

**Topic Modeling:-** Process to find the topics from documents in an unsupervised manner.

Also popularly known as documents to topic districution and topic to term distribution.

1. Document categorization.
2. Document summrizatiom
3. Dimensionality reduction
4. Informal retrival
5. Recommendation engines

**Techniques:**

LDA – Latent Dirichlet Allocation

NNMF - Non Negative Mattrix Factorization

LSA – Latent Semantic Allocation

LDA : is generative probablistics algorithm. Two important assumptoin documents are mixers of Topics and second topics are mixture of term.

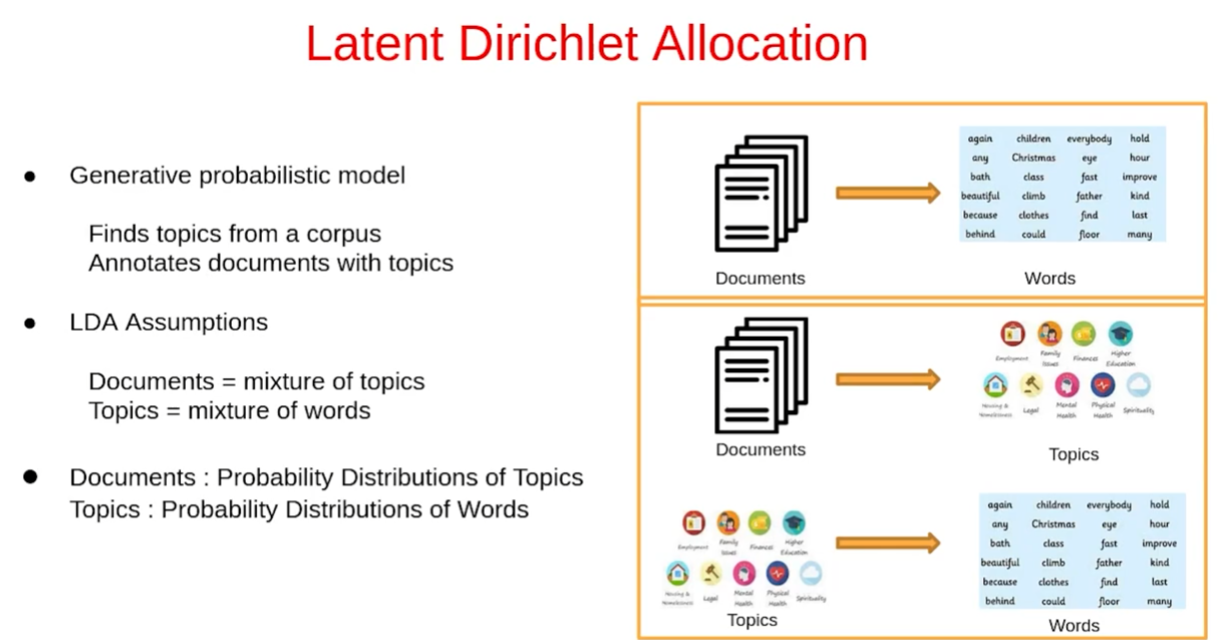
GPA – find topics from corpus and assign topics to documents

🡪 mixtures

Documents 🡪 topics 🡪 terms

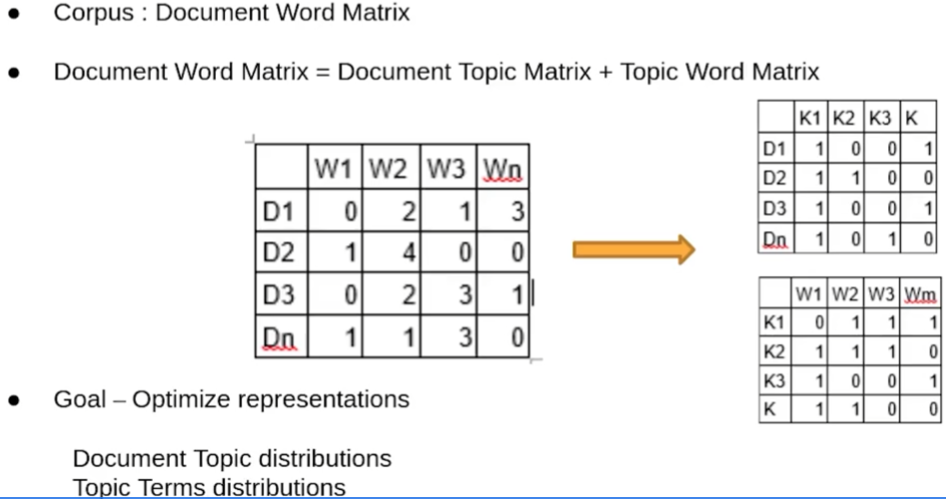
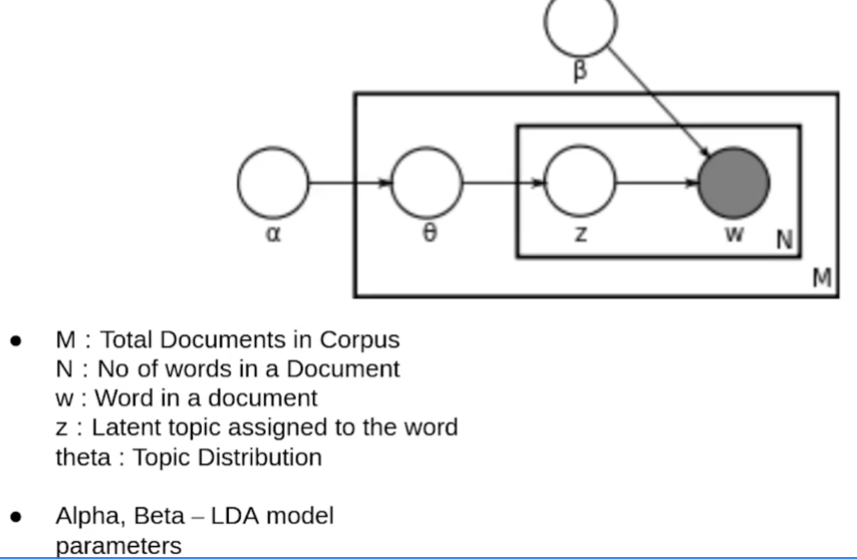
Documents = probablity distributions of topics

Topics = probablity distributions of words.



Corpus = document word matrix

document word matrix = document topic matrix + topic word matrix

Corpus : d1 = ( w1, w2,w3, ….wn)

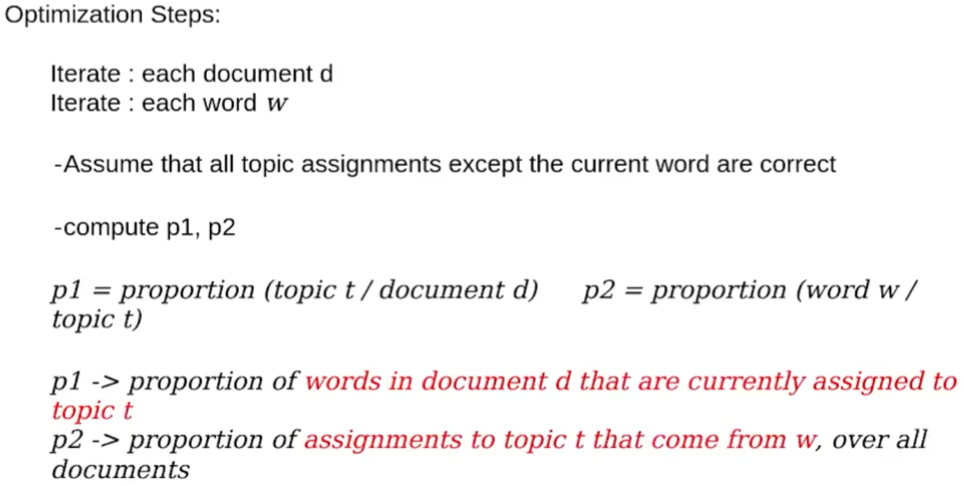
Firstly assign random topics to each word :

d1 = ( w1(k4), w2(k3),w3(k1), ….wn(k2))

two outcomes:-

Documents is mixture of topics.

Topics is mixture of terms.



* Reassign word w of documents d a new topic

Where we choose topics k wirh a new proability p = p1\* p2

* Repeated large number of times until steady state