Data Mining

K-means clustering

Mariia Rybalka Elchin Valiyev Abbas Khan Maxim Radomskyi

Group 15

Outline

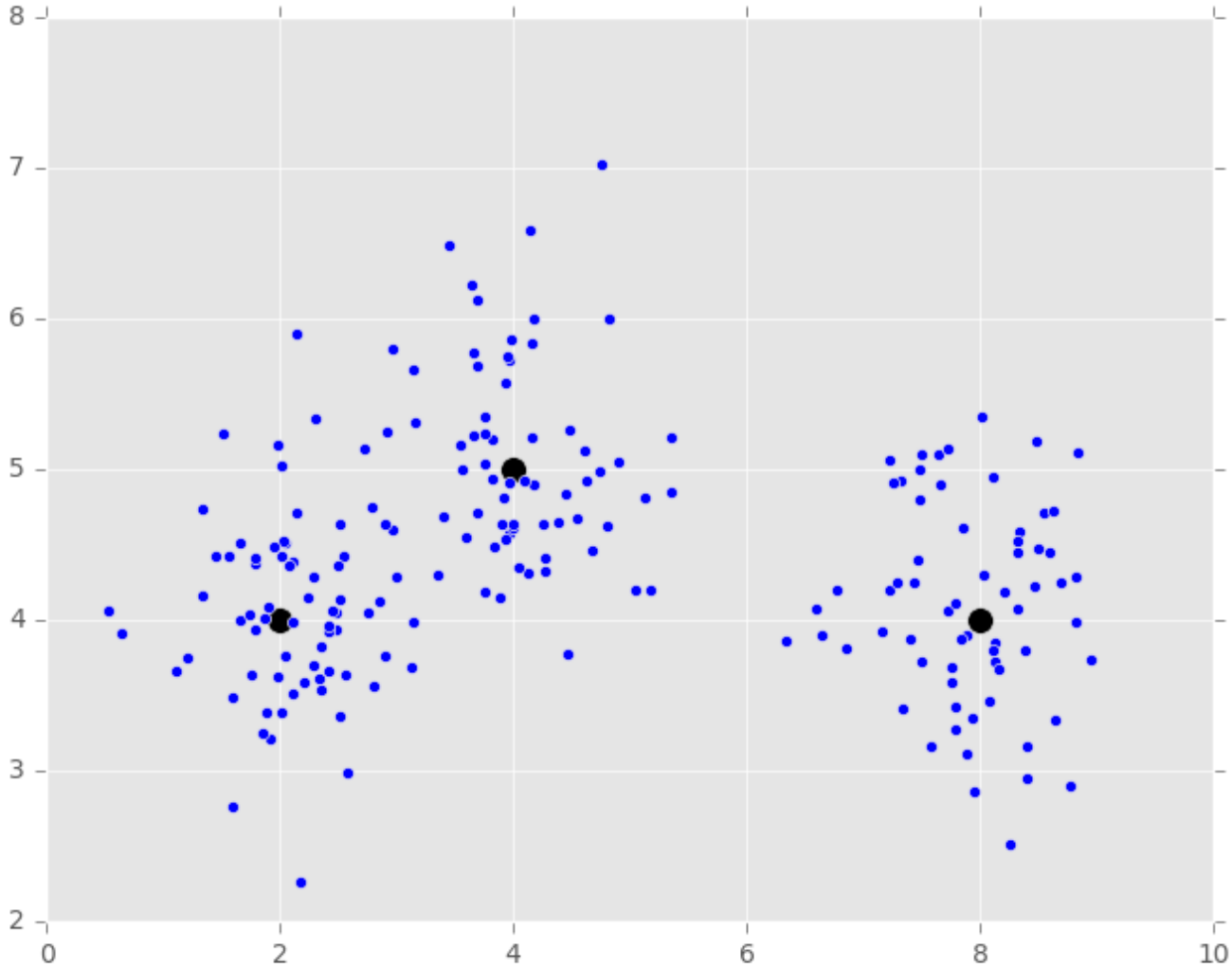
K-means

```
double[][] centers = initializeCenters(data, k); // initializing centers
double[][] oldCenters;

do {
        oldCenters = centers;
        int[] assignments = EStep(data, oldCenters); // finding closest center
        centers = MStep(data, assignments, k);  // calculating new centers

//checking if new centers and old centers are different
} while (!equal(oldCenters, centers));
```

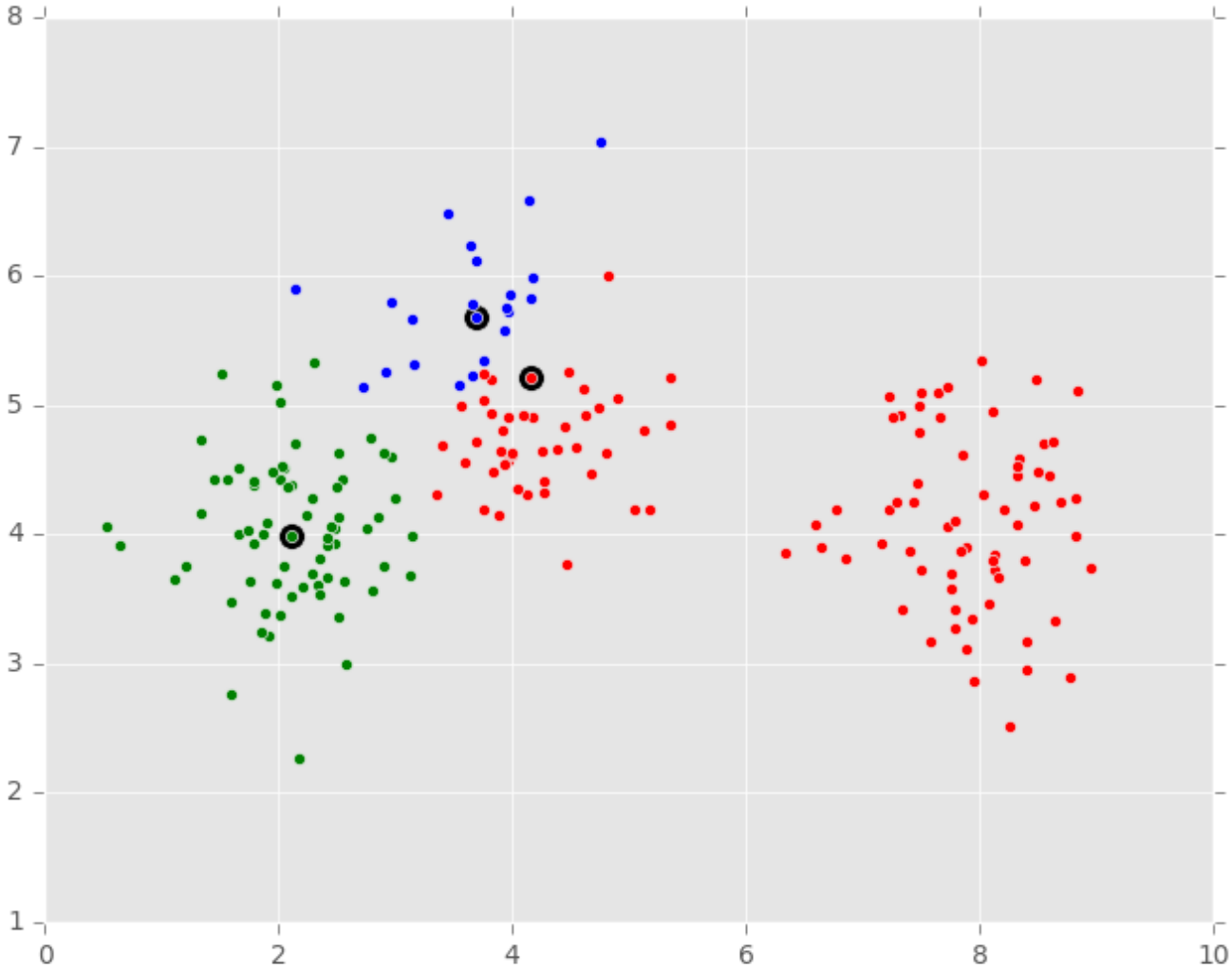
Euclidean distance

$$d(x, y) = \sqrt{\sum_{i}^{n} (x_i - y_i)^2}$$

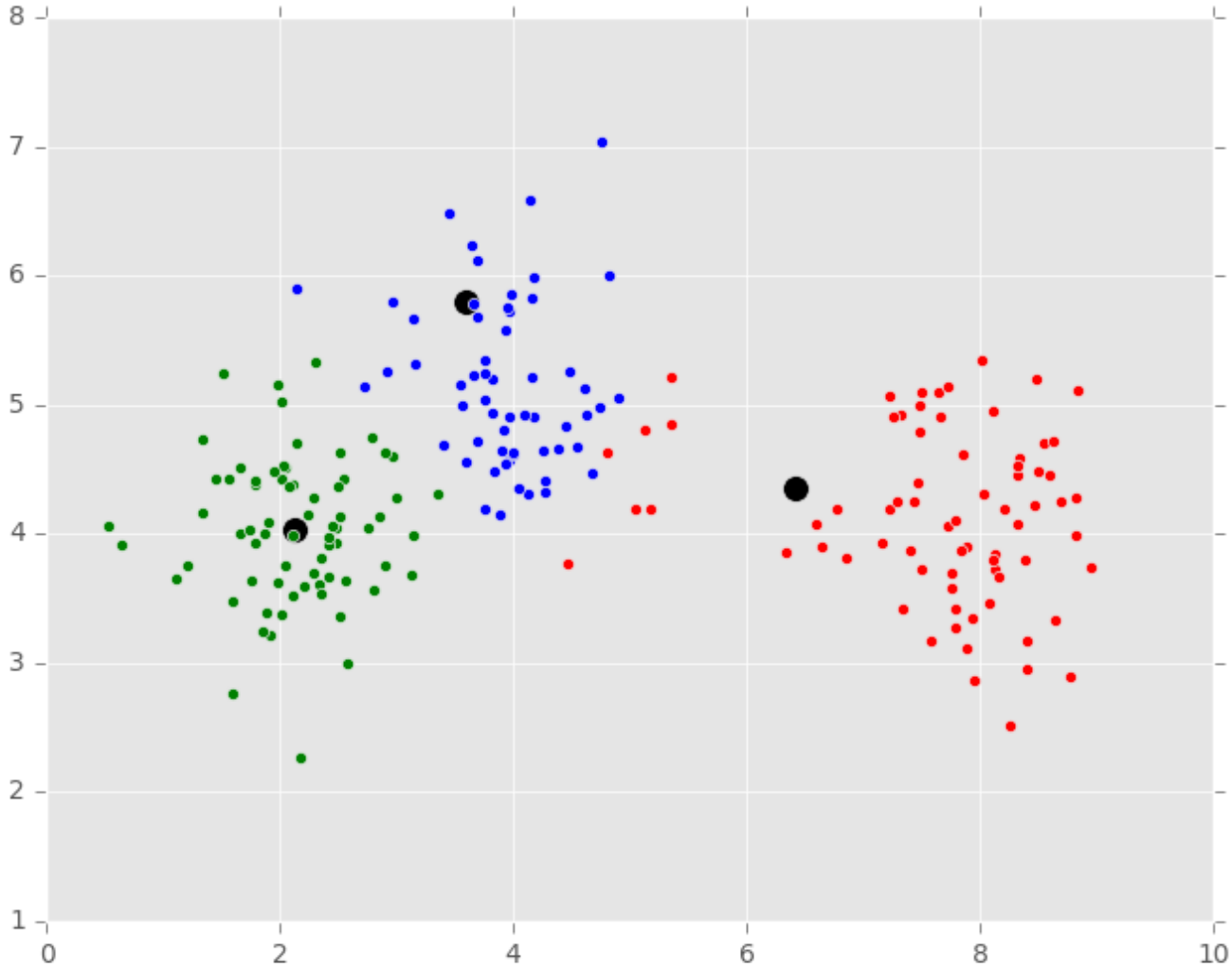
Euclidean distance



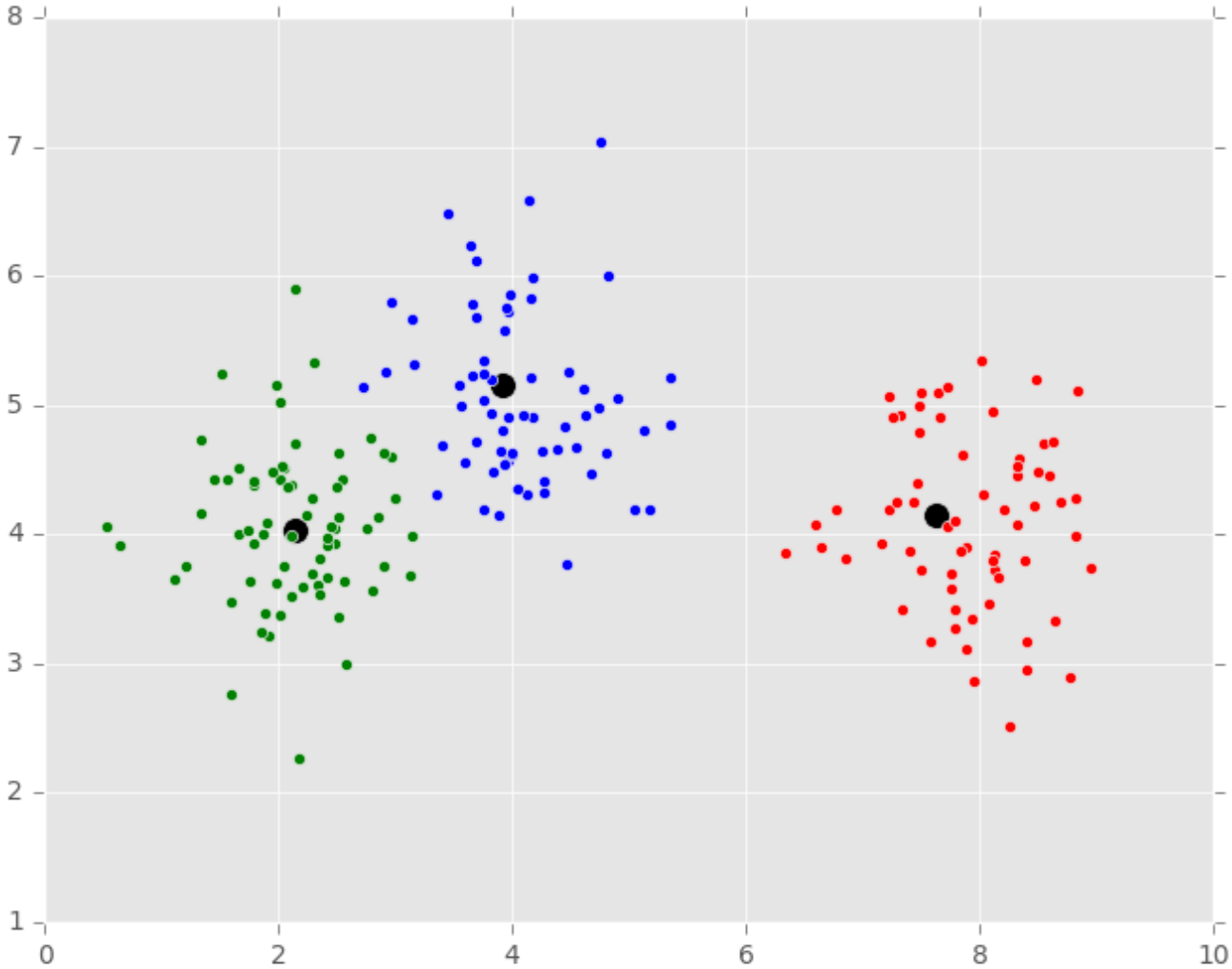
Euclidean distance



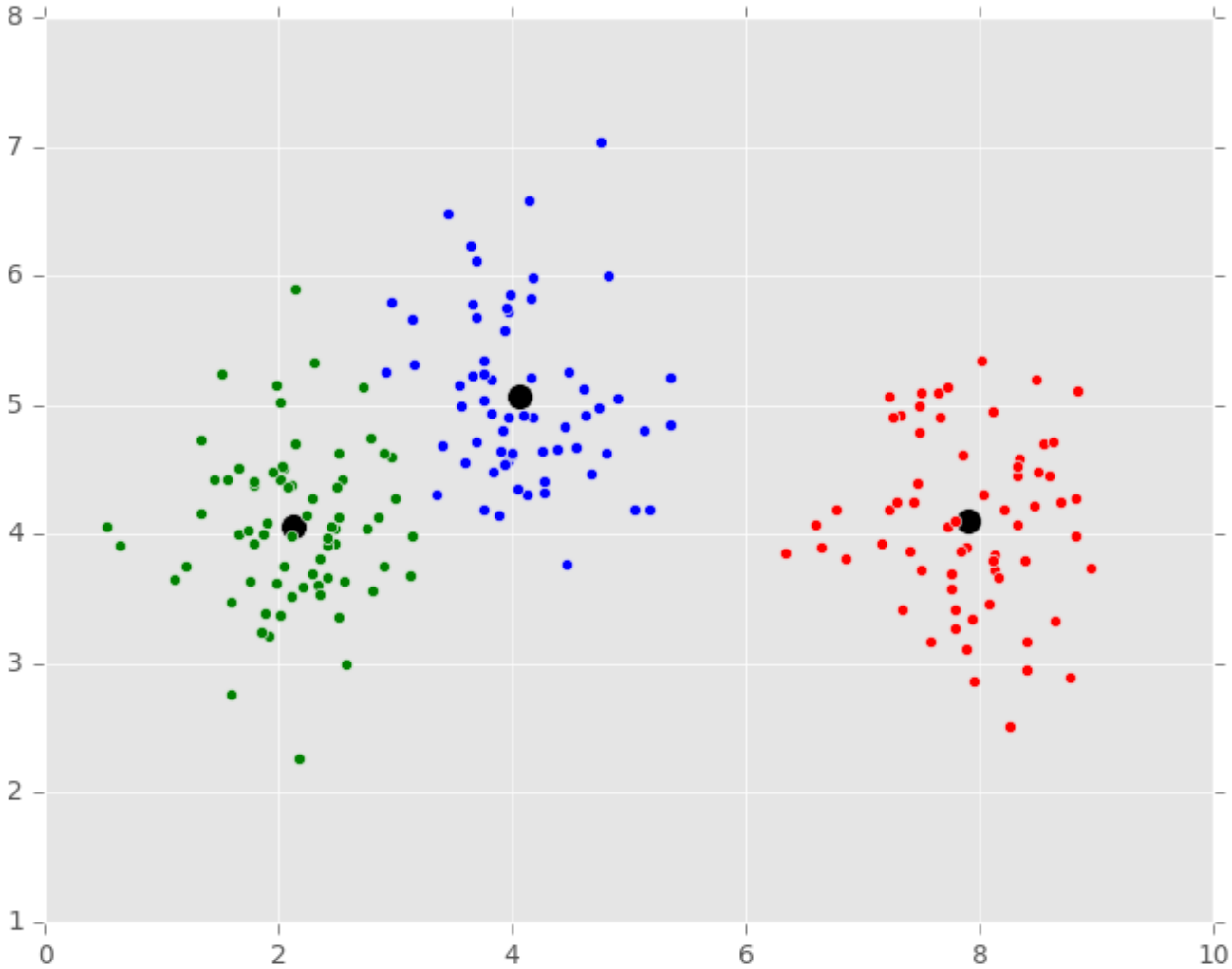
Euclidean distance



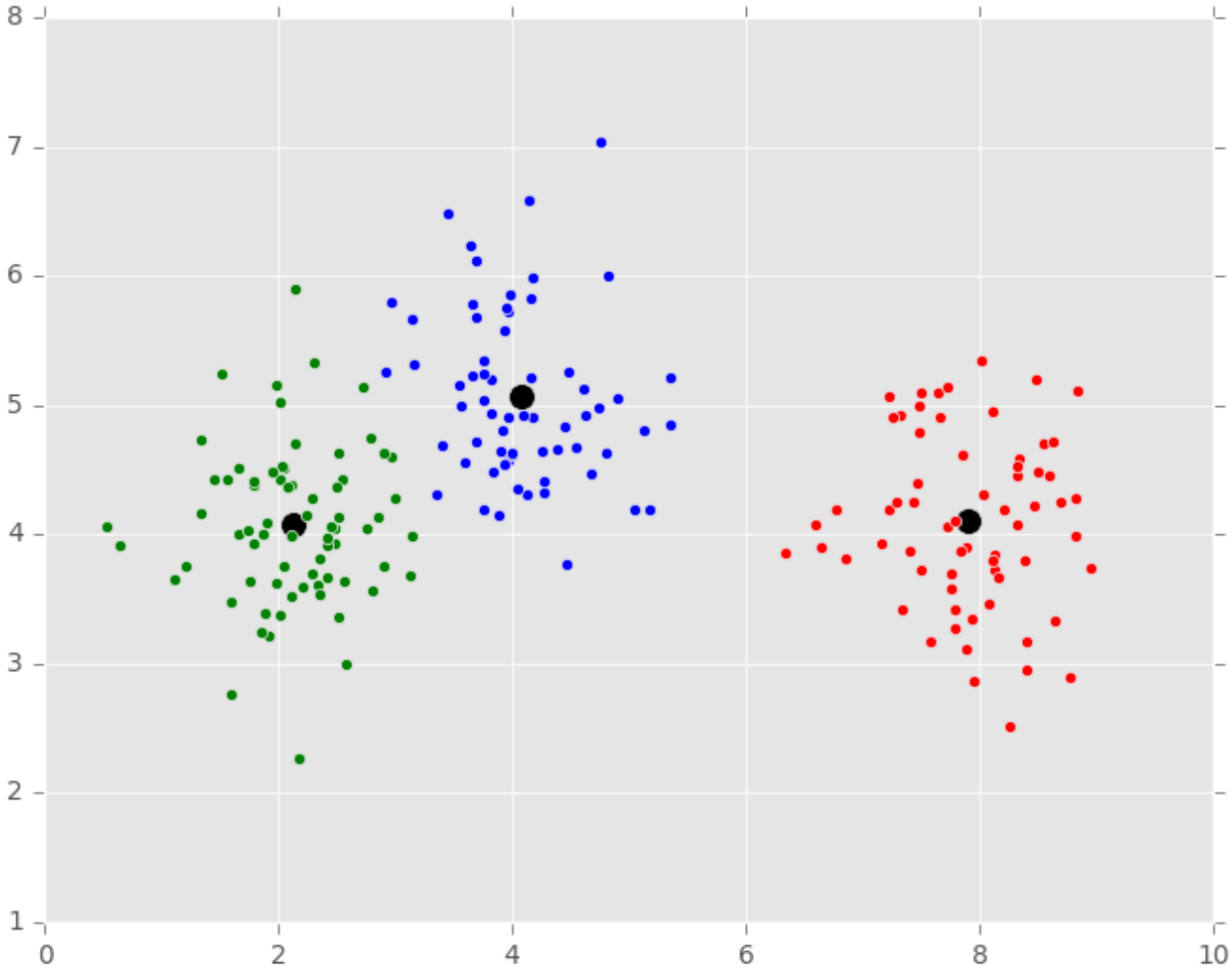
Euclidean distance



Euclidean distance



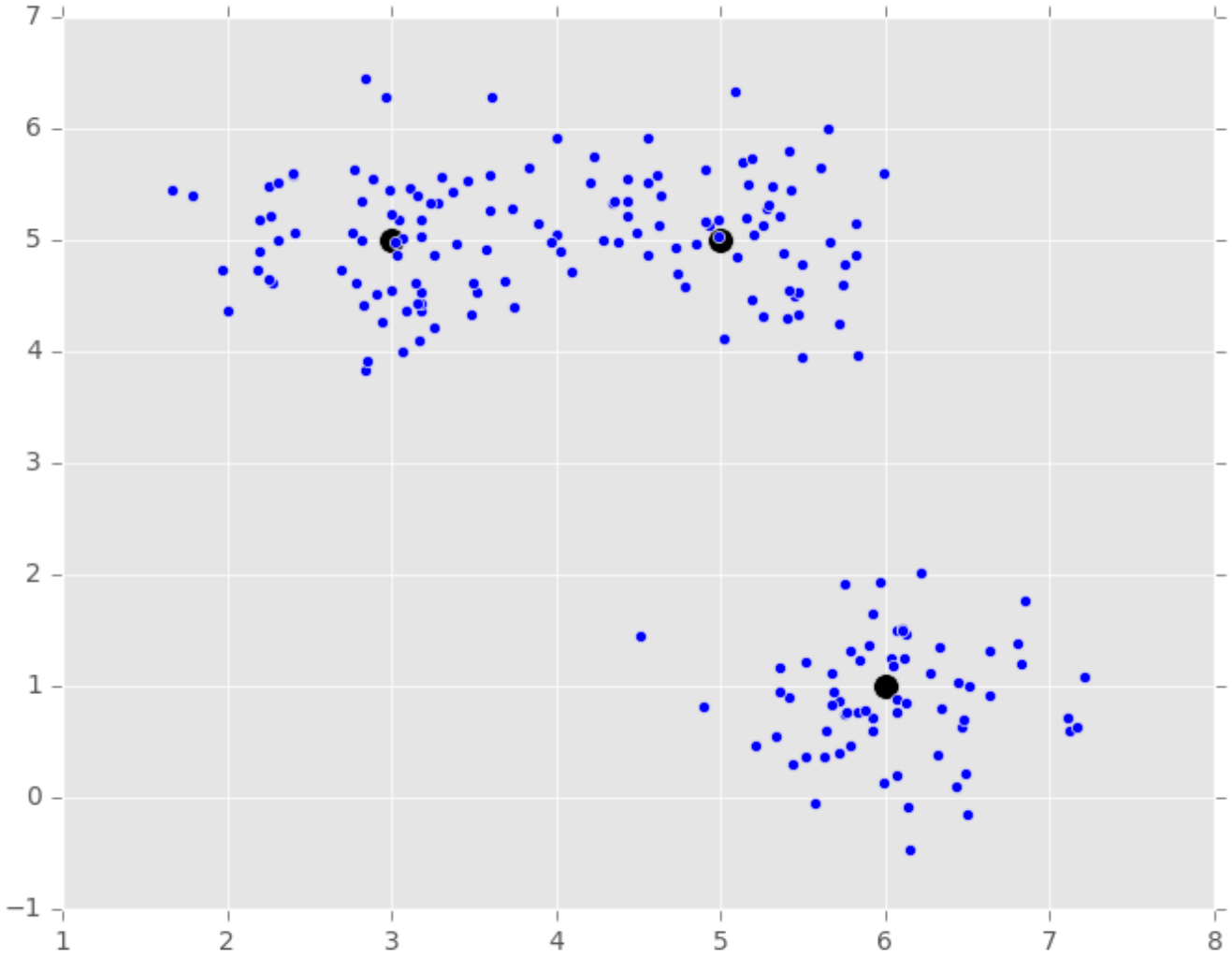
Euclidean distance



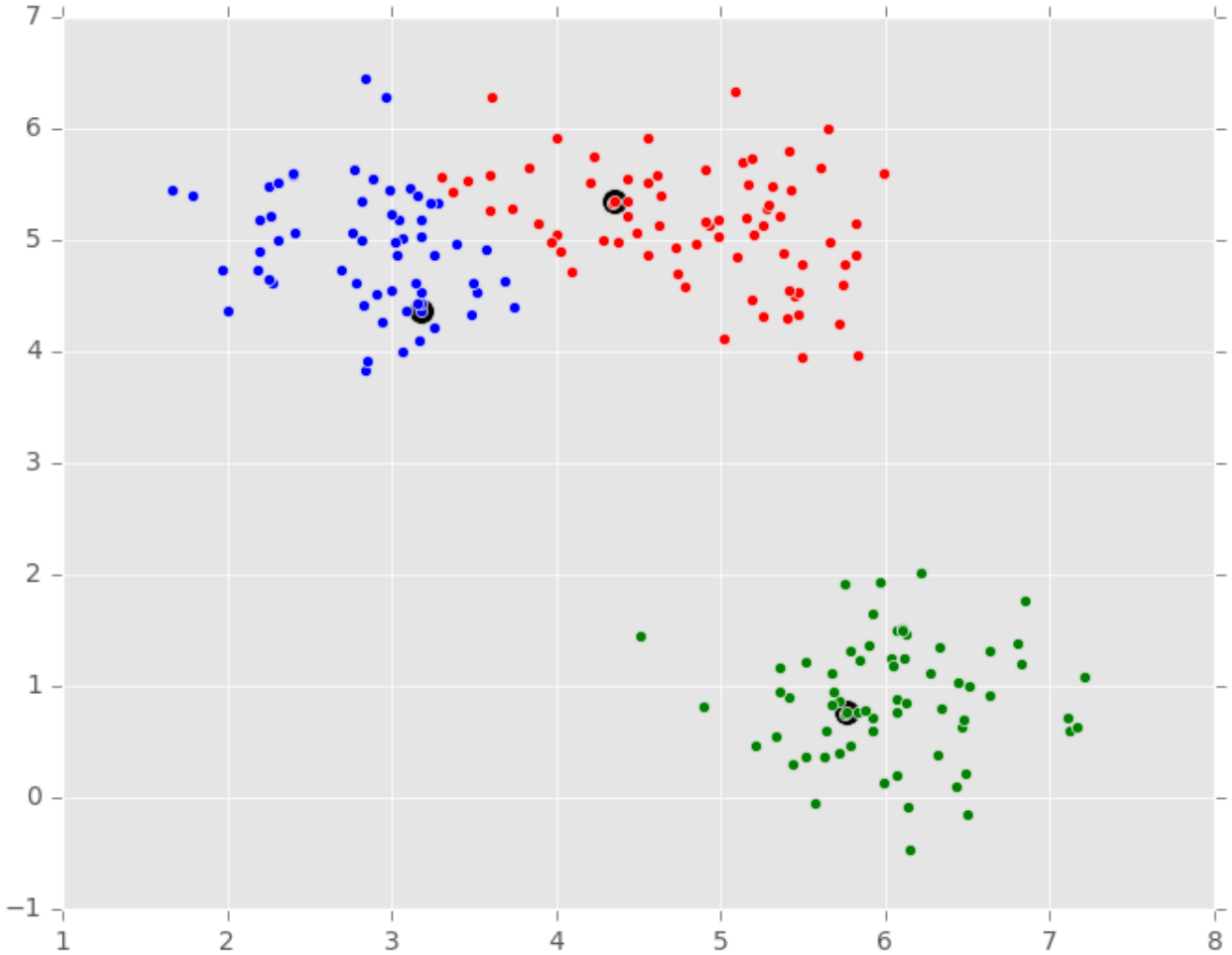
Manhattan distance

$$d(x, y) = \sum_{i}^{n} |x_i - y_i|$$

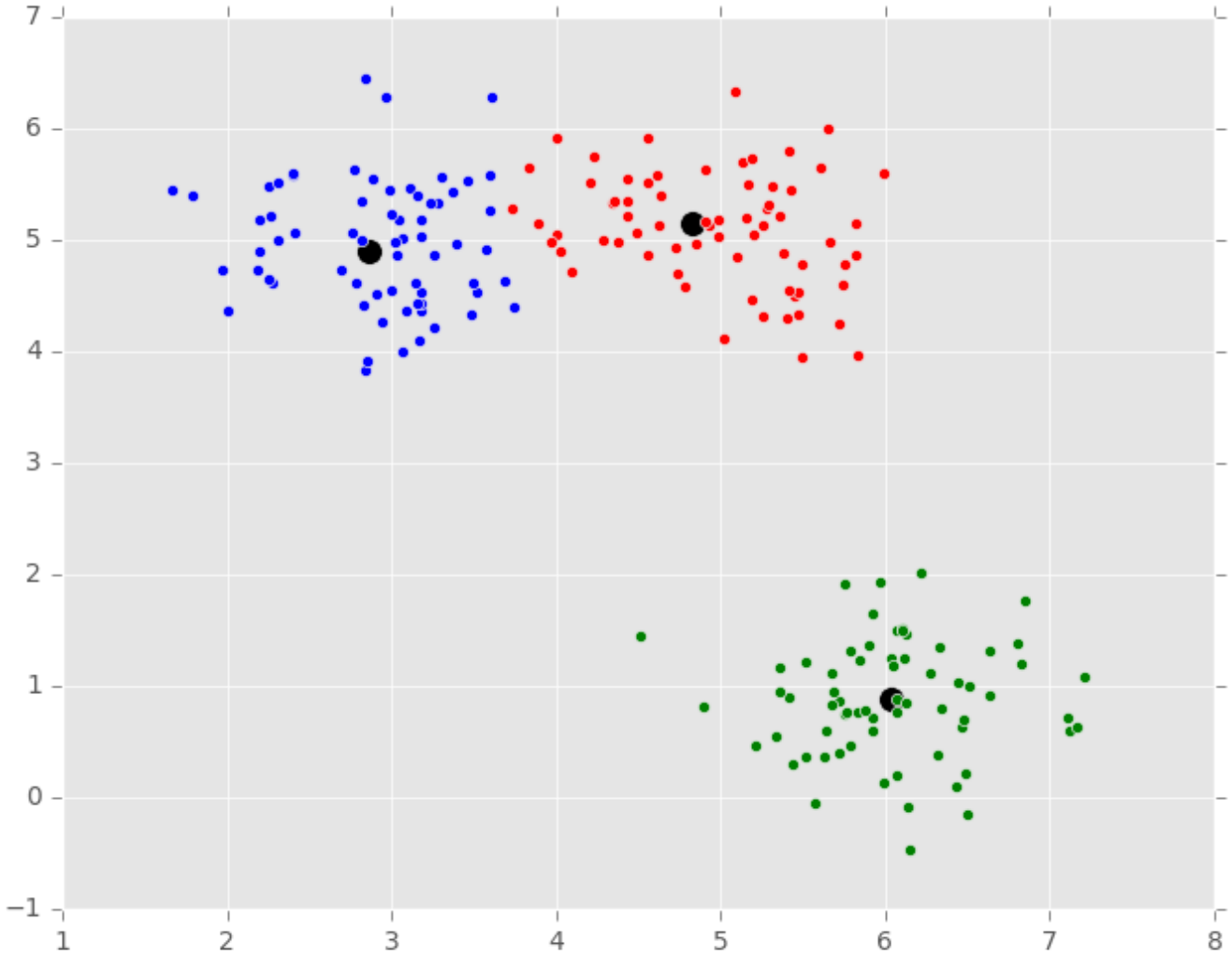
Manhattan distance



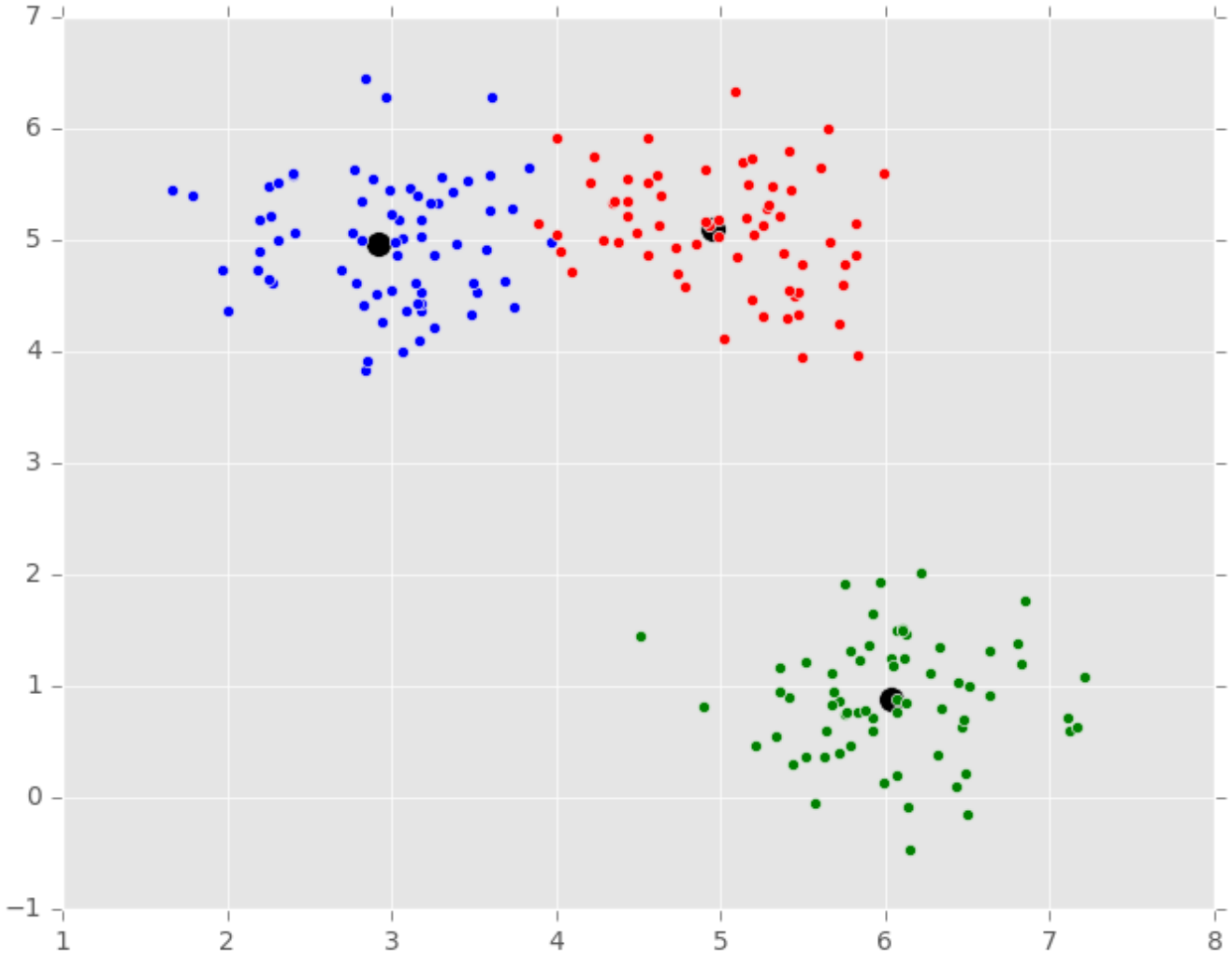
Manhattan distance



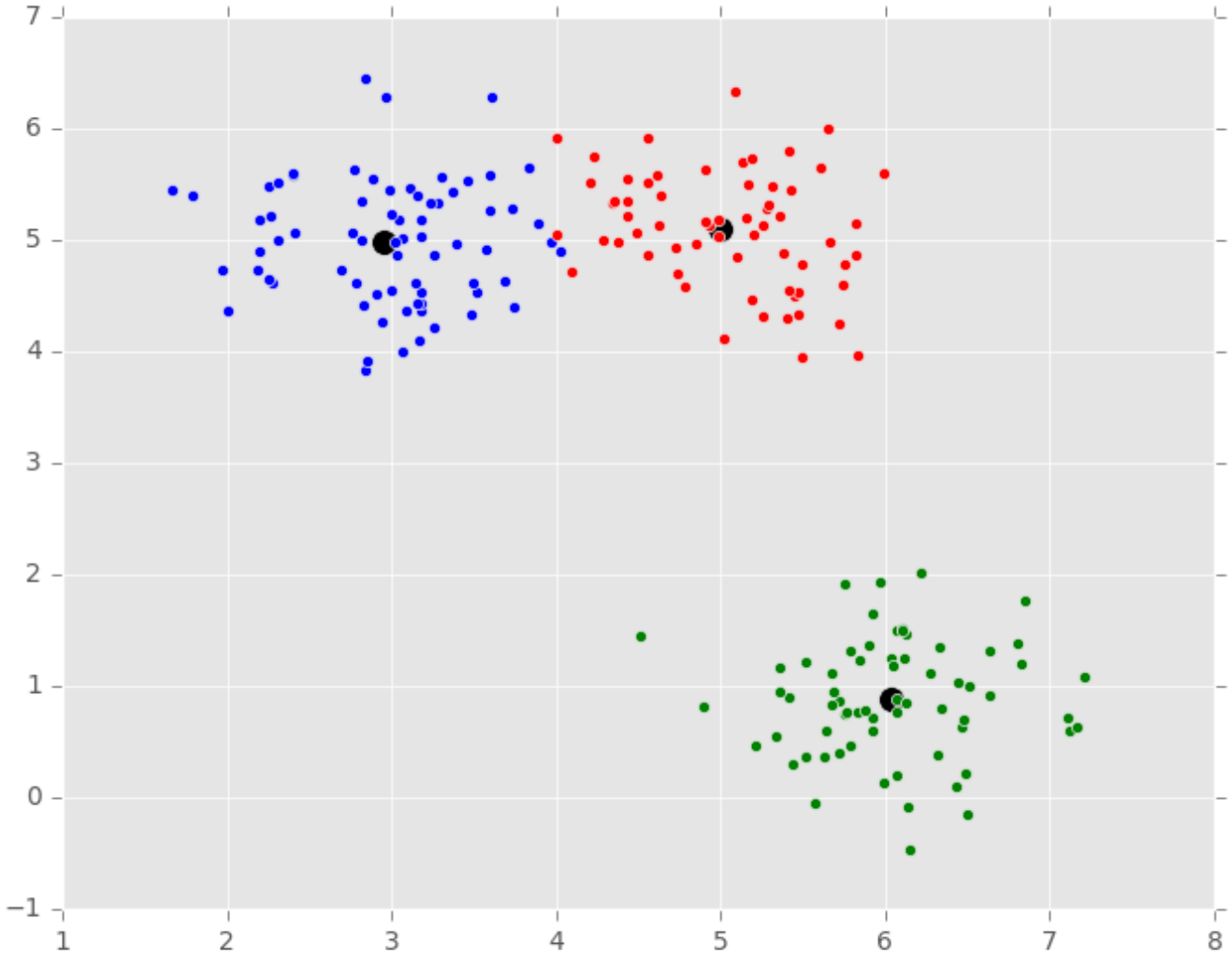
Manhattan distance



Manhattan distance



Manhattan distance



Manhattan distance