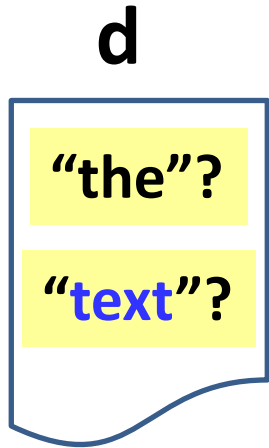


Generative Topic Model Revisited

Why can't this model be used for clustering?



w

$P(w | \theta_1)$

θ_1

the 0.04
mining 0.035
association 0.03
clustering 0.005
...
the 0.000001

$p(\theta_1) + p(\theta_2) = 1$

$P(\theta_1) = 0.5$

$P(w | \theta_2)$

θ_2

the 0.03
a 0.02
is 0.015
we 0.01
food 0.003
...
text 0.000006

Topic
Choice

$P(\theta_2) = 0.5$