

16 OCT

REVIEW: UNSUPERVISED LEARNING

K-MEANS

INPUT: SET OF PTS
"SEED" CENTROIDS
NUMBER K

[ASSIGNS PTS TO NEAREST CENT.
[RECOMPUTE CENTROID
→ LOOP UNTIL CONVG
(NO REASSIGN, MIN SSE,
ETC.)

ADV:

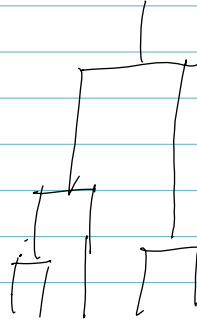
FAST
LOT OF EVID. OF PRACT USE

DISAD:

SENS TO OUTLIERS
SENS TO INIT SEEDS ↗
LOCAL OPTIMAL ↘
USER MUST SEL K

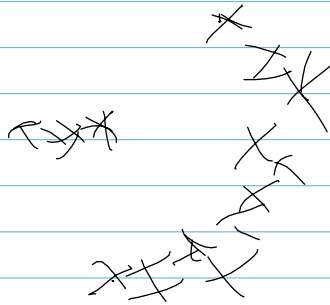
HIERARCH CLUSTERING

CLUSTER CONTAINS INSIDE
EACH OTHER



⊖ GENERAL n^2 RUNTIME

⊕ HANDLES CLUSTERS THAT
ARE CONCAVE



DATA STANDARDIZATION

NUMERIC ATTRIBUTES

RANGE (\min, \max)
Z-SCORE (σ, μ)

INTERVAL vs. RATIO

BINARY vs. NOMINAL

COMM FINDING

COMPONENT - BASED

MODULARITY MAXIMIZATION
(M)

GREEDY

LOUVAIN — FASTER,
MAXIMIZES
M BETTER

LABEL PROPAGATION

— VERY FAST

— LABEL PROP LIKE
IN DIFFUS.

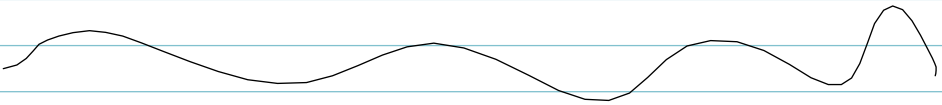
— DEGENER. SOLN,

$$P(D_1 | I) = 0.8$$

$$P(D_2 | I) = 0.5$$

$$P(\neg D_1 \wedge \neg D_2 | I) = 0.1$$

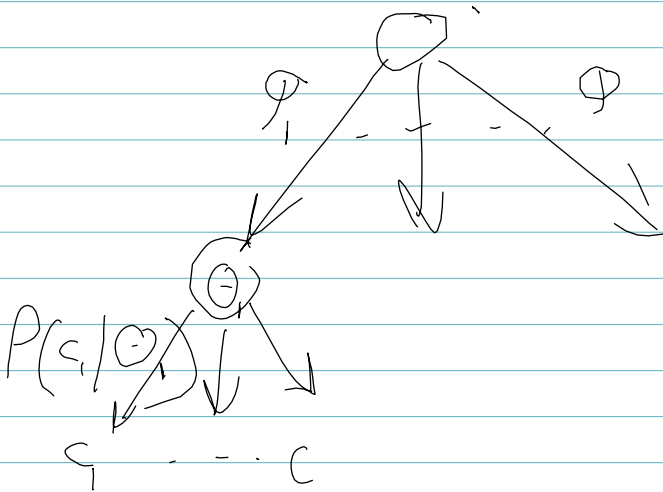
$(0.2 \cdot 0.5)$



$$P(D_2 | \neg D_1 \wedge I) = 0$$

$$p_j = P(c_j | \theta)$$

$$P(d_i | \theta) = \sum_{j=1}^{|C|} P(c_j | \theta) P(d_i | c_j, \theta)$$



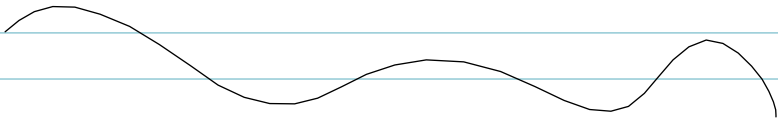
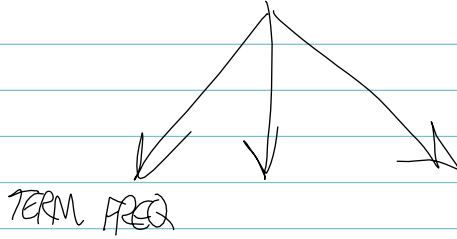
$$P(w | c, \theta)$$

~~$$P(d|c_j)$$~~

$$P(d|c_j, \Theta_i) =$$

A hand-drawn graph on lined paper. The vertical axis is labeled "LEARN" and the horizontal axis is labeled "STUDY". The axes are drawn as simple lines meeting at an origin. There are small tick marks at the ends of both axes. The word "LEARN" is written vertically along the y-axis, and "STUDY" is written horizontally along the x-axis.

PICK TOPIC



$$P(\theta_j | d)$$

$$P(w_i | \theta_j)$$

