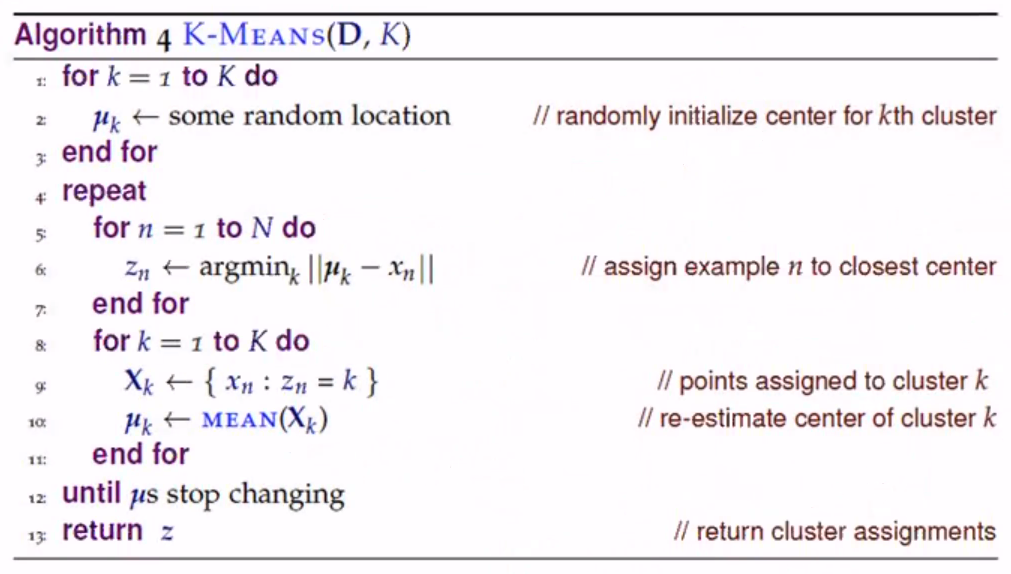
**Machine Learning : K-means clustering.**

**Types of machine learning**

* **Supervised learning** 
  + **Decision Tree Classifier, Majority Classifier, Nearest Neighbor Classifier, Perceptron Classifier**
  + Labeled data.
  + Direct feedback.
  + Predict outcome/future.
* **Unsupervised learning**
  + No labels.
  + No feedback.
  + “Find hidden structure.”
* **Reinforcement learning**
  + Decision process.
  + Reward system.
  + Learn series of actions.

**K-means clustering**

* Unsupervised learning



**D = Data set of points**

**K = Number of clusters**

**1st Loop = Creates initial set of clusters**

**2nd Loop = Refines the set of clusters**

**Complexity**

**K : clusters**

**N : data points**

**d : dimensions**

**t : iterations**

**O(KNdt)**

**Will it converge (stop changing)?**

* Yes because t is finite.

**K-means hyperparameters**

* **K**
  + Choose using elbow method.
* **Distance**
  + Euclidean
  + Manhattan

clusters = kmeans(dataset,k)

for k in range(max(k)):

for cluster in clusters:

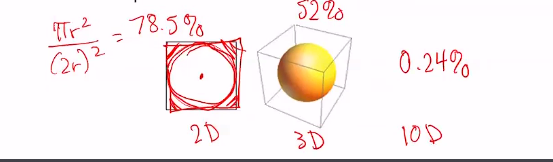
mean = clusterMean(cluster)

for datapoint in cluster:

sse[k] +=

**The curse of dimensionality**

* Cost
* ML performance is bad the more features you add.
* As complexity increases, overfit increases mean you need more training data.
* As you get high dimension you end up not having near neighbors and distance does not have any meaning.



**At higher D points will land in the corner and not near where you want**

**Distance to corner =**

