

- homogeneity, completeness, V-measure.

- homogeneity: 각 cluster가 단일 클래스의 Data만 가지는 정도.
- completeness: 같은 클래스의 데이터가 하나의 클러스터로 모여있는 정도
- V-measure: homogeneity와 completeness의 조화 평균.

$$h = 1 - \frac{H[C|K]}{H[C]}$$

$$c = 1 - \frac{H[K|C]}{H[K]}$$

$$v = 2 \cdot \frac{h \cdot c}{h + c}$$

- $H[C]$: 클래스 엔트로피. 여러 클래스에 분산되어있는 경우.
- $H[C|K]$: 클래스당이 클래스의 엔트로피.
- $H[K]$: 클러스터 엔트로피. 여러 클러스터에 분산되어있는 경우.
- $H[K|C]$: 클래스별로 분류한 후, 클러스터 엔트로피.

2) 정확도 (클러스터의 갯수와 실제 갯수)를 보는 경우.

- 실루엣 계수 (Silhouette Coefficient)

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

- a : 같은 클러스터에 속한 원소들의 평균 거리.
- b : 다른 클러스터 중 가장 가까운 클러스터까지의 평균 거리.

Adaptive Propagation

- 모든 data가 특정된 기준에 따라 자신을 대표할 대표 데이터를 선택한다. 만약 스스로가 자기 자신을 대표하게 되면 클러스터의 중심이 된다.

- responsibility $r(i, k)$
 - k 번째 data가 i 번째 data의 대표가 되어야 하는 근거.
- availability $a(i, k)$
 - i 번째 data가 k 번째 data를 대표로 선택해야 하는 근거.
- 다음 수식을 수행할 때까지 반복.

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} (a(i, k') + s(i, k'))$$

$$a(i, k) \leftarrow \min(0, r(h, k) + \sum_{i' \neq i} r(i', k))$$

* $s(i, k)$ 는 유클리드 거리로 정의하는 값이다. $s(i, k) = -\|x_i - x_k\|^2$

- 위 Algorithm을 계산한 후, r, a 가 0이상 변하지 않고 수렴하면 계산이 종료되고, 종료 시점까지 $r(k, k) + a(k, k) > 0$ 이 data가 클러스터의 중심이 된다.