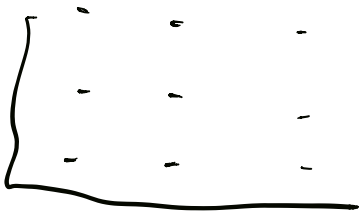
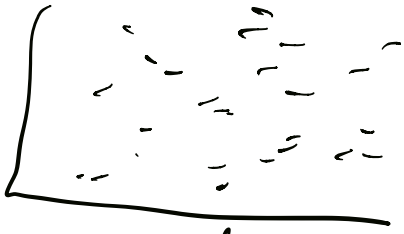


uniform
(a)



randomly generated (b)

(1) DBSCAN 在 (a) 和 (b) performance 怎樣?

(2) K-mean & DBSCAN 在 (b) 的 performance

Documentation data 的 similarity 方法
1. Jaccard 2. cosine 3. correlation

Similarity & differences:

1. Clustering & Association
   Vs.
   2. Simi
   2 Differences.

2. SOM
   SNN-Density based Vs. EM

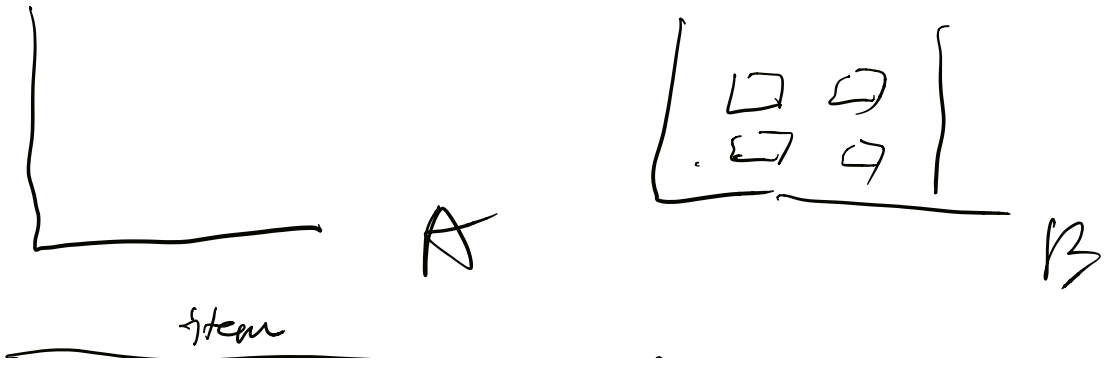
3. DENCLUE Vs. Fuzzy-K mean

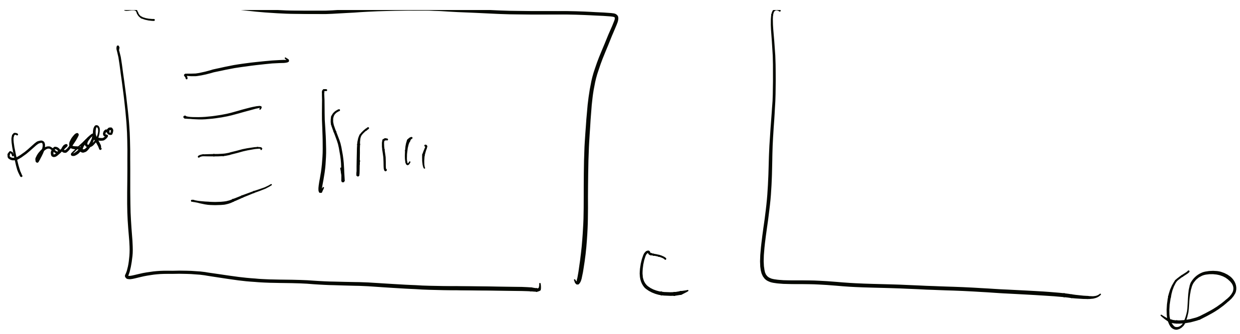
4. Anomaly: clustery-based vs.
Statistical - based

5. K-mean.

Classification.

1.





maximal frequent.

---

1. Clustering.

hierarchical or partitional.

fuzzy vs. crisp ?

complete or incomplete ?

— or — ?

3种.

评价 classification techniques 合不合适?

① 数据:

attribute are temporal [illegible]

② attributes are highly correlated

T & F

① some attributes have -1 correlation, does it negatively affect precision and recall

② if $corr(a_1, a_2) = 0.5$

$corr(a_2, a_3) = 0.5$

问 $corr(a_1, a_3) = ?$

③ Classification 2 classes.

其中一个占2%，问 猜哪个

entropy = 1

= 0.3

or = 0.

~~[redacted]~~

④ SEE 和 entrophy

entropy 这 1 is more accurate than 2

那么，SEE 高的一定会 indicate 相同的结果吗？

---

(2) Complete link.

(1) 画图　similarity
single link (min)

Sampling:

representative?　2题
Biased

① sample size = 10, representative?

②

③ 2 * n 一定超过 n 吗

④ None is true