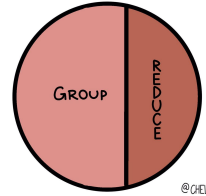


SIMPLIFY



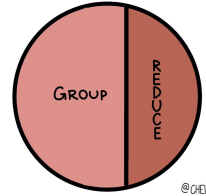
@CHELSEAPARLETT

# K-Means and Expectation Maximization

Chelsea Parlett-Pelleriti

# Unsupervised Machine Learning

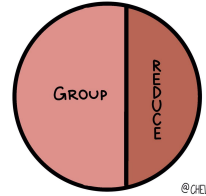
SIMPLIFY



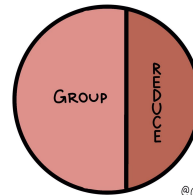
@GELSEAPARLETT

# Clustering

SIMPLIFY



@GELSEAPARLETT



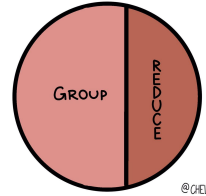
# K-Means

1. Choose **k** random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers
4. Repeat 2 and 3 until either:
  - a. Cluster membership does not change
  - b. Centers change only a tiny amount

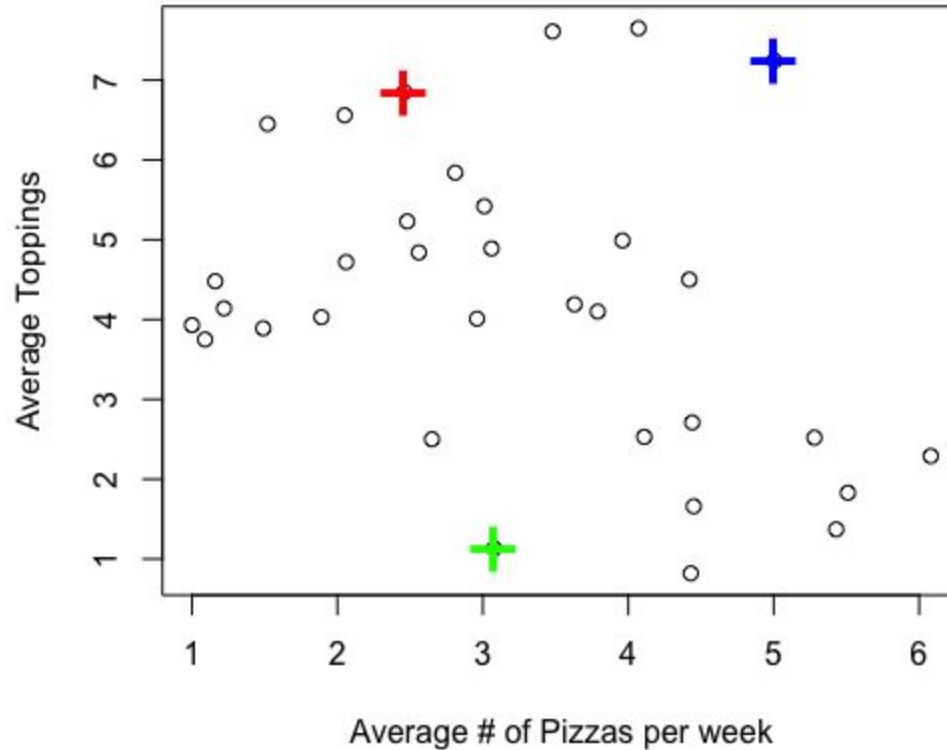
# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers

SIMPLIFY



@GELSEAPARLETT

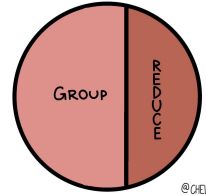


1

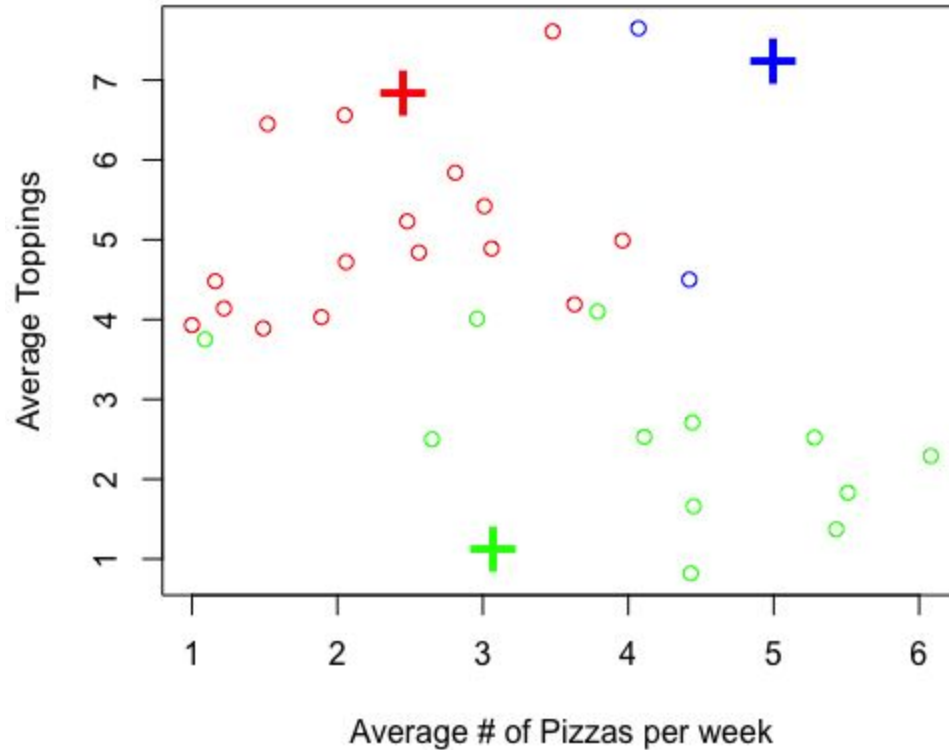
# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers

SIMPLIFY



@GELSEAPARLETT

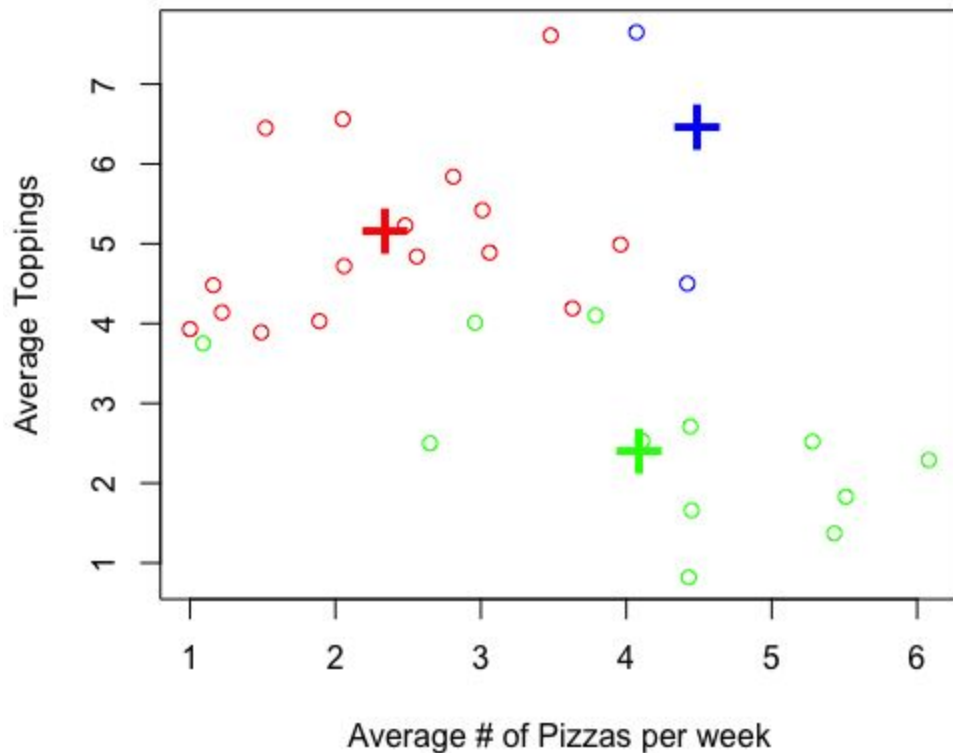


2

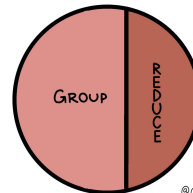
# 3

## K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers



SIMPLIFY

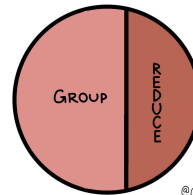


@GELSEAPARLETT

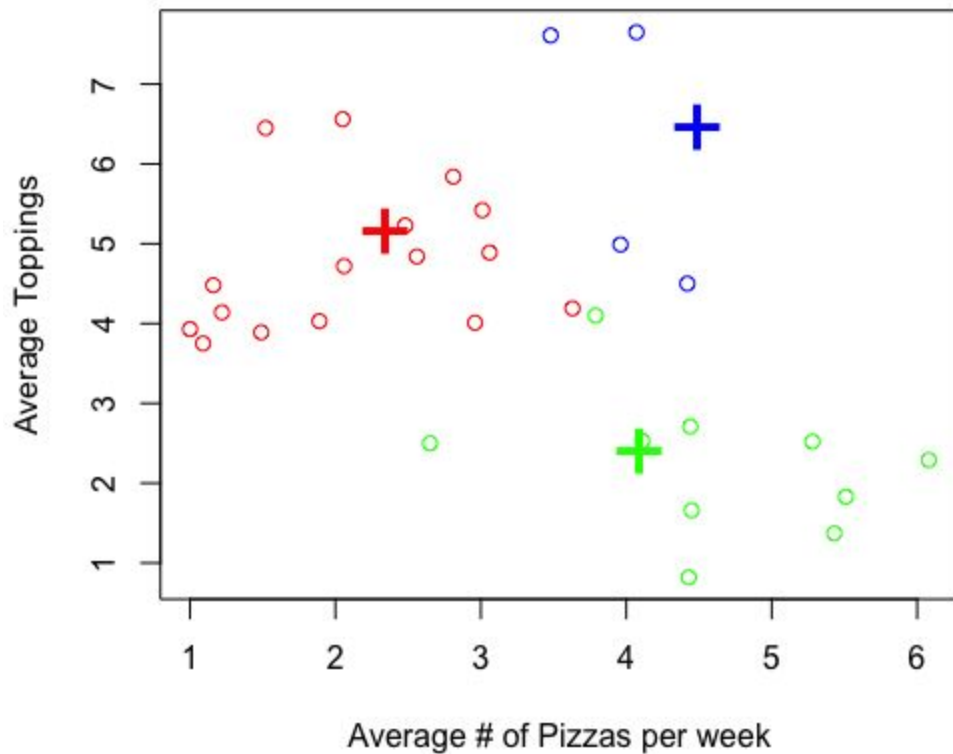
# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers

SIMPLIFY



@GELSEAPARLETT



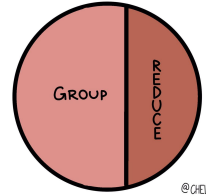
2



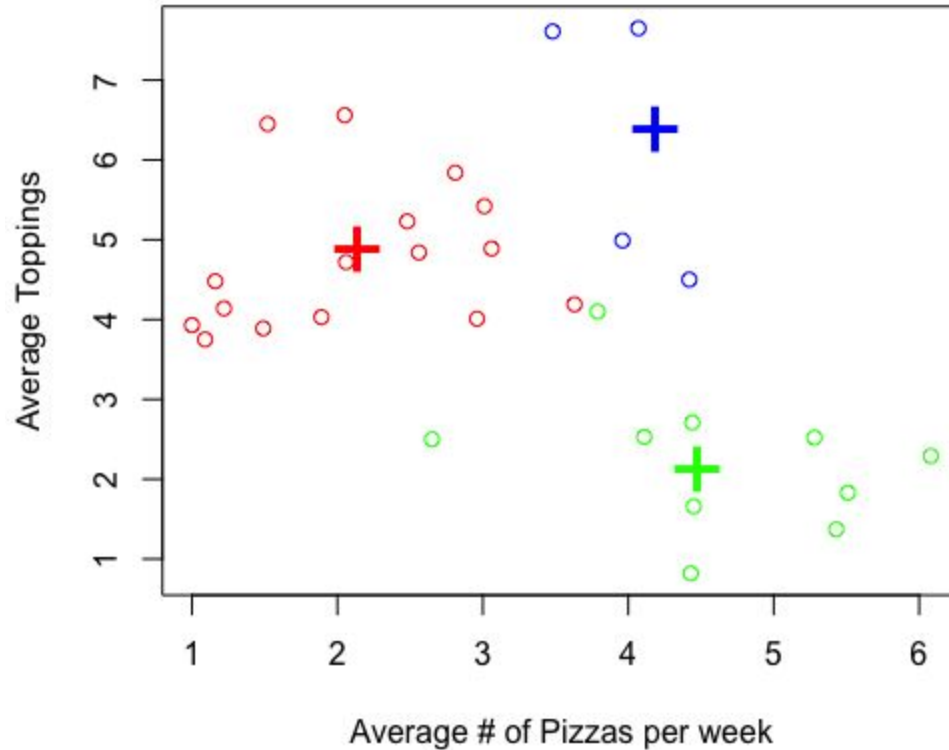
# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers

SIMPLIFY

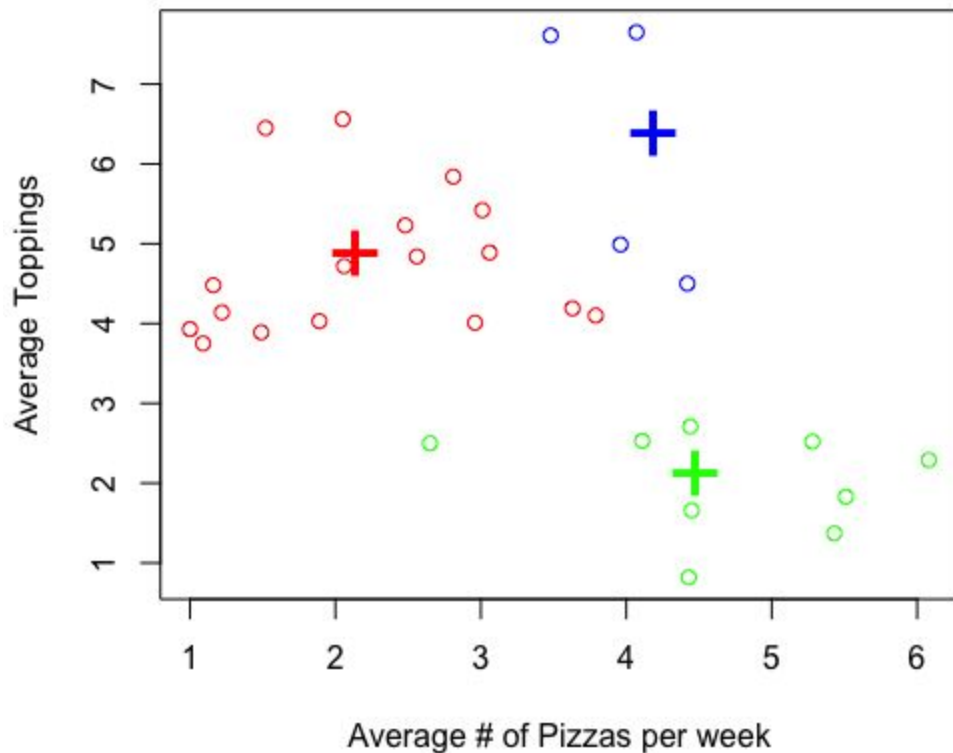


@GELSEAPARLETT

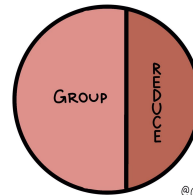


# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers



SIMPLIFY

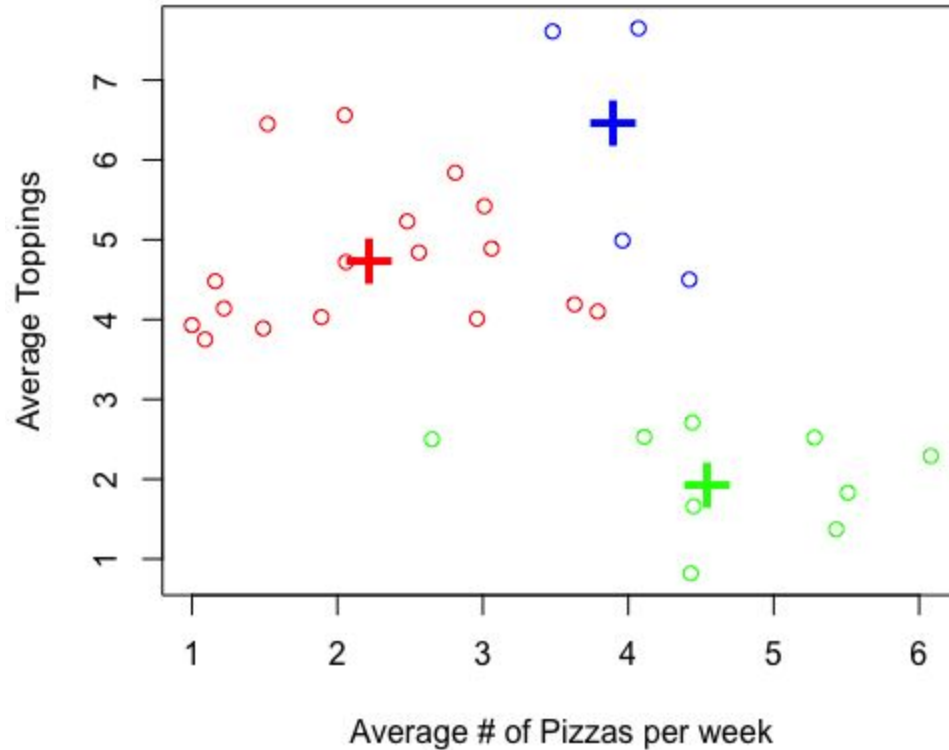


@GELSEAPARLETT

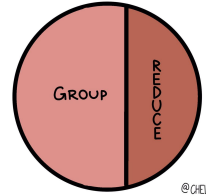
2

# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers



SIMPLIFY

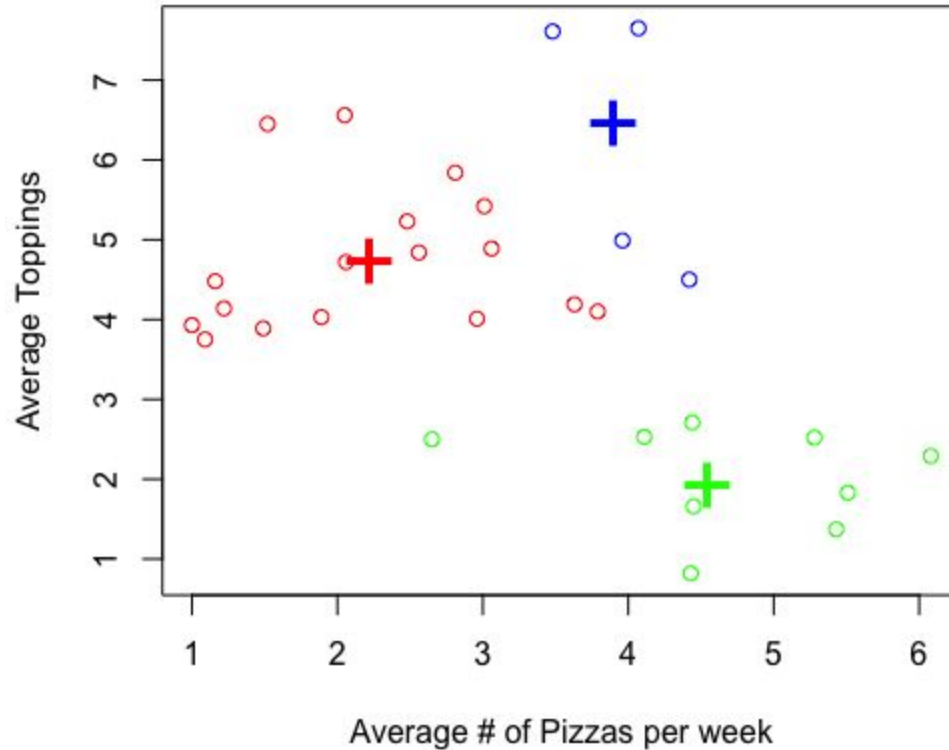


@GELSEAPARLETT

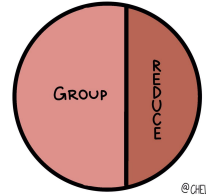
3

# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers



SIMPLIFY

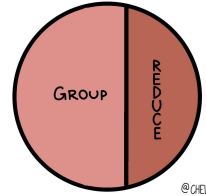


@GELSEAPARLETT

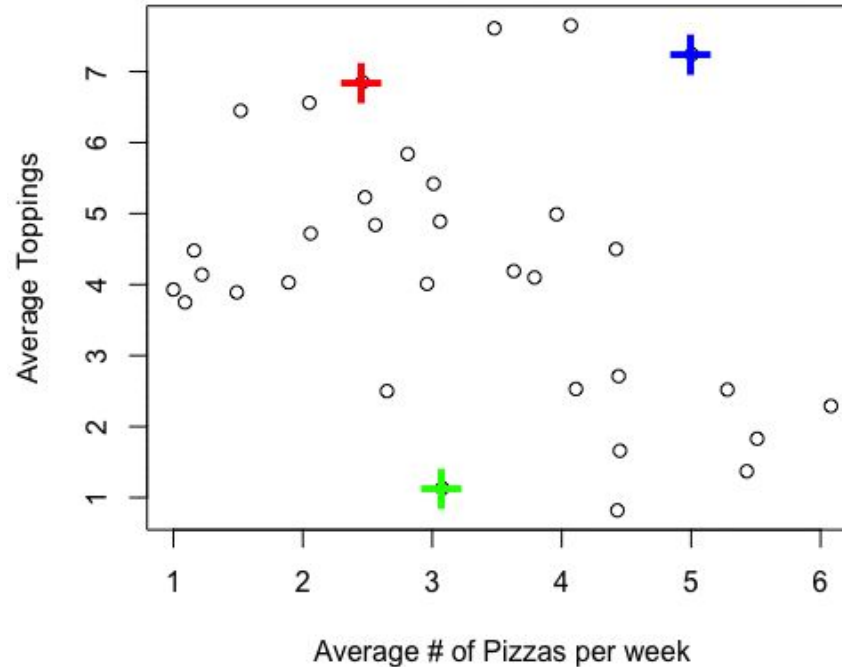
# K-Means

1. Choose  $k$  random points to be cluster centers
2. For each data point, assign it to the cluster whose center is closest
3. Using these assignments, recalculate the centers

SIMPLIFY



@GELSEAPARLETT



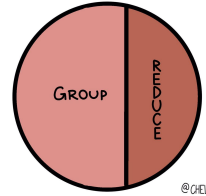
# K-Means

Assumptions

Spherical Clusters

Roughly the same # in each cluster

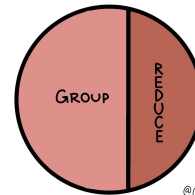
SIMPLIFY



@GELSEAPARLETT

# Evaluating Unsupervised Models

SIMPLIFY



Cohesion:

Separation:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$