SLAM 101

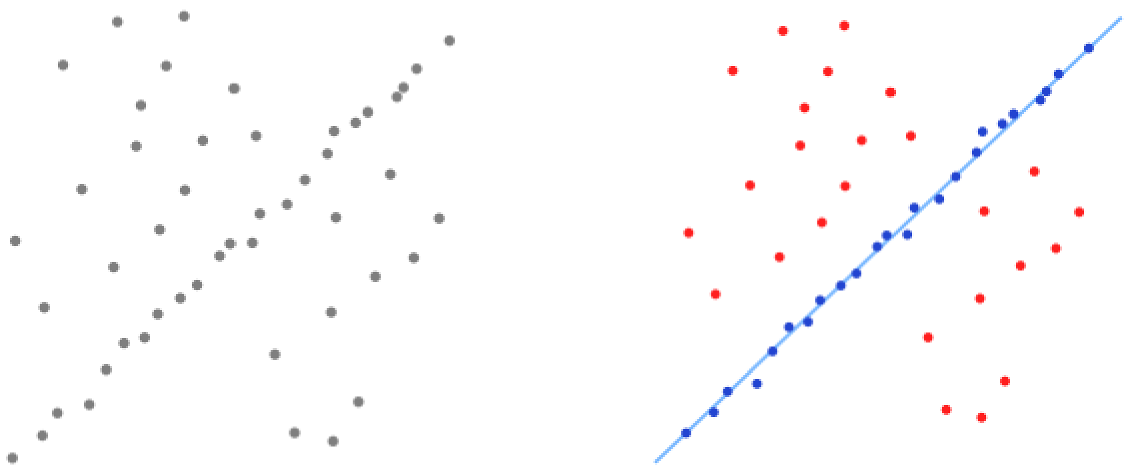
Lecture 00 Machine Learning2

Ayoung Kim

RANSAC

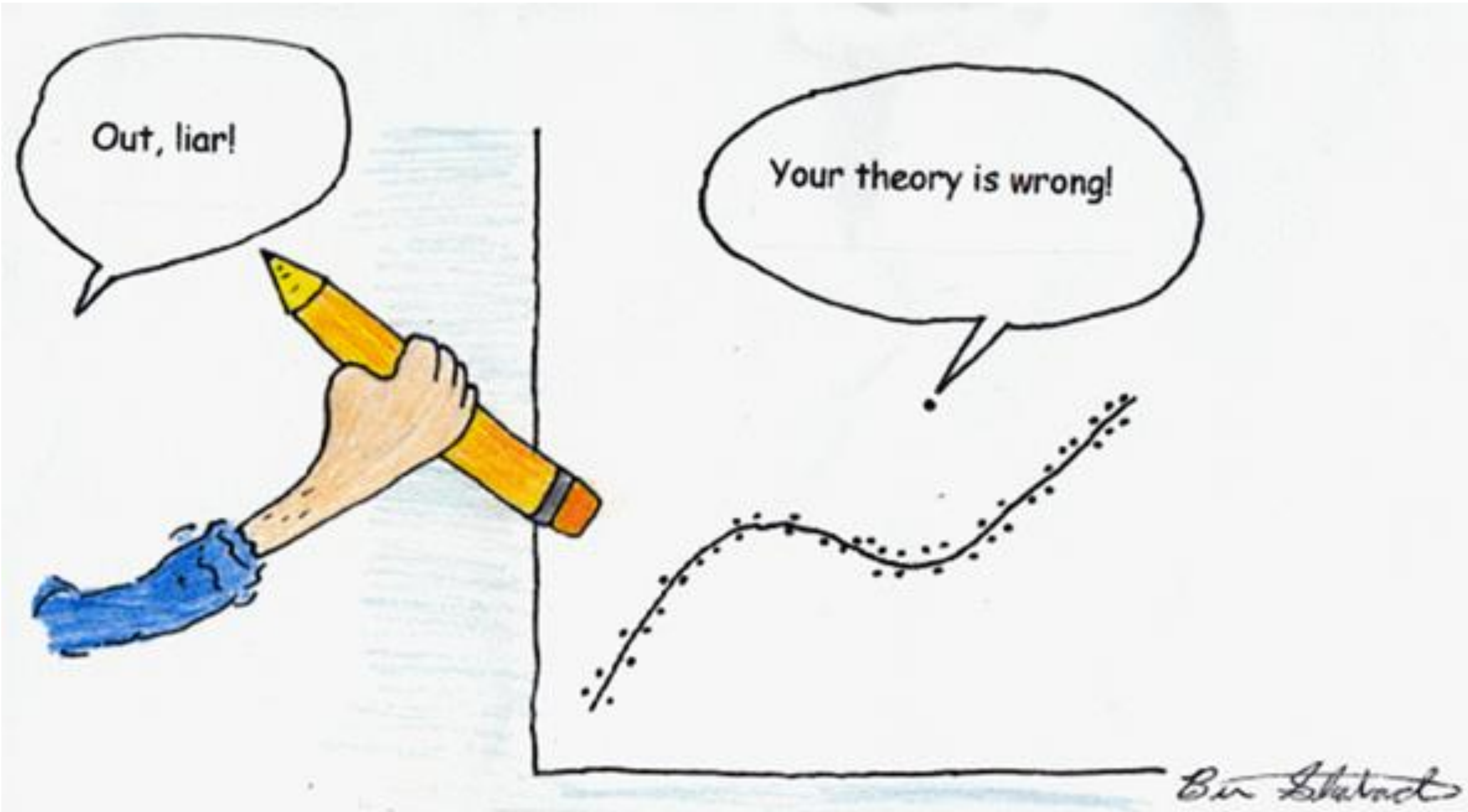
RANSAC

- RANdom SAmple Consensus (RANSAC) (Fischler, Bolles 1981)
  - Estimate parameters of a mathematical model
  - Given a set of observed data which contains outliers



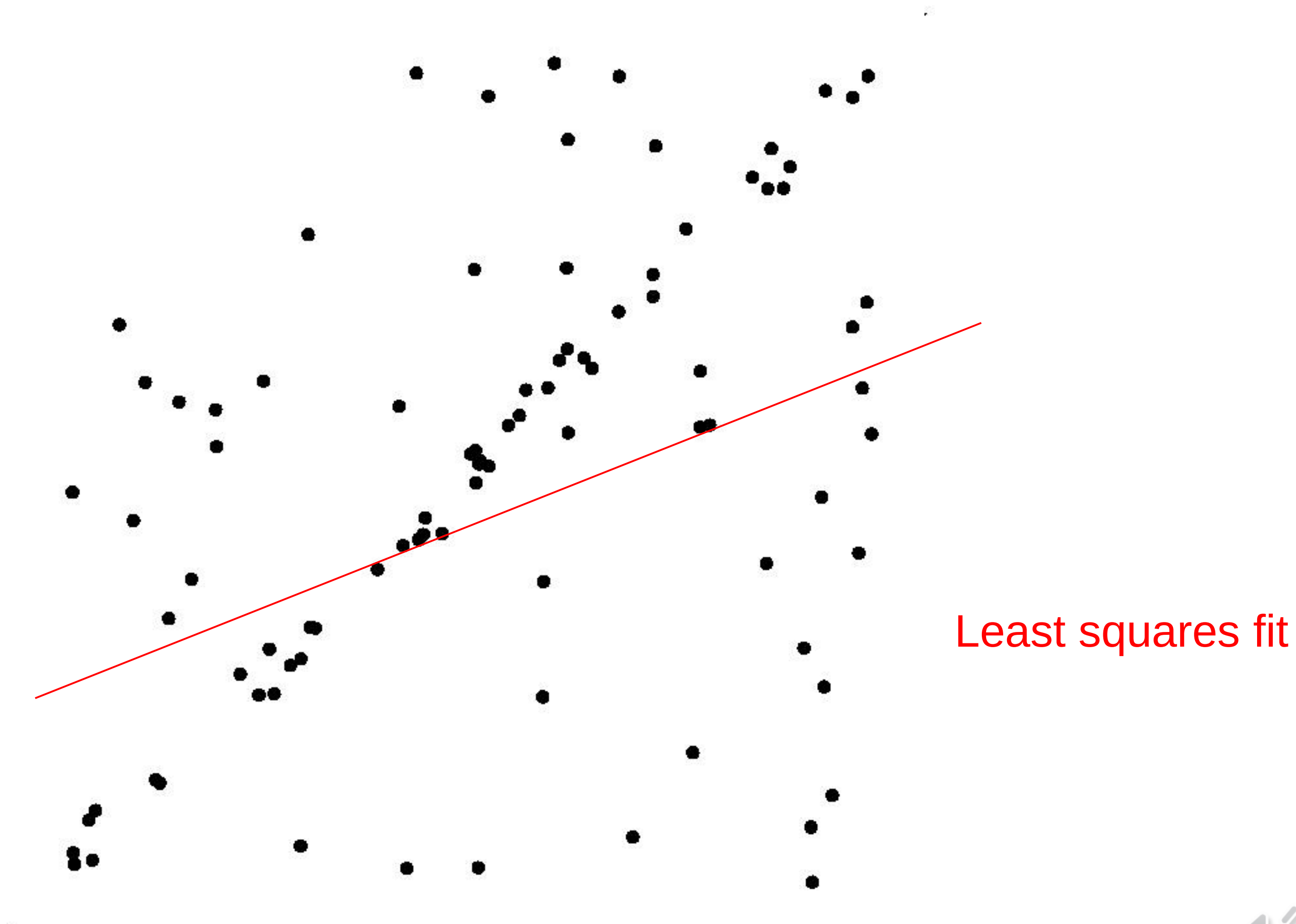
- Popular in computer vision
  - Finding correspondences
  - Robust to outliers

Robustness



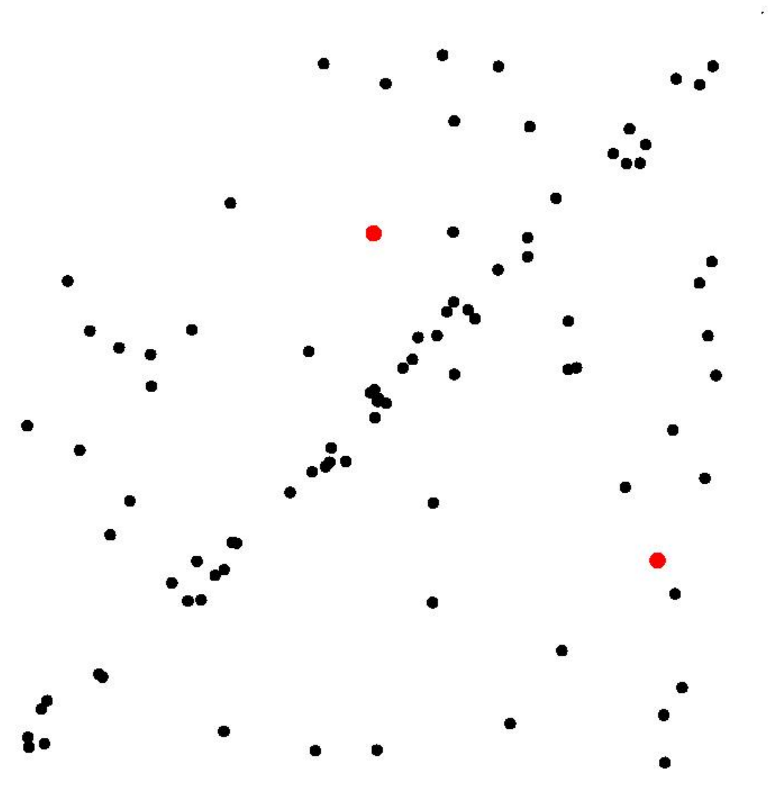
RANSAC – Algorithm

- Limitation in LS



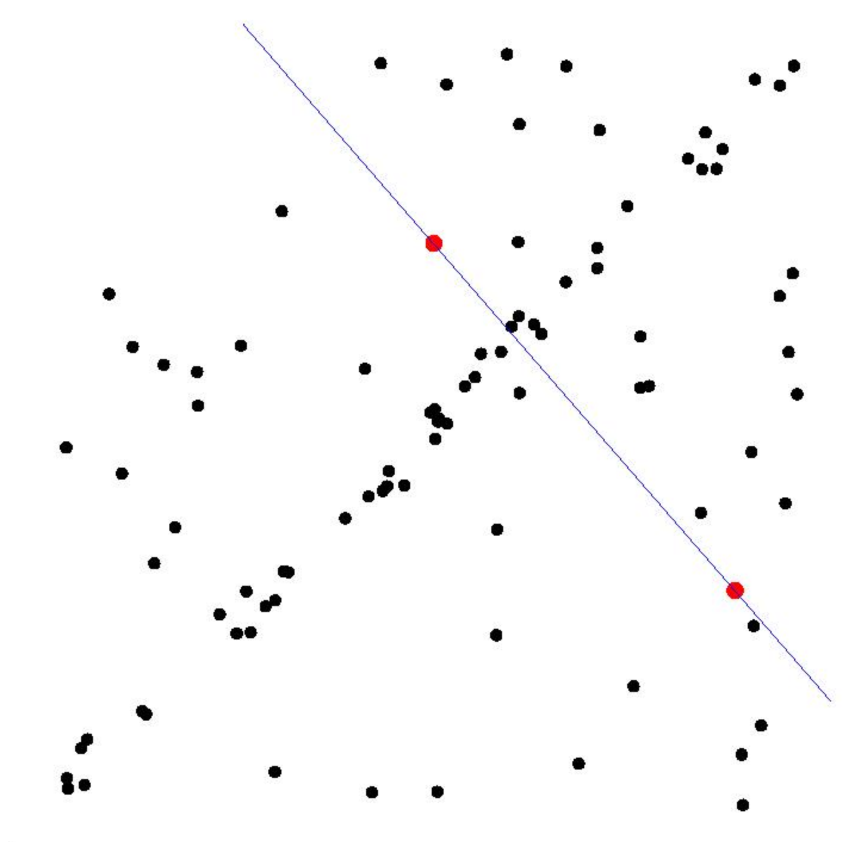
RANSAC – Algorithm

- Select m point (m = D.O.F)
  - Line = 2 DOF



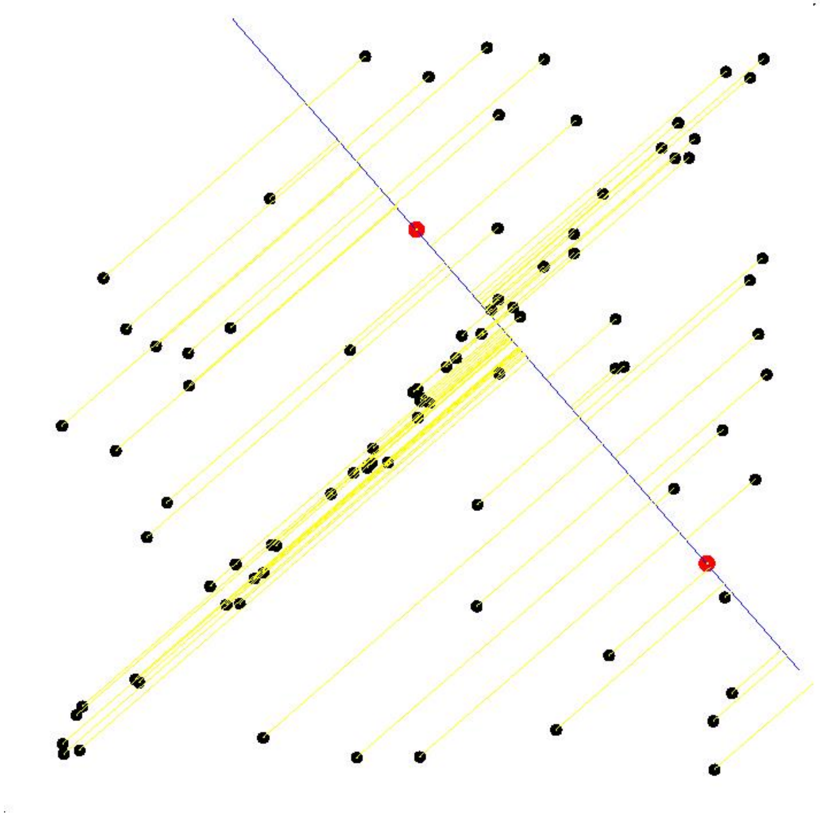
RANSAC – Algorithm

- Calculate model parameters that fit the data in the sample



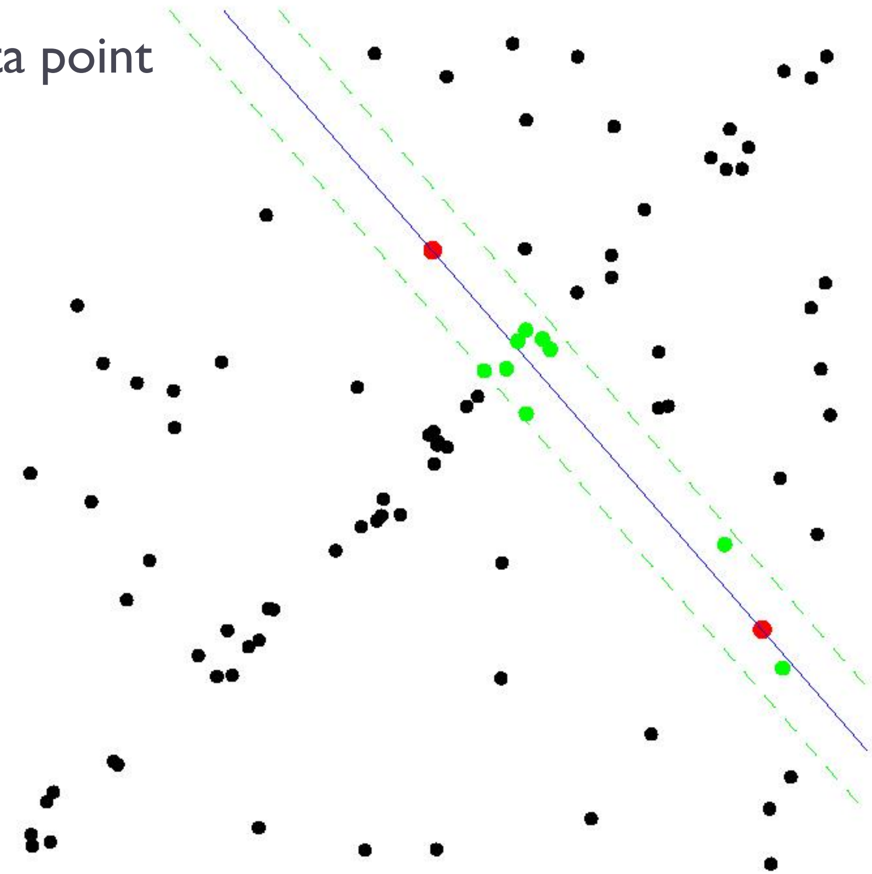
RANSAC – Algorithm

- Count "how may agrees"
  - Calculate error function for each data point



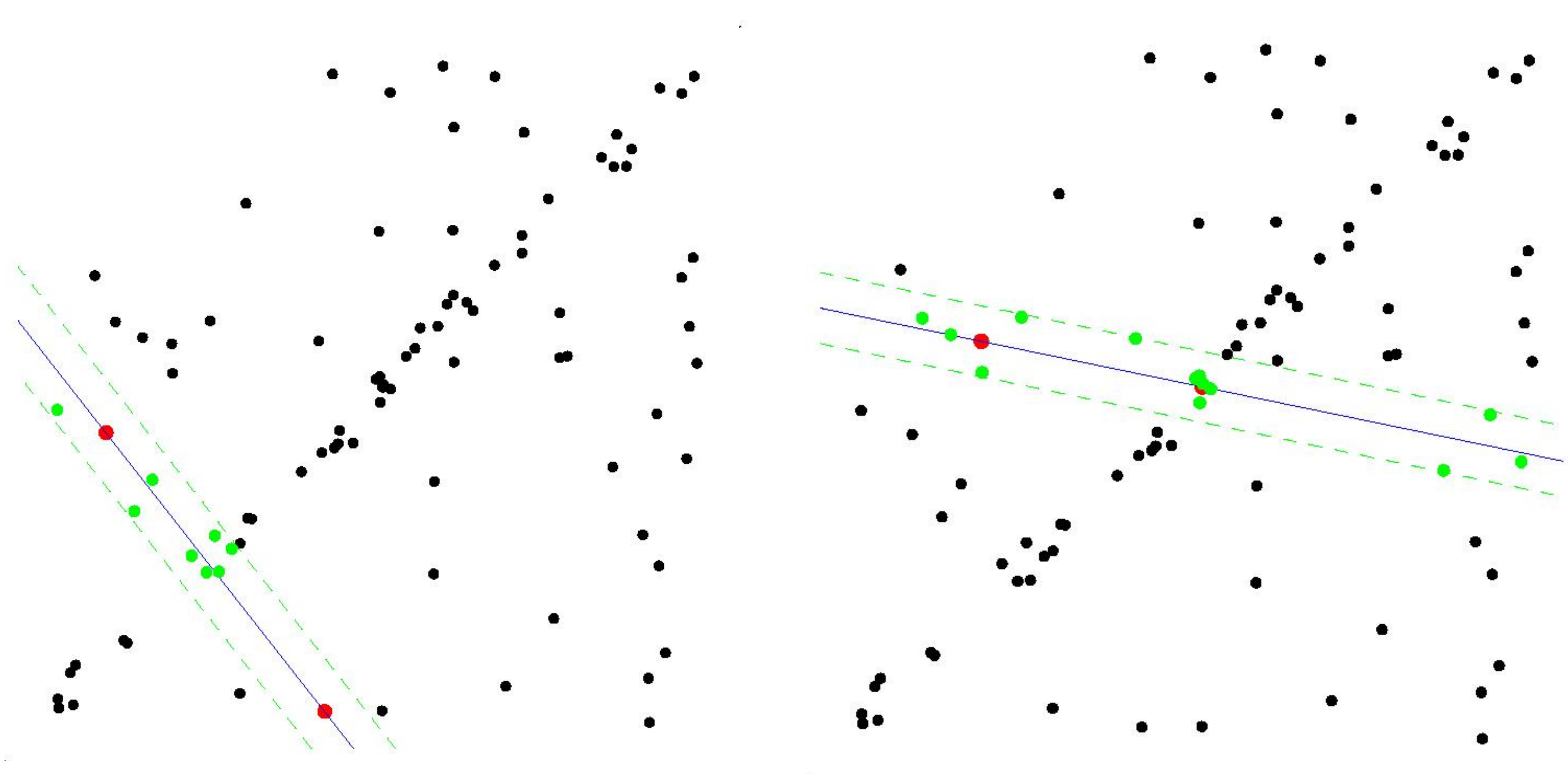
RANSAC – Algorithm

- Count "how may agrees"
  - Calculate error function for each data point
  - Select data that support this hypoethesis



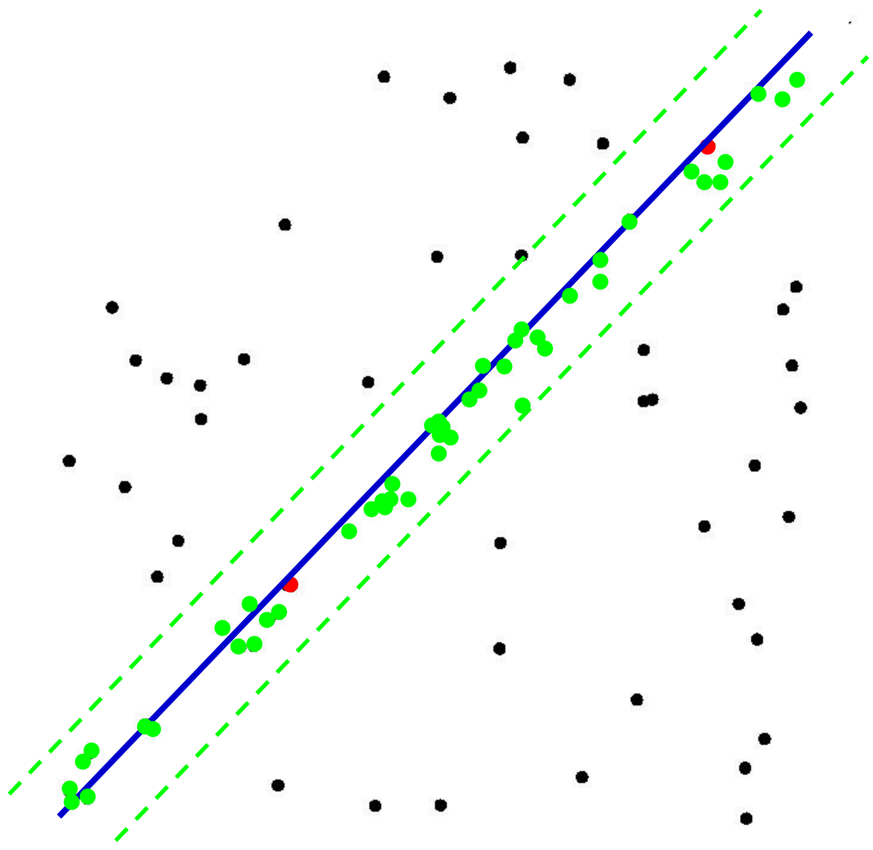
RANSAC – Algorithm

- Try with another sample



RANSAC – Algorithm

- Randomly repeating this
  - You'll hit the solution



RANSAC

- Algorithm

```
Given:
    data - a set of observed data points
    model - a model that can be fitted to data points
    n - the minimum number of data values required to fit the model
    k - the maximum number of iterations allowed in the algorithm
    t - a threshold value for determining when a data point fits a model
    d - the number of close data values required to assert that a model fits well to data
Return:
    bestfit - model parameters which best fit the data (or nil if no good model is found)
iterations = 0
bestfit = nil
besterr = something really large
while iterations < k {
    maybeinliers = n randomly selected values from data
    maybemodel = model parameters fitted to maybeinliers
    alsoinliers = empty set
    for every point in data not in maybeinliers {
        if point fits maybemodel with an error smaller than t
             add point to alsoinliers
    }
    if the number of elements in alsoinliers is > d {
        % this implies that we may have found a good model
        % now test how good it is
        bettermodel = model parameters fitted to all points in maybeinliers and alsoinliers
        thiserr = a measure of how well model fits these points
        if thiserr < besterr {
            bestfit = bettermodel
            besterr = thiserr
        }
    }
    increment iterations
}
return bestfit
```

How Many Iteration?

- Iteration number k
- It is a function of inlier ratio
  - Inlier ratio

$$w = \frac{\text{num of inliers}}{\text{num of samples}}$$

$$\begin{aligned} w^m &= \text{prob. that all points are inliers} \\ 1 - w^m &= \text{at least one point is outlier} \end{aligned}$$

  - We want to ensure prob p that select a inlier pair

$$\begin{aligned} 1 - p &= (1 - w^m)^k \\ k &= \frac{\log 1 - p}{\log(1 - w^m)} \end{aligned}$$

How Many Iteration?

- How many iteration

$$k = \frac{\log 1 - p}{\log(1 - w^m)}$$

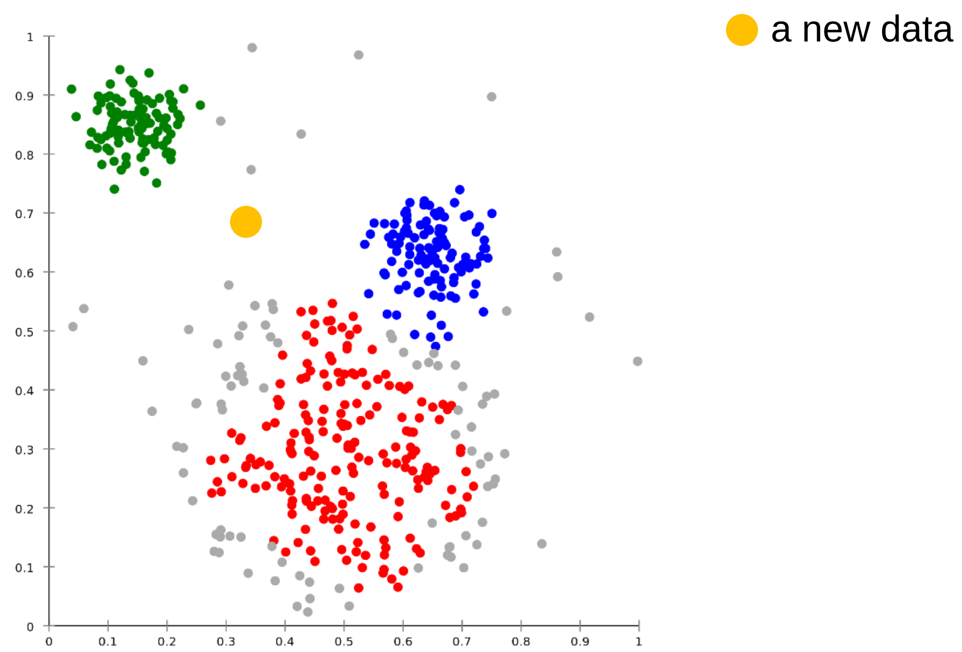
  - p = 99%, m = s (dof needed) = sample size

Sample size	Proportion of outliers ϵ						
s	5%	10%	20%	25%	30%	40%	50%
2	2	3	5	6	7	11	17
3	3	4	7	9	11	19	35
4	3	5	9	13	17	34	72
5	4	6	12	17	26	57	146
6	4	7	16	24	37	97	293
7	4	8	20	33	54	163	588
8	5	9	26	44	78	272	1177

Clustering

Clustering

- Grouping dataset into groups
  - 2D example of clustering



K-means Criteria

- Cluster map
  - A cluster map is a function

$$C : \{1, 2, \cdots, n\} \mapsto \{1, 2, \cdots, K\} \text{ that partitions the data into } K \text{ clusters}$$

  - Assumes
    - K is known
    - Adopt squared Euclidean distance $||x - y||^2 = \sum_{j=1}^{n} (x^{(i)} - y^{(i)})^2$
  - Objective
    - Seek to minimize the within cluster scatter

$$W(C) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i:C(i)=k} \left[\frac{1}{n_k} \sum_{j:C(j)=k} ||x_i - x_j||^2 \right] \qquad n_k = \sum 1_{C(i)=k}$$

avg distance to points in the same cluster

Problem Analysis

- Algorithm
  - The K-means criterion is a combinational optimization problem
  - The number of possible cluster map C is

$$\frac{1}{K!}\sum_{k=1}^{K}(-1)^{K-k}\binom{K}{k}k^{n}$$

by Jain and Dubes, 1988

  - If n=10 K =4 → 34,105
  - If n = 19 K = 4 → 10^10

- Bad news…
  - There is no known efficient search strategy for this space
  - Therefore we resort to an iterative suboptimal algorithm

K-means Clustering Algorithm

- Algorithm

Initialize $\bar{x}_k, k = 1, 2, \cdots, K$

Repeat

$$* \; C(i) = \arg\min{}_k ||x_i - \bar{x}_k||$$

$$* \; \bar{x}_k = \frac{1}{n_k} \sum x_i$$

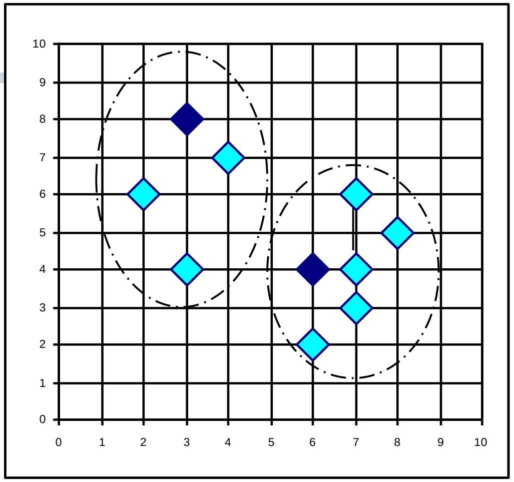
- Illustration
  - Select initial centroids at random.
  - Assign each object to the cluster with the nearest centroid.
  - Compute each centroid as the mean of the objects assigned to it.
  - Repeat previous steps until no change

Issues with K-means

- Issues
  - Local optimum
  - Simulated annealing and genetic algorithms for global optimum
  - Need to specify K, the number of clusters, in advance
  - Trouble with noisy data and outliers
  - Not suitable to discover clusters with non-convex shapes

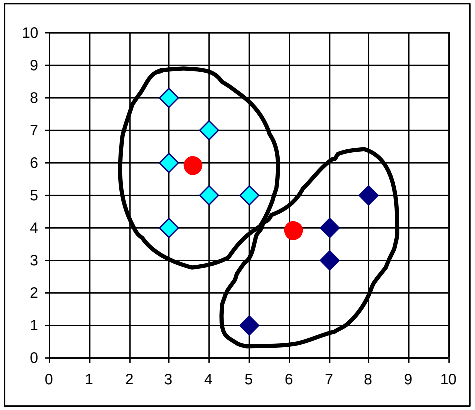
Other Partitioning Methods

- K-medoids (1990)
  - Instead of mean use median
  - For a data {13, 23, 11, 16, 15, 10, 26}
    - Mean = 16.28
    - Median = 15 (middle number)



K-medoids

- PAM (Partitioning Around Medoids)
  - Find K representative objects of the data set.
  - Each of the K objects is called a Medoid,
  - The most centrally located object within a cluster.



K-means

Density Based Clustering

- Two parameters:
  - ϵ : Maximum radius of neighborhood
  - MinPts: Minimum number of points in an Eps-neighborhood of a point

- Neighbor
  - $N_\epsilon(p) \quad = \quad \{q \in D | dist(p,q) \leq \epsilon\}$

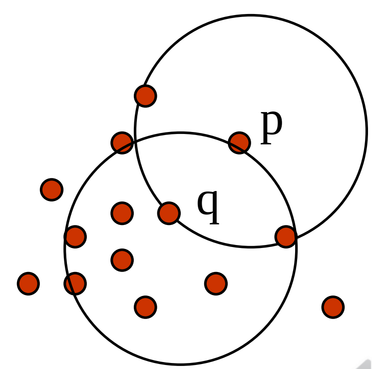
- Directly density-reachable
  - A point p is directly density-reachable from a point q iff

    1) p belongs to $N_\epsilon(q)$

    2) q is a core point $|N_\epsilon(q)| \geq$MinPts

MinPts = 5

Eps = 1 cm



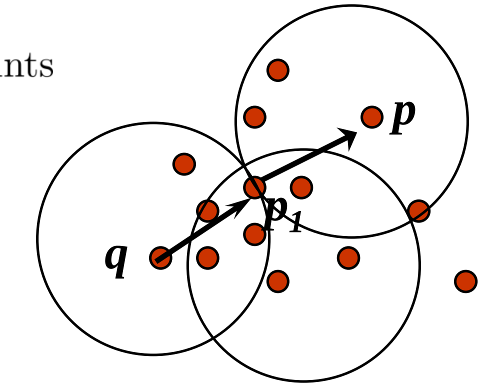
Density Based Clustering

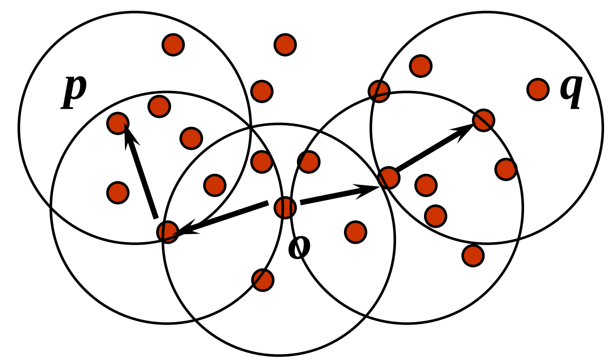
- ## Density-reachable

A point p is density-reachable from a point q if there is a chain of points

$$p_1, \cdots, p_n, p_1 = q, p_n = p$$

such that p_{i+1} is directly density-reachable from p_i





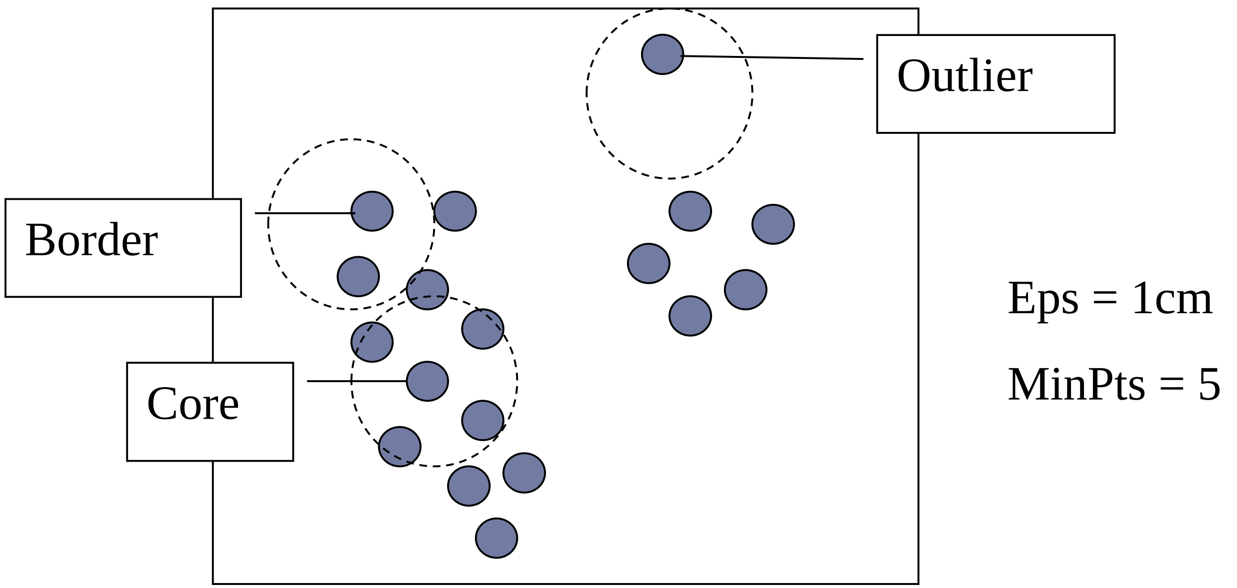
- ## Density-connected

A point p is density-connected to a point q

if there is a point o such that both, p and q are density-reachable from o

DBSCAN

- DBSCAN
  (Density Based Spatial Clustering of Applications with Noise)
  - Relies on a density-based notion of cluster
    - A cluster is defined as a maximal set of density-connected points
  - Discovers clusters of arbitrary shape in spatial databases with noise



DBSCAN

- The Algorithm
  - Arbitrarily select a point p
  - Retrieve all points density-reachable from p wrt Eps and MinPts.
  - If p is a core point, a cluster is formed.
  - If p is a border point, no points are density-reachable from p and DBSCAN visits the next point of the database.
  - Continue the process until all of the points have been processed.

Thank you very much !!