Definition: Clustering

- **Clustering** is another application of our fundamental notion of similarity.

- The basic idea is that we want to find groups of objects (consumers, businesses, etc.), where the objects within groups are similar,
  - but the objects in different groups are not so similar.
Clustering Around Centroids
Clustering Around Centroids
Example: $k$-means Clustering
Example: $k$-means Clustering
Clustering

• $k$-means clustering is efficient

• $k =$?
Generating Cluster Descriptions

BODY_round:

- color_red:
  - color_f.gold:
    - not_J
    - not_J
  - not_J

BODY_light:
  - FIN_dry:
    - not_J
    - J

NOSE_SHERRY:
  - not_J
  - J
Case Study: Credit Line Optimization

- Credit granting businesses face a challenging environment due to the wide variety of customer behaviors
Case Study: Credit Line Optimization

Framework for credit customer optimization based on clustering and predictions:

- Customer clusters are formed by using clustering from past credit performance data

- *Within each of the $K$ clusters*, the *expected net present value* to the credit company is estimated

- External data is used to predict for *new accounts* the probabilities of membership for each performance cluster

- The prediction is done using classification and regression trees

Clustering + Expected Value Framework
Thanks!
Questions?