Natural Language Processing DSECL ZG565

Dr. Chetana Gavankar, Ph.D,
IIT Bombay-Monash University Australia
Chetana.gavankar@pilani.bits-pilani.ac.in

BITS Pilani
Hyderabad Campus

innovate achieve lead

BITS Pilani
Pilani Campus

Session 7- Latent Dirichlet Allocation
Date – 6th January 2024

These slides are prepared by the instructor, with grateful acknowledgement of Jurafsky and Martin and many others who made their course materials freely available online.

Session Content

- Objective of Latent Dirichlet Allocation (LDA)
- Intuition behind LDA
- LDA Generative model
- Mathematical foundations for LDA : Bernoulli Trial, Binomial distribution, Multinomial distribution
- Mathematical foundations for LDA : Beta distribution, Conjugate Prior and Dirichlet distributions
- Dirichlet Visualization using Simplex
- Probabilistic Graphical LDA Model
- Mathematical modelling of LDA
- Gibbs Sampling Algorithm
- Case Study
- Implementing LDA in Python

Objectives of Topic Modelling

- Use these annotations to organize, summarize and search the documents.
- Topic Model can be defined as an unsupervised technique to discover topics across various text documents

Sample output from the LDA

- Four topics learned from the S&P 500 stock market data
- Goal is to find groups of stocks that tend to move together.

Topic 1	Topic 2	Topic 3	Topic 4
Southwestern Energy	Penneys	Capital One	Simon Property
Range Resources	Macys	BNY Mellon	Kimco Realty
Cabot Oil & Gas	Kohls	Discover	Equity Residential
EOG Resources	Nordstrom	Northern Trust	AvalonBay Communities
Chesapeake Energy	Target	Janus	Apartment Investment
Pioneer Resources	Limited	JPMorgan Chase	Vornado Realty Trust
Devon Energy	Lowes	State Street	Boston Properties
Peabody Energy	Home Depot	Wells Fargo	Public Storage
Anadarko Petroleum	American Express	PPL	Host Hotels
Massey Energy	Abercrombie	T. Rowe Price	HCP Inc.

- The topic model does not provide any label to these group of words.

Sample output from the LDA

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Genetics

Evolutionary biology

Data Analysis

Overall schematic

Each topic is defined as a Multinomial distribution over the vocabulary

Corpus

Documents($d_1 \ldots d_n$)

K-Topics (Hyper Parameter)

Each document has a distribution over K topics

Vocabulary($W_1 \ldots W_m$)

LDA Generative model

- Each **topic** is a distribution over words
- Each **document** is a mixture of corpus-wide topics
- Each **word** is drawn from one of these topics
- We only observe the words within the documents and the other structure are **hidden variables**.

The posterior distribution

LDA model

A 100-topic LDA model was fitted to **17,000 articles from the *Science*** journal.
At right are **the top 15 most frequent words** from the most frequent topics.
At left are the **inferred topic proportions** for the example article from previous slide.

"Genetics"	"Evolution"	"Disease"	"Computers"
human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

LDA Generative Process

LDA assumes that new documents are created in the following way:

1. Determine number of words in document
2. Choose a topic mixture for the document over a fixed set of topics (i.e. 20% topic A, 30% topic B, 50% topic C)
3. Generate the words in the document by:
   - First pick a topic based on the document's multinomial distribution above.
   - Next pick a word based on the topic's multinomial distribution.

LDA Generative Process

-Say we have a group of articles and we assume that all of those articles can be characterized by three topics: Animals, Cooking, and Politics.

-Each of those topics can be described by the following words:

- Animals: dog, chicken, cat, nature, zoo
- Cooking: oven, food, restaurant, plates, taste, delicious
- Politics: Republican, Democrat, Congress, ineffective, divisive

-Say we want to generate a new document that is 80% about animals and 20% about cooking.

1. We choose the length of the article (say, 1000 words)
2. We choose a topic based on our specified mixture (so, out of our 1000 words, roughly 800 will come from the topic "animals"
3. We choose a word based on the word distribution for each topic (i.e.

Intuition behind LDA

-Suppose you have a corpus of documents

-You want LDA to learn the topic representation of K topics in each document and the word distribution of each topic.

-LDA backtracks from the document level to identify topics that are likely to have generated the corpus.

Intuition behind LDA

1. Randomly assign each word in each document to one of the K topics.
2. For each document d:
   - Assume that all topic assignments except for the current one are correct.
   - Calculate two proportions:
     1. Proportion of words in document d that are currently assigned to topic t = p(topic t | document d)
     2. Proportion of assignments to topic t over all documents that come from this word w = p(word w | topic t)
   - Multiply those two proportions and assign w a new topic based on that probability. p(topic t | document d) * p(word w | topic t)
3. Eventually we'll reach a steady state where assignments make sense

Bernoulli Trial

- Any single trial with two possible outcomes can be modeled as a Bernoulli trial: team wins/loses, pitch is a strike/ball, coin comes up heads or tails, etc.
- A Bernoulli trial uses Bernoulli distribution to calculate the probability of either outcome.

Bernoulli trial

Bernoulli: A Special Case of the Binomial Distribution

Binomial Trial: Chance of getting n heads in a row(n=3)

Bernoulli Trial: Chance of getting a heads on a single flip

Bernoulli - Distribution Notation

- The probability mass function of the Bernoulli distribution is

$$f(x)=P(X=k)=\Theta^{k}(1-\Theta)^{1-k}, \qquad k=\{0,1\}$$

- The only parameter of the bernoulli distribution is θ, which defines the probability of success during a bernoulli trial.

Binomial distribution

Binomial distribution

If p is the probability of heads, the probability of getting exactly k heads in n independent yes/no trials is given by the binomial distribution $Bin(n,p)$:

$$
\begin{aligned}
P(k \text{ heads}) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}
\end{aligned}
$$

Multinomial Distribution

- Suppose we roll our die of words having k sides(vocabulary) where each side takes probabilities $\Theta_1, \ldots, \Theta_k$ respectively.
- Probability mass function

$$f(x) = \frac{n!}{x_1! x_2! \cdots x_k!} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_k^{x_k}$$

k - number of sides on the die
n - number of times the die will be rolled

Multinomial Distribution

Suppose that we observe an experiment that has k possible outcomes $\{O_1, O_2, \ldots, O_k\}$ independently n times.

Let $p_1, p_2, \ldots, p_k$ denote probabilities of $O_1, O_2, \ldots, O_k$ respectively.

Let X_i denote the number of times that outcome O_i occurs in the n repetitions of the experiment.

Then the joint probability function of the random variables $X_1, X_2, \ldots, X_k$ is

$$p(x_1, \ldots, x_n) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

Multinomial Distribution

<u>Note:</u> $p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$

is the probability of a sequence of length n containing

x_1 outcomes O_1

x_2 outcomes O_2

$\dots$

x_k outcomes O_k

$$\frac{n!}{x_1!x_2!\dots x_k!} = \binom{n}{x_1 \quad x_2 \quad \dots \quad x_k}$$

is the number of ways of choosing the positions for the x_1 outcomes O_1, x_2 outcomes O_2, …, x_k outcomes O_k

Beta Distribution

- The probability distribution function for the beta distribution

$$f(\theta;\alpha,\beta) = \frac{\theta^{(\alpha-1)}(1-\theta)^{(\beta-1)}}{B(\alpha,\beta)}$$

Beta Distribution

- The beta distribution can be thought of as a probability distribution of probabilities.

The Building Blocks of inferring the parameters

- Parameter estimation

$$\underbrace{p(\theta|D)}_{posterior} = \frac{\overbrace{p(D|\theta)}^{likelihood}\,\overbrace{p(\theta)}^{prior}}{\underbrace{p(D)}_{evidence}}$$

Conjugate Prior

In Bayesian probability theory,

- if the posterior distributions $p(\theta \mid x)$ are in the same probability distribution family as the prior probability distribution $p(\theta)$, the prior and posterior are then called **conjugate distributions,** and the prior is called a **conjugate prior** for the likelihood function

$$\underbrace{p(\theta|D)}_{posterior} = \frac{\overbrace{p(D|\theta)}^{likelihood}\,\overbrace{p(\theta)}^{prior}}{\underbrace{p(D)}_{evidence}}$$

- conjugate prior gives a closed-form expression for the posterior; otherwise numerical integration may be necessary.
- Also may give intuition, by more transparently showing how a likelihood function updates a prior distribution

Beta distribution as Conjugate Prior for Binomial distribution

Given a **prior** $P(\theta \mid \alpha,\beta) = \text{Beta}(\alpha,\beta)$, and **data** $D=(H,T)$, what is our posterior?

$$
\begin{aligned}
P(\theta|\alpha,\beta,H,T) &\propto P(H,T|\theta)P(\theta|\alpha,\beta) \\
&\propto \theta^H(1-\theta)^T\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \theta^{H+\alpha-1}(1-\theta)^{T+\beta-1}
\end{aligned}
$$

With normalization

$$
\begin{aligned}
P(\theta|\alpha,\beta,H,T) &= \frac{\Gamma(H+\alpha+T+\beta)}{\Gamma(H+\alpha)\Gamma(T+\beta)}\theta^{H+\alpha-1}(1-\theta)^{T+\beta-1} \\
&= \text{Beta}(\alpha+H,\beta+T)
\end{aligned}
$$

Dirichlet distributions

- Dirichlet distributions are probability distributions over multinomial parameter vectors
  - They are called Beta distributions when k = 2
- The Dirichlet probability density function is defined as

$$Dir(\vec{\theta}|\vec{\alpha}) = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}\theta_i^{\alpha_i-1}$$

- $$Dir(\vec{\theta}|\vec{\alpha}) = \frac{1}{B(\alpha)}\prod_{i=1}^{K}\theta_i^{\alpha_i-1}$$ where $$\frac{1}{B(\alpha)} = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}$$

Visualization of the simplex

- This is often referred as simplex and a most convenient way to visualize this is using a certain shapes depending upon the number of topics.
- Suppose K=2 topics which can be modeled as 1-simplex and can be visualized using a line.

Visualization of the simplex

- Suppose K=3 topics which can be modeled as 2-simplex and can be visualized using a triangle.

- If we have K topics this can be generated using K-1 simplex.

Shape of the Dirichlet distribution

Dirichlet(0.999, 0.999, 0.999)

Dirichlet(2, 5, 15)

Dirichlet(5, 5, 5)

Dirichlet distributions

Each point on a k dimensional simplex is a multinomial probability distribution:

Statistical Inference

Topic Modelling Example - LDA

Science Politics Sports

LDA

Sports
Sports
Sports
Sports
Science
Science
Science
Science
Politics
Politics
Science
Politics

Quiz: Which one for topics?

Quiz: Which one for topics?

science	politics	sports
0.07	0.03	0.9

science	politics	sports
0.45	0.05	0.5

science	politics	sports
0.8	0.1	0.1

science	politics	sports
0.05	0.8	0.15

science	politics	sports
0.05	0.9	0.05

A distribution of distributions

More topics? More dimensions

2 topics 3 topics 4 topics …

n-dimensional simplex

Word distributions

Galaxy	Planet	Ball	Referendum
0.4	0.4	0.1	0.1

Galaxy	Planet	Ball	Referendum
0.3	0.1	0.5	0.1

Galaxy	Planet	Ball	Referendum
0.1	0.1	0.1	0.7

Two Dirichlet distributions

Documents-Topics

Topics-Words

Blueprint for the LDA machine

Mathematical modelling of LDA

Probability of a document

$$P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta) = \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j)\, P(W_{j,t} \mid \varphi_{Z_{j,t}})$$

Mathematical modelling of LDA

$$\prod_{j=1}^{M} P(\theta_j;\alpha)$$

science	politics	sports
0.7	0.1	0.2

$$\prod_{i=1}^{K} P(\varphi_i;\beta)$$

Galaxy	Planet	Ball	Referendum
0.4	0.4	0.1	0.1

Galaxy	Planet	Ball	Referendum
0.1	0.1	0.1	0.7

Galaxy	Planet	Ball	Referendum
0.3	0.1	0.5	0.1

$$\prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j)$$

$$P(W_{j,t} \mid \varphi_{Z_{j,t}})$$

Topics		Words
science	→	planet
science	→	galaxy
sports	→	ball
science	→	planet
science	→	galaxy
politics	→	referendum
sports	→	galaxy
sports	→	ball
science	→	referendum

Mathematical modelling of LDA

$$\prod_{j=1}^{M} P(\theta_j;\alpha)$$

science	politics	sports
0.7	0.1	0.2

$$\prod_{i=1}^{K} P(\varphi_i;\beta)$$

Galaxy	Planet	Ball	Referendum
0.4	0.4	0.1	0.1

Galaxy	Planet	Ball	Referendum
0.1	0.1	0.1	0.7

Galaxy	Planet	Ball	Referendum
0.3	0.1	0.5	0.1

$$\prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j)$$

Topics
science
science
sports
science
science
politics
sports
sports
science

$$P(W_{j,t} \mid \varphi_{Z_{j,t}})$$

Mathematical modelling of LDA

$$\prod_{j=1}^{M} P(\theta_j; \alpha) \qquad \prod_{i=1}^{K} P(\varphi_i; \beta) \qquad \prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j) \qquad P(W_{j,t} \mid \varphi_{Z_{j,t}})$$

P(same articles) = low

Mathematical modelling of LDA

$$\prod_{j=1}^{M} P(\theta_j;\alpha) \qquad \prod_{i=1}^{K} P(\varphi_i;\beta) \qquad \prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j) \qquad P(W_{j,t} \mid \varphi_{Z_{j,t}})$$

P(same articles) = low

P(same articles) = very low

Best settings on the machine

Mathematical modelling of LDA

The winning arrangements

Documents-Topics

Topics-Words

Lengh of the articles?
Poisson distribution

Latent Dirichlet Allocation

Directed graphical model

Graphical LDA Model

Our goal is to **infer** or **estimate** the hidden variables, i.e. computing their distribution conditioned on the documents. $\longrightarrow$ $p(topics,\ proportions,\ assignments \mid documents)$

- Nodes are RVs; edges indicate dependence.
- Shaded nodes are observed, and unshaded nodes are hidden.
- Plates indicate replicated variables.

Graphical LDA Model

K – total number of topics
β_k – topic, a distribution over the vocabulary
D – total number of documents
Θ_d – per-document topic proportions
N – total number of words in a document (it fact, it should be N_d)
$Z_{d,n}$ – per-word topic assignment
$W_{d,n}$ – observed word
α, η – Dirichlet parameters

- Several **inference algorithms** are available (e.g. sampling based)
- A few **extensions** to LDA were created:
  - Bigram Topic Model

Graphical LDA Model

$$p(\beta,\theta,\mathbf{z},\mathbf{w}) = \left(\prod_{i=1}^{K} p(\beta_i|\eta)\right)\left(\prod_{d=1}^{D} p(\theta_d|\alpha)\prod_{n=1}^{N} p(z_{d,n}|\theta_d)p(w_{d,n}|\beta_{1:K},z_{d,n})\right)$$

1) Draw each topic $\beta_i \sim Dir(\eta)$ for $i = 1, \ldots, K$

2) For each document:

First, Draw topic proportions $\theta_d \sim Dir(\alpha)$

For each word within the document:

a) Draw $Z_{d,n} \sim Multi(\theta_d)$

b) Draw $W_{d,n} \sim Multi(\beta_{z_{d,n}})$

LDA Model

This joint distribution defines a posterior $p(\theta, z, \beta \mid w)$.

From a collection of documents we have to infer:

1. Per-word topic assignment $z_{d,n}$.
2. Per-document topic proportions θ_d.
3. Per-corpus topic distributions β_k.

Then use posterior expectations ($E\{\beta|w\}$ for the corpus, $E\{\theta_d|w\}$ for each document) to perform the task at hand: information retrieval, document similarity, exploration, and others.

Train LDA - Gibbs Sampling

Science Politics Sports

Assign Topics

Doc 1
ball
ball
ball
planet
galaxy
Doc 2
referendum
planet
planet
referendum
referendum
Doc 3
planet
planet
galaxy
planet
ball
Doc 4
planet
galaxy
referendum
planet
ball
Topic 1
Topic 2
Topic 3

Doc 1
ball Topic 3
ball Topic 3
ball Topic 3
plane Topic 1
galax Topic 3
Doc 2
referen Topic 2
plane Topic 1
plane Topic 2
referen Topic 2
referen Topic 2
Doc 3
plane Topic 1
plane Topic 1
galax Topic 1
plane Topic 1
ball Topic 3
Doc 4
plane Topic 1
galax Topic 1
referenc Topic 2
plane Topic 1
ball Topic 3
Topic 1
Topic 2
Topic 3

Doc 1
ball
ball
ball
planet
galaxy
Doc 2
referendum
planet
planet
referendum
referendum
Doc 3
planet
planet
galaxy
planet
ball
Doc 4
planet
galaxy
referendum
planet
ball
80% Topic 3
20% Topic 1
80% Topic 2
20% Topic 1
80% Topic 1
20% Topic 3
60% Topic 1
20% Topic 2
20% Topic 3
Topic 1
Topic 2
Topic 3

Property 1:
Articles are as monochromatic as possible

Doc 1
ball
ball
ball
planet
galaxy
Doc 2
referendum
planet
planet
referendum
referendum
Doc 3
planet
planet
galaxy
planet
ball
Doc 4
planet
galaxy
referendum
planet
ball
Articles
Words
planet
planet
planet
planet
planet
planet
planet
planet
referendum
referendum
referendum
referendum
ball
ball
ball
ball
ball
galaxy
galaxy
galaxy

Property 2:
Words are as monochromatic as possible

Goal: Color each word with blue, green, red

1. Each article is as monochromatic as possible
2. Each word is as monochromatic as possible

ball

Topic 1 | Topic 2 | Topic 3

How much is Topic 1 in Doc 1?

2

How much is 'ball' in Topic 1?

0

Product: 0

How much is Topic 2 in Doc 1?

0

How much is 'ball' in Topic 2?

1

Product: 0

How much is Topic 3 in Doc 1?

2

How much is 'ball' in Topic 3?

3

Product: 6

Topic 1

How much is Topic 1 in Doc 1?

$2+\alpha$

How much is 'ball' in Topic 1?

$0+\beta$

Topic 2

How much is Topic 2 in Doc 1?

$0+\alpha$

How much is 'ball' in Topic 2?

$1+\beta$

Topic 3

How much is Topic 3 in Doc 1?

$2+\alpha$

How much is 'ball' in Topic 3?

$3+\beta$

Topic 1	Topic 2	Topic 3
How much is Topic 1 in Doc 1?	How much is Topic 2 in Doc 1?	How much is Topic 3 in Doc 1?
$2+\alpha$	$0+\alpha$	$2+\alpha$
How much is 'ball' in Topic 1?	How much is 'ball' in Topic 2?	How much is 'ball' in Topic 3?
$0+\beta$	$1+\beta$	$3+\beta$

Doc 1
ball
ball
ball
planet
galaxy
Doc 2
referendum
planet
planet
referendum
referendum
Doc 3
planet
planet
galaxy
planet
ball
Doc 4
planet
galaxy
referendum
planet
ball
80% Topic 3
20% Topic 1
80% Topic 2
20% Topic 1
80% Topic 1
20% Topic 3
60% Topic 1
20% Topic 2
20% Topic 3
Science
Topic 1
planet (7)
galaxy (2)
Politics
Topic 2
referendum (4)
planet (1)
Sports
Topic 3
ball (5)
galaxy (1)

Probability of a document

$$P(\mathbf{W}, \mathbf{Z}, \theta, \phi; \alpha, \beta) = \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{i=1}^{K} P(\phi_i; \beta) \prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j)\, P(W_{j,t} \mid \phi_{Z_{j,t}})$$

Topics Words

Gibbs sampling

Gibbs Sampling Algorithm

- Step1: Assign a random topic [1…T] for each word
- Step2: For each word token, a new topic is sampled as per $P(z_i=j|z_{-I},w_i,d_i)$ and the matrices C_{wt} (word-topic) and C_{dt} (document-topic) are updated.
- One iteration over all word token in the document is a Gibbs Sample
- Each iteration may have correlation with the next hence these samples are saved at spaced intervals.

LDA Summary

Documents are probability distributions over latent topics.

Topics are probability distributions over words.

LDA takes a number of documents. It assumes that the words in each document are related. It then tries to figure out the "recipe" for how each document could have been created. We just need to tell the model how many topics to construct and it uses that "recipe" to generate topic and word distributions over a corpus. Based on that output, we can identify similar documents within the corpus.

LDA Summary

ADVANTAGES

LDA is an effective tool for topic modeling.

Easy to understand conceptually

Has been shown to produce good results over many domains.

New applications

LIMITATIONS

Must know the number of topics K in advance

Dirichlet topic distribution cannot capture correlations among topics

Difference between document clustering and topic modeling

Topic Modeling

T_1	T_2	. . .	T_m
$Word_{11}, pr_{11}$	$Word_{21}, pr_{21}$	. . .	$Word_{m1}, pr_{m1}$
$Word_{12}, pr_{12}$	$Word_{22}, pr_{22}$	. . .	$Word_{m2}, pr_{m2}$
.	.	.	.
.	.	.	.
.	.	.	.
$Word_{1k}, pr_{1k}$	$Word_{2k}, pr_{2k}$	. . .	$Word_{mk}, pr_{mk}$

Document Clustering

C_1	C_2	. . .	C_m
D_a	D_i	. . .	D_p
D_b	D_j	. . .	D_q
.	.	.	.
.	.	.	.
.	.	.	.
D_c	D_k	. . .	D_s

https://iksinc.online/2016/05/16/topic-modeling-and-document-clustering-whats-the-difference/

Implementation Tools

Tooling

gensim: topic modeling for humans
- Free python library
- Memory independent
- Distributed computing

http://radimrehurek.com/gensim

Stanford Topic Modeling Toolbox

http://nlp.stanford.edu/software/tmt

MAchine **L**earning for **L**anguag**E** **T**oolkit (MALLET) is a Java-based package for:
- statistical natural language processing
- document classification
- Clustering
- topic modeling
- information extraction
- and other machine learning applications to text.

http://mallet.cs.umass.edu

References

Bernoulli trial, binomial and multinomial distribution:

https://www.askiitians.com/iit-jee-algebra/probability/bernoulli-trials-and-binomial-distribution/

Beta Distribution:

- https://www.youtube.com/watch?v=v1uUgTcInQk&feature=emb_logo

Conjugate Prior:

- https://www.youtube.com/watch?time_continue=2&v=aPNrhR0dFi8&feature=emb_logo
- https://www.youtube.com/watch?v=qpNAXnmy0GU

Topic Models

- https://www.youtube.com/watch?v=fCmIceNqVog

Gibbs Sampling

- https://www.youtube.com/watch?v=u7l5hhmdc0M
- https://medium.com/analytics-vidhya/topic-modeling-using-lda-and-gibbs-sampling-explained-49d49b3d1045

References

LDA

- [1] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation.“ Journal of machine Learning research (2003): 993-1022.
- https://www.youtube.com/watch?v=3mHy4OSyRf0
- https://www.coursera.org/learn/text-mining/lecture/dmpQ0/2-5-topic-mining-and-analysis-motivation-and-task-definition
- https://www.youtube.com/watch?v=NYkbqzTlW3w
- https://github.com/adashofdata/nlp-in-python-tutorial
- https://www.youtube.com/watch?time_continue=3&v=Cpt97BpI-t4&feature=emb_logo
- https://github.com/bhattbhavesh91
- https://www.youtube.com/watch?v=T05t-SqKArY
- https://www.youtube.com/watch?v=BaM1uiCpj_E
- https://livebook.manning.com/book/grokking-machine-learning/