu_B =$$

$$\mu_C =$$

$$\mu_D =$$

Four Points

$$\mu_A = (-3.67, 3.33)$$
$$\mu_B = (-3.67, -3.33)$$
$$\mu_C =$$
$$\mu_D =$$

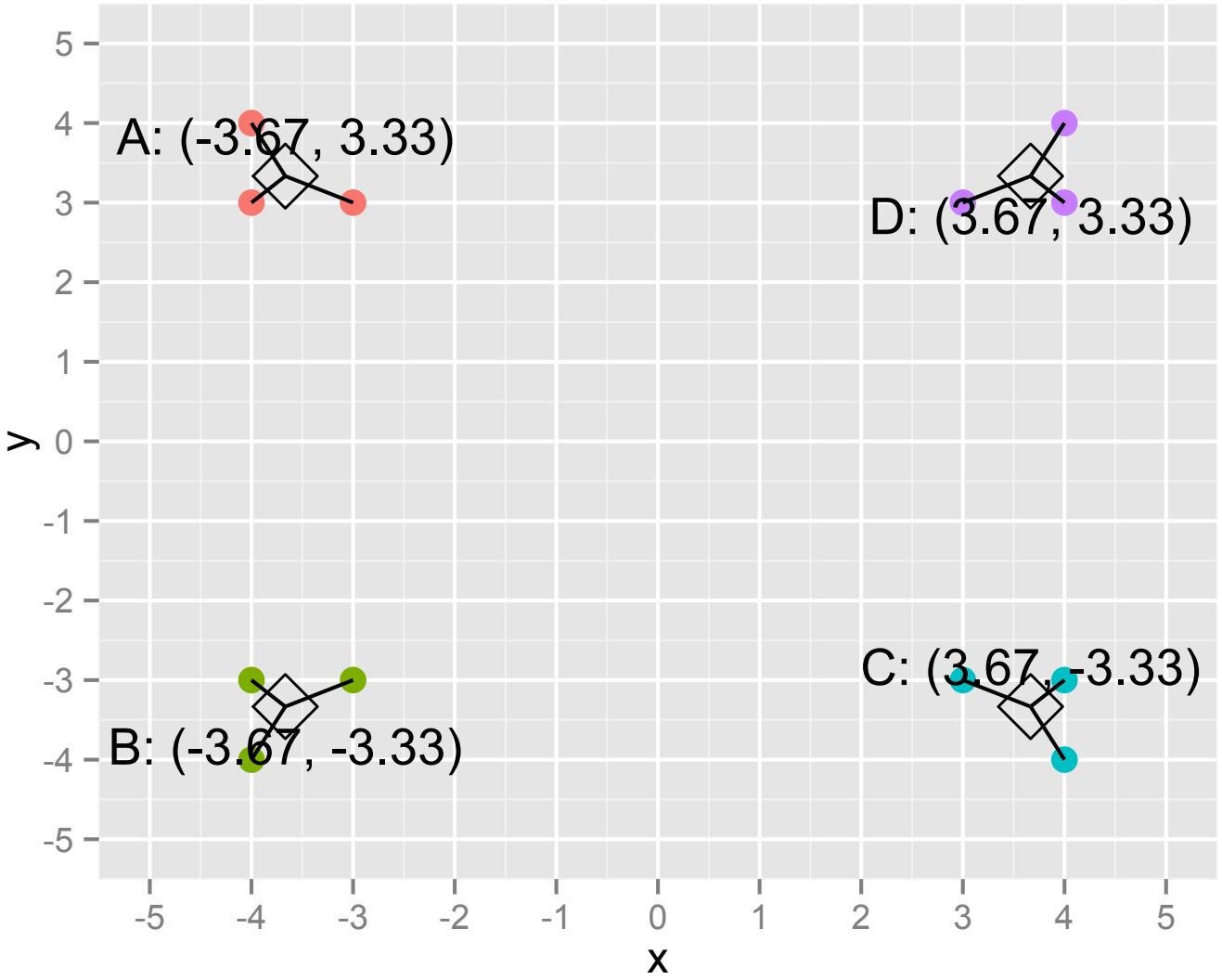
Four Points

$$\mu_A = (-3.67, 3.33)$$
$$\mu_B = (-3.67, -3.33)$$
$$\mu_C = (3.67, -3.33)$$
$$\mu_D =$$

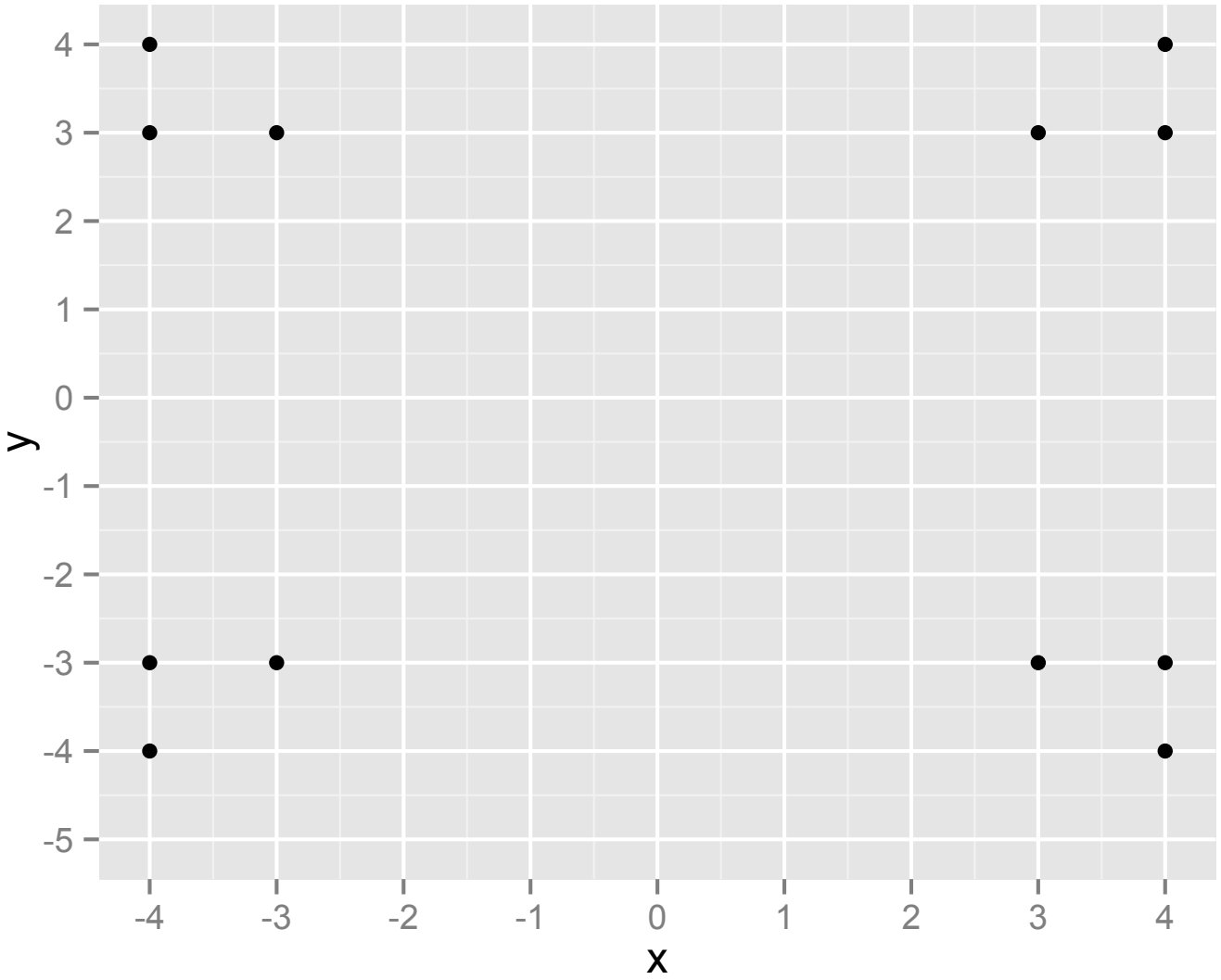
Four Points

$$\mu_A = (-3.67, 3.33)$$
$$\mu_B = (-3.67, -3.33)$$
$$\mu_C = (3.67, -3.33)$$
$$\mu_D = (3.67, 3.33)$$

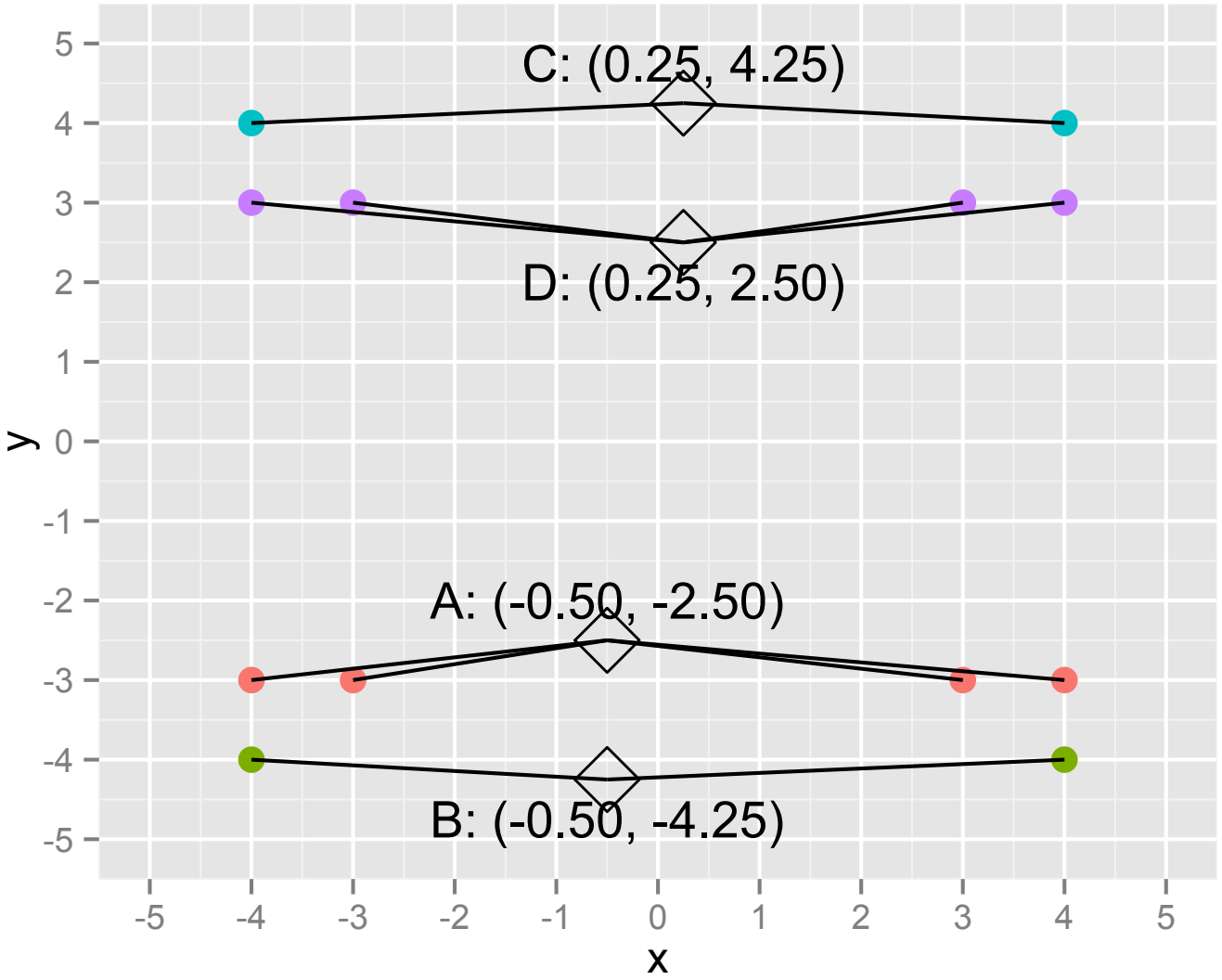
Four Points


A: (-3.67, 3.33)
B: (-3.67, -3.33)
C: (3.67, -3.33)
D: (3.67, 3.33)
y
x
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

Bad Initialization



Bad Initialization



Bad Initialization