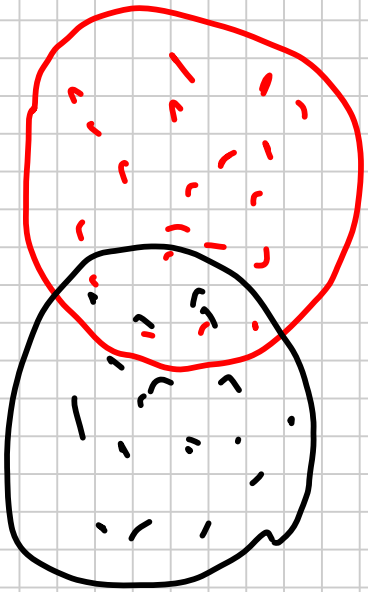


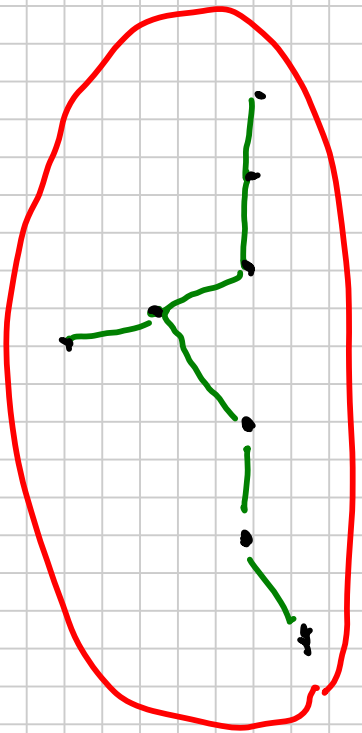
IR 19.5.04

Notizite!

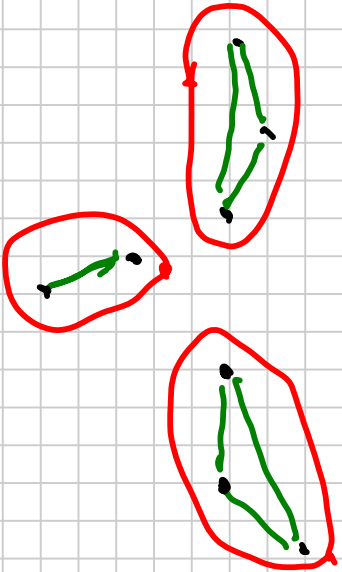
19.05.2004

Clustering

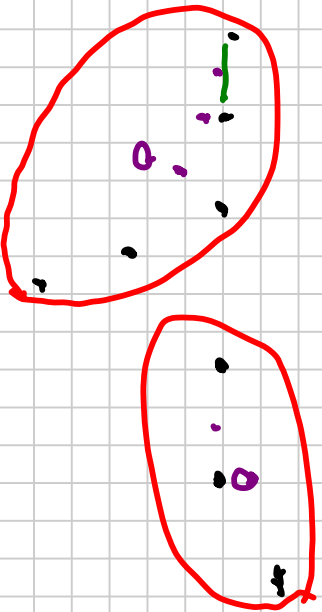




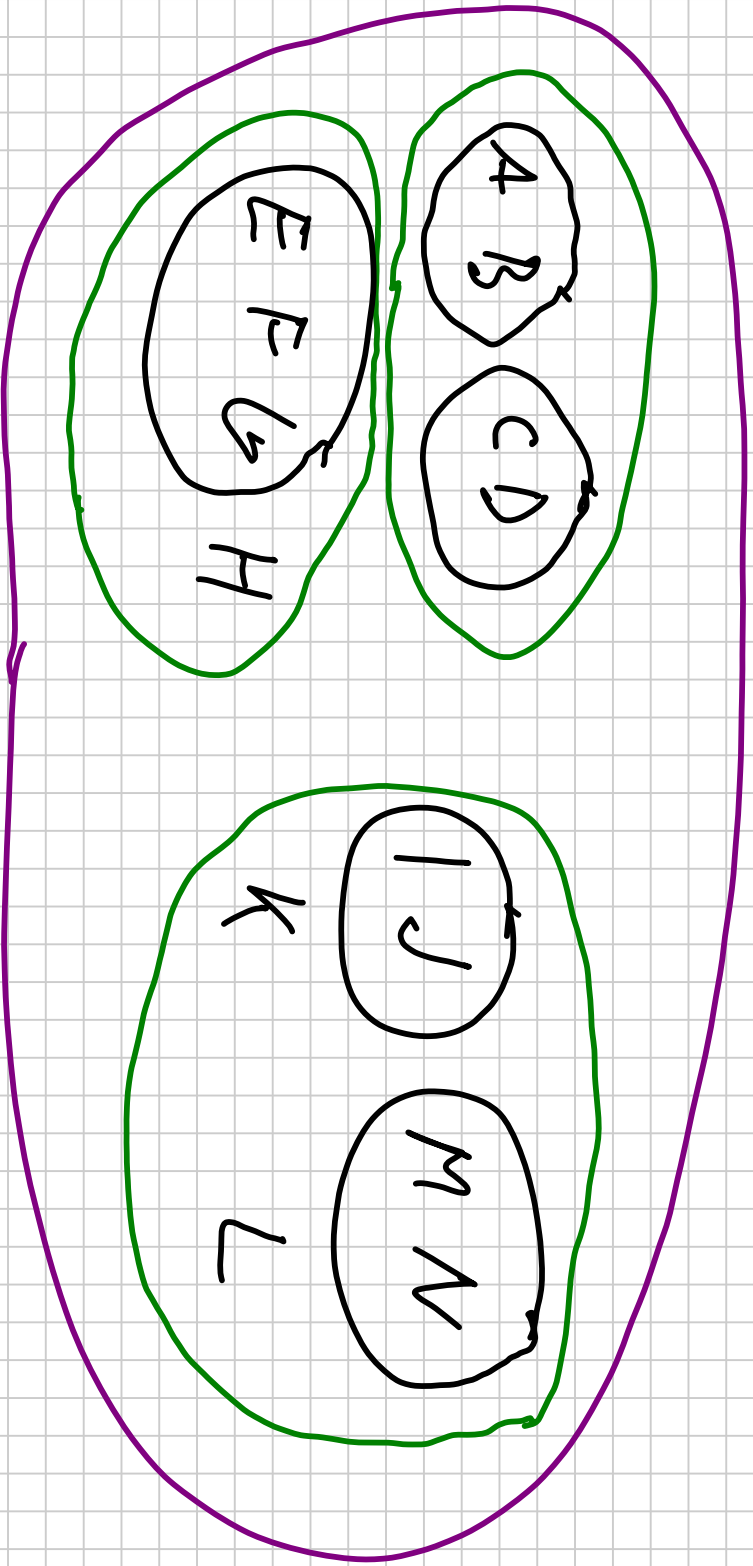
single link



complete link



average link



Probabilistisches Clustering

$$\text{Bayes: } P(a|B) = \frac{P(a \cap B)}{P(B)} = \frac{P(B|a) \cdot P(a)}{P(B)}$$

$$P(c^j|\vec{x}) = \frac{P(\vec{x}^T|c^j) \cdot P(c^j)}{P(\vec{x}^T)} = \frac{P(\vec{x}^T|c^j) \cdot P(c^j)}{\sum_{a=1}^k P(c^a) \cdot P(\vec{x}^T|c^a)}$$

$$P(\vec{x}^T|c^j) = \prod_i P(x_i|c^j) = \prod_{x_i=1} P(x_i=1|c^j) \cdot \prod_{x_i=0} P(x_i=0|c^j)$$

$P(x_i=1|c^j)$: W, dass x_i in einem beliebigen Dok
des Clusters c^j auftritt