9. Unsupervised Techniques III: Topic Models (flipped)

DS-GA 1015, Text as Data
Arthur Spirling

April 6, 2021

Housekeeping

Housekeeping

1 HW2 graded, back soon. HW3 on the way.

Housekeeping

1 HW2 graded, back soon. HW3 on the way.

2 Guide to final paper now in <Resources> folder.

Housekeeping

1. HW2 graded, back soon. HW3 on the way.
2. Guide to final paper now in <Resources> folder.
3. No lecture/flipped session next week: will use lecture time (Tues, 11am-1250pm) as office hours.

Recap

Recap

What is a topic, literally?

Recap

What is a topic, literally?

What is a topic, substantively?

Recap

What is a topic, literally?

What is a topic, substantively?

What is a topic model for?

Plate Diagram for LDA

Plate Diagram for LDA

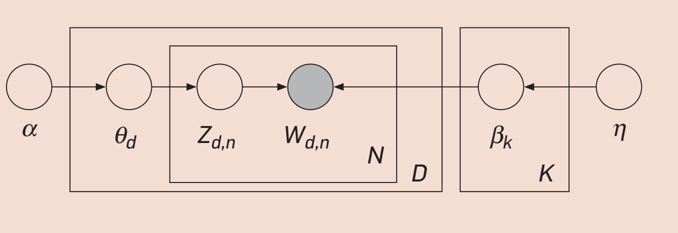
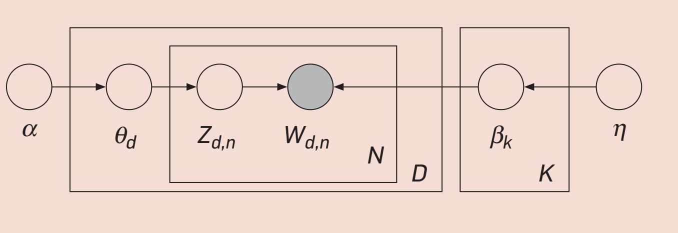
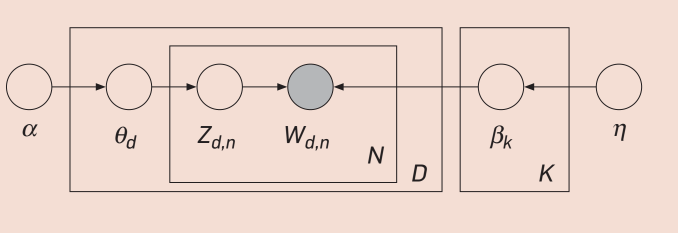


Plate Diagram for LDA



What is the difference between solid and empty nodes?

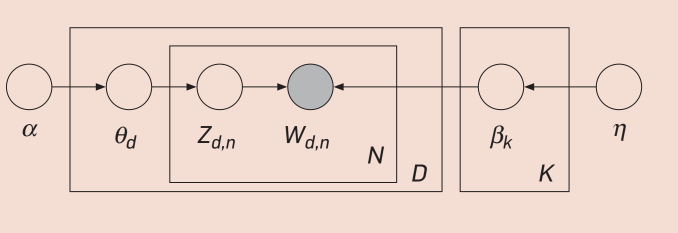
Plate Diagram for LDA



What is the difference between solid and empty nodes?

What does plate mean?

Plate Diagram for LDA

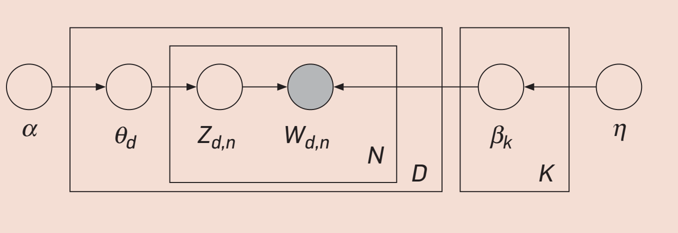


What is the difference between solid and empty nodes?

What does plate mean?

What is z?

Plate Diagram for LDA

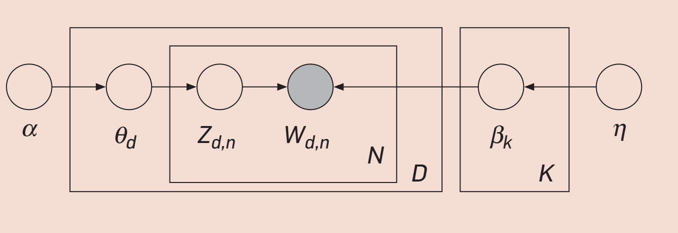


What is the difference between solid and empty nodes?

What does plate mean?

What is z? θ?

Plate Diagram for LDA

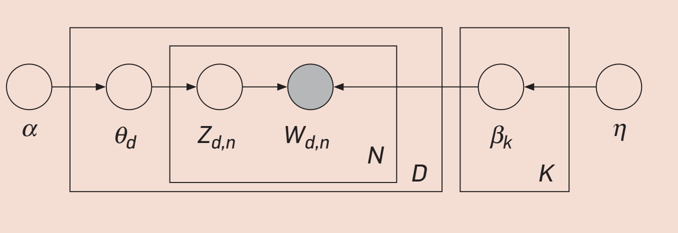


What is the difference between solid and empty nodes?

What does plate mean?

What is z? θ? β?

Plate Diagram for LDA



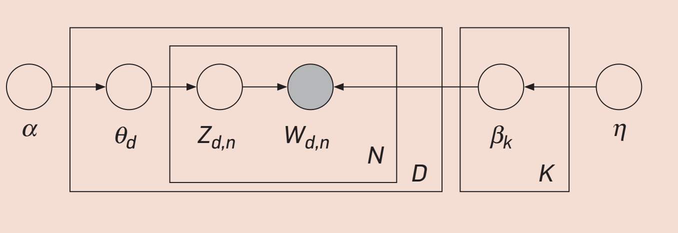
What is the difference between solid and empty nodes?

What does plate mean?

What is z? θ? β?

$\rightarrow$ z is specific topic for this word in this document;

Plate Diagram for LDA



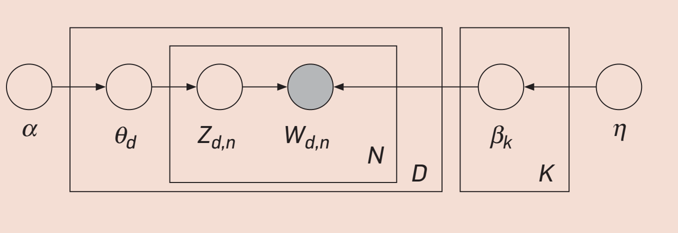
What is the difference between solid and empty nodes?

What does plate mean?

What is z? θ? β?

$\rightarrow$ z is specific topic for this word in this document; θ is distribution over topics for this document;

Plate Diagram for LDA



What is the difference between solid and empty nodes?

What does plate mean?

What is z? θ? β?

$\rightarrow$ z is specific topic for this word in this document; θ is distribution over topics for this document; β is distribution over words (for this topic)

LDA...

LDA...

LDA is a very popular topic model:

LDA...

LDA is a very popular topic model: a probabilistic procedure to generate topics. Thus, a 'generative' model.

LDA. . .

LDA is a very popular topic model: a probabilistic procedure to generate topics. Thus, a 'generative' model.

There are D documents in the corpus.

LDA...

LDA is a very popular topic model: a probabilistic procedure to generate topics. Thus, a 'generative' model.

There are D documents in the corpus. There are V terms in these D documents.

LDA...

LDA is a very popular topic model: a probabilistic procedure to generate topics. Thus, a 'generative' model.

There are D documents in the corpus. There are V terms in these D documents. For now suppose we know the K topic distributions: there are K multinomials containing V elements each.

LDA...

LDA is a very popular topic model: a probabilistic procedure to generate topics. Thus, a 'generative' model.

There are D documents in the corpus. There are V terms in these D documents. For now suppose we know the K topic distributions: there are K multinomials containing V elements each.

The multinomial distribution for the ith topic is denoted β_i, and $|\beta_i| = V$, meaning that the 'size' of this multinomial is equal to the number of different words in the corpus.

So, a little more formally...

So, a little more formally. . .

For each document. . .

So, a little more formally...

For each document...

1. Randomly choose a distribution over topics (multinomial of length K)

So, a little more formally...

For each document...

1. Randomly choose a distribution over topics (multinomial of length K)
2. Then, for every word in the document...

So, a little more formally...

For each document...

1. Randomly choose a distribution over topics (multinomial of length K)
2. Then, for every word in the document...
   1. Probabilistically draw one of the K topics from the distribution over topics from step 1. E.g. draw β_j

So, a little more formally...

For each document...

1. Randomly choose a distribution over topics (multinomial of length K)
2. Then, for every word in the document...
    1. Probabilistically draw one of the K topics from the distribution over topics from step 1. E.g. draw β_j
    2. Probabilistically draw one of the V words from β_j

Aside: Dirichlet distribution

Aside: Dirichlet distribution

The Dirichlet distribution is a conjugate prior for the multinomial ('categorical' if you only have one trial) distribution.

Aside: Dirichlet distribution

The Dirichlet distribution is a conjugate prior for the multinomial ('categorical' if you only have one trial) distribution. Makes certain calculations easier.

Aside: Dirichlet distribution

The Dirichlet distribution is a conjugate prior for the multinomial ('categorical' if you only have one trial) distribution. Makes certain calculations easier.

It is parameterized by a vector of positive real numbers $\boldsymbol{\alpha}$. In principle, one can have $\alpha_1, \ldots, \alpha_k$ be different concentration parameters, but LDA uses special symmetric Dirichlet where all the values of $\boldsymbol{\alpha}$ are the same.

Aside: Dirichlet distribution

The Dirichlet distribution is a conjugate prior for the multinomial ('categorical' if you only have one trial) distribution. Makes certain calculations easier.

It is parameterized by a vector of positive real numbers $\boldsymbol{\alpha}$. In principle, one can have $\alpha_1, \ldots, \alpha_k$ be different concentration parameters, but LDA uses special symmetric Dirichlet where all the values of $\boldsymbol{\alpha}$ are the same.

In terms of θ... larger values of α (assuming we are in symmetric case) mean we think (*a priori*) that documents are generally an even mix of the topics.

Aside: Dirichlet distribution

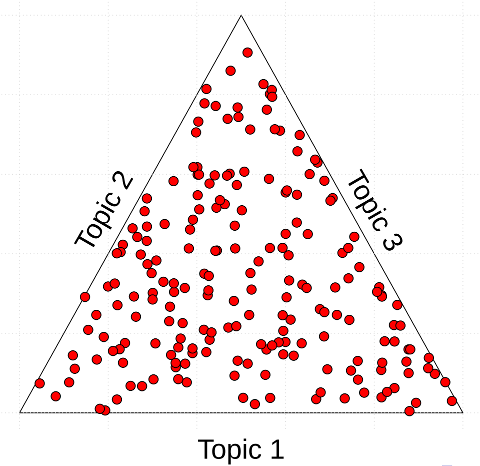
The Dirichlet distribution is a conjugate prior for the multinomial ('categorical' if you only have one trial) distribution. Makes certain calculations easier.

It is parameterized by a vector of positive real numbers $\boldsymbol{\alpha}$. In principle, one can have $\alpha_1, \ldots, \alpha_k$ be different concentration parameters, but LDA uses special symmetric Dirichlet where all the values of $\boldsymbol{\alpha}$ are the same.

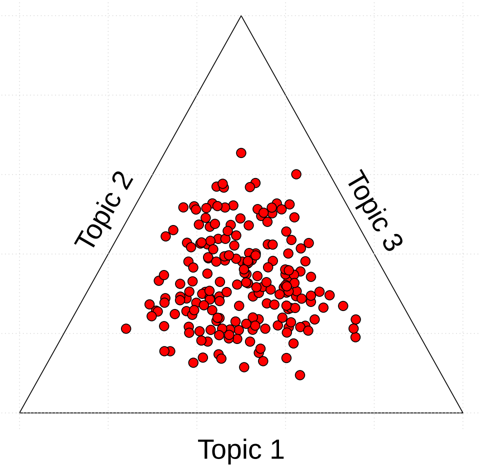
In terms of θ... larger values of α (assuming we are in symmetric case) mean we think (*a priori*) that documents are generally an even mix of the topics. If α is small (less than 1) we think a given document is generally from one or a few topics.

Example of Dirichlet

200 documents, 3 topics, $\alpha = 1$
(uniform)

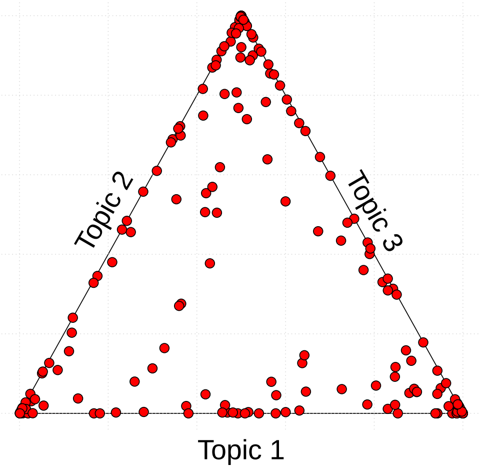


200 documents, 3 topics, $\alpha = 5$



Example of Dirichlet

200 documents, 3 topics, $\alpha = 0.2$



Exercise

Would you pick a high, low or uniform (i.e. $= 1$) α prior for the following:

1. State of the Union addresses
2. Sports news bulletins (e.g. headline summaries of results)
3. Tweets from `@NYUDataScience`
4. News stories from NYT

Answers?

Would you pick a high, low or uniform (i.e. $= 1$) α prior for the following:

1. State of the Union addresses

A larger than one– multiple topics covered

2. Sports news bulletins (e.g. headline summaries of results)

A lower than 1–mostly about one sport or other

3. Tweets from `@NYUDataScience`

A lower than 1–typically each document is about one topic

4. News stories from NYT

A symmetric–some stories talk about one thing, some touch on multiple issues

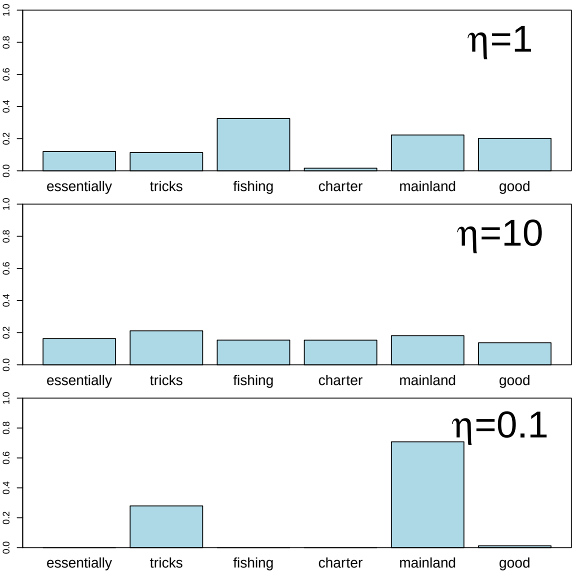
Varying Dirichlet prior for β

Varying Dirichlet prior for β

Vocab: `essentially`, `tricks`, `fishing`, `charter`, `mainland`, `good`

Varying Dirichlet prior for β

Vocab: `essentially`, `tricks`, `fishing`, `charter`, `mainland`, `good`



Results

Results

For a user-selected k, a typical implementation of LDA will return...

Results

For a user-selected k, a typical implementation of LDA will return...

The word distribution for each topic.

Results

For a user-selected k, a typical implementation of LDA will return...

The word distribution for each topic. (estimated βs)

Results

For a user-selected k, a typical implementation of LDA will return...

The word distribution for each topic. (estimated βs)

The topic distribution for each document.

Results

For a user-selected k, a typical implementation of LDA will return...

The word distribution for each topic. (estimated βs)
The topic distribution for each document. (estimated θs)

Results

For a user-selected k, a typical implementation of LDA will return...

The word distribution for each topic. (estimated βs)
The topic distribution for each document. (estimated θs)

Some implementations allow you to estimate e.g. α, in which case this is also returned.

Results

For a user-selected k, a typical implementation of LDA will return...

The word distribution for each topic. (estimated βs)
The topic distribution for each document. (estimated θs)

Some implementations allow you to estimate e.g. α, in which case this is also returned. And perhaps some kind of fit statistic(s).

A Manifesto Example

A Manifesto Example

69 UK manifestos.

A Manifesto Example

69 UK manifestos. Some preprocessing.

A Manifesto Example

69 UK manifestos. Some preprocessing. Used `topicmodels` to fit five topics.

A Manifesto Example

69 UK manifestos. Some preprocessing. Used `topicmodels` to fit five topics. Has Gibbs sampling and variational options.

A Manifesto Example

69 UK manifestos. Some preprocessing. Used `topicmodels` to fit five topics. Has Gibbs sampling and variational options.

The (some selected) word distributions for each topic.

A Manifesto Example

69 UK manifestos. Some preprocessing. Used `topicmodels` to fit five topics. Has Gibbs sampling and variational options.

The (some selected) word distributions for each topic. Sum down the columns is one.

A Manifesto Example

69 UK manifestos. Some preprocessing. Used `topicmodels` to fit five topics. Has Gibbs sampling and variational options.

The (some selected) word distributions for each topic. Sum down the columns is one.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
conservative	0.00188	0.00088	0.00185	0.00221	0.00168
party	0.00145	0.00067	0.00066	0.00577	0.00093
general	0.00073	0.00033	0.00018	0.00192	0.00040
election	0.00079	0.00053	0.00022	0.00235	0.00076
manifesto	0.00059	0.00078	0.00032	0.00099	0.00048
⋮	⋮	⋮	⋮	⋮	⋮

Continued. . .

Continued...

'Top' 6 most frequent words in each topic:

Continued...

'Top' 6 most frequent words in each topic: might help interpretation (!)

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	people	new	[markup]	new	must
2	local	government	people	labour	government
3	government	people	new	government	labour
4	new	continue	work	people	shall
5	tax	can	[markup]	shall	can
6	liberal	conservative	support	britain	policy

Continued...

'Top' 6 most frequent words in each topic: might help interpretation (!)

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	people	new	[markup]	new	must
2	local	government	people	labour	government
3	government	people	new	government	labour
4	new	continue	work	people	shall
5	tax	can	[markup]	shall	can
6	liberal	conservative	support	britain	policy

Up to analyst to label the topics!

Continued...

'Top' 6 most frequent words in each topic: might help interpretation (!)

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	people	new	[markup]	new	must
2	local	government	people	labour	government
3	government	people	new	government	labour
4	new	continue	work	people	shall
5	tax	can	[markup]	shall	can
6	liberal	conservative	support	britain	policy

Up to analyst to label the topics!

Meaningless 'junk' topics not unusual:

Continued...

'Top' 6 most frequent words in each topic: might help interpretation (!)

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	people	new	[markup]	new	must
2	local	government	people	labour	government
3	government	people	new	government	labour
4	new	continue	work	people	shall
5	tax	can	[markup]	shall	can
6	liberal	conservative	support	britain	policy

Up to analyst to label the topics!

Meaningless 'junk' topics not unusual: debate as to whether one has to interpret every topic.

Exercise

This is the output for four topics (of a 100 topic model) produced for a sample of the Associated Press corpus (1988–1990).

Name/describe the topics.

New	Million	Children	School
Film	Program	Women	Students
Show	Tax	People	Schools
Music	Budget	Child	Education
Movie	Billion	Years	Teachers
Play	Federal	Families	High
Musical	Year	Work	Public
Best	Spending	Parent	Teacher
Actor	New	Says	Bennett
First	State	Family	Manigat
York	Plan	Welfare	Namphy
Opera	Money	Men	State
Theater	Programs	Percent	President
Actress	Government	Care	Elementary
Love	Congress	Life	Haiti

Exercise

This is the output for four topics (of a 100 topic model) produced for a sample of the Associated Press corpus (1988–1990).

Name/describe the topics.

"ARTS"	"BUDGET"	"CHILDREN"	"EDUCATION"
New	Million	Children	School
Film	Program	Women	Students
Show	Tax	People	Schools
Music	Budget	Child	Education
Movie	Billion	Years	Teachers
Play	Federal	Families	High
Musical	Year	Work	Public
Best	Spending	Parent	Teacher
Actor	New	Says	Bennett
First	State	Family	Manigat
York	Plan	Welfare	Namphy
Opera	Money	Men	State
Theater	Programs	Percent	President
Actress	Government	Care	Elementary
Love	Congress	Life	Haiti

Picking k

Picking k

In social science, keep fitting until topic of interest peels off. . .

In computer science, keep fitting until perplexity suggests otherwise.

Picking k

In social science, keep fitting until topic of interest peels off...

In computer science, keep fitting until perplexity suggests otherwise.

Q What does perplexity measure? (intuitively)

Ballpark k from Mimno & Blei, 2011

Ballpark k from Mimno & Blei, 2011

Table 1: Statistics for models used as examples.

Name	Docs	Tokens	Vocab	Topics
News	1.8M	76M	121k	1000
Blogs	13k	2.2M	90k	100
Parliament	540k	55M	52k	300

Example: an "education" topic

Example: an "education" topic

for $k = 5$

local
people
system
tax
education
ensure
support
new
liberal
rights

for $k = 100$

work
new
children
people
system
standards
education
every
schools
labour

Example: an "education" topic

for $k = 5$

local
people
system
tax
education
ensure
support
new
liberal
rights

"bad"

for $k = 100$

work
new
children
people
system
standards
education
every
schools
labour

Example: an “education” topic

for $k = 5$

local
people
system
tax
education
ensure
support
new
liberal
rights

“bad”

for $k = 100$

work
new
children
people
system
standards
education
every
schools
labour

“good”

Fitting LDA

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables.

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables. The joint density is

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables. The joint density is

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$$

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables. The joint density is

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$$

Q What is $\mathbf{x}$ and $\mathbf{z}$ in a topic model?

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables. The joint density is

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$$

Q What is $\mathbf{x}$ and $\mathbf{z}$ in a topic model?

Q I want $p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z},\mathbf{x})}{p(\mathbf{x})}$.

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables. The joint density is

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$$

Q What is $\mathbf{x}$ and $\mathbf{z}$ in a topic model?

Q I want $p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z},\mathbf{x})}{p(\mathbf{x})}$. What name do we give to that quantity in Bayesian statistics?

Fitting LDA

Let $\mathbf{x}$ be our observations, and $\mathbf{z}$ be our latent variables. The joint density is

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$$

Q What is $\mathbf{x}$ and $\mathbf{z}$ in a topic model?

Q I want $p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z},\mathbf{x})}{p(\mathbf{x})}$. What name do we give to that quantity in Bayesian statistics?

Q What is $p(\mathbf{z}|\mathbf{x})$, in terms of our topic model (intuitively)?

Variational Inference: the problem

Variational Inference: the problem

So we want

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z}, \mathbf{x})}{p(\mathbf{x})}$$

the denominator requires calculating $p(\mathbf{x}) = \int p(\mathbf{z}, \mathbf{x}) dz$.

Variational Inference: the problem

So we want

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z}, \mathbf{x})}{p(\mathbf{x})}$$

the denominator requires calculating $p(\mathbf{x}) = \int p(\mathbf{z}, \mathbf{x}) dz$.

Q What do we call this quantity?

Variational Inference: the problem

So we want

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z}, \mathbf{x})}{p(\mathbf{x})}$$

the denominator requires calculating $p(\mathbf{x}) = \int p(\mathbf{z}, \mathbf{x}) dz$.

Q What do we call this quantity? Why is it hard to obtain?

Variational Inference: the problem

So we want

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z}, \mathbf{x})}{p(\mathbf{x})}$$

the denominator requires calculating $p(\mathbf{x}) = \int p(\mathbf{z}, \mathbf{x}) dz$.

Q What do we call this quantity? Why is it hard to obtain?

We want the evidence, so that our calculation of the value of the latent variable posteriors is correct.

Variational Inference: the problem

So we want

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z}, \mathbf{x})}{p(\mathbf{x})}$$

the denominator requires calculating $p(\mathbf{x}) = \int p(\mathbf{z}, \mathbf{x}) dz$.

Q What do we call this quantity? Why is it hard to obtain?

We want the evidence, so that our calculation of the value of the latent variable posteriors is correct.

Q How and why does this differ to e.g. naive Bayes classification?

Variational Inference: the solution

Variational Inference: the solution

1 Make sure the posterior and prior are all conjugate:

Variational Inference: the solution

1. Make sure the posterior and prior are all conjugate: nice, closed form integral.

Variational Inference: the solution

1. Make sure the posterior and prior are all conjugate: nice, closed form integral. Do the integral.

Variational Inference: the solution

1 Make sure the posterior and prior are all conjugate: nice, closed form integral. Do the integral. Often not an option.

Variational Inference: the solution

1. Make sure the posterior and prior are all conjugate: nice, closed form integral. Do the integral. Often not an option.
2. Draw samples from the posterior via MCMC, and means we don't have to calculate evidence directly.

Variational Inference: the solution

1. Make sure the posterior and prior are all conjugate: nice, closed form integral. Do the integral. Often not an option.
2. Draw samples from the posterior via MCMC, and means we don't have to calculate evidence directly. Potentially very difficult and very slow. But it will get there eventually.

Variational Inference: the solution

1 Make sure the posterior and prior are all conjugate: nice, closed form integral. Do the integral. Often not an option.

2 Draw samples from the posterior via MCMC, and means we don't have to calculate evidence directly. Potentially very difficult and very slow. But it will get there eventually.

3 Find an approximating distribution $q(\mathbf{z}) \approx p(\mathbf{z}|\mathbf{x})$.

Variational Inference: the solution

1. Make sure the posterior and prior are all conjugate: nice, closed form integral. Do the integral. Often not an option.
2. Draw samples from the posterior via MCMC, and means we don't have to calculate evidence directly. Potentially very difficult and very slow. But it will get there eventually.
3. Find an approximating distribution $q(\mathbf{z}) \approx p(\mathbf{z}|\mathbf{x})$. Iteratively improve the closeness between $q(\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$. This is variational inference.

Variational Inference: more details

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation.

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation. And we can do that in such a way that it doesn't involve having to calculate $p(\mathbf{x})$. See Blei et al, "Variational Inference: A Review for Statisticians"

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation. And we can do that in such a way that it doesn't involve having to calculate $p(\mathbf{x})$. See Blei et al, "Variational Inference: A Review for Statisticians"

Q When to use VI vs MCMC? Which one is more accurate? Which one is faster? What does that imply?

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation. And we can do that in such a way that it doesn't involve having to calculate $p(\mathbf{x})$. See Blei et al, "Variational Inference: A Review for Statisticians"

Q When to use VI vs MCMC? Which one is more accurate? Which one is faster? What does that imply?

A MCMC is guaranteed to get to the conditional you want.

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation. And we can do that in such a way that it doesn't involve having to calculate $p(\mathbf{x})$. See Blei et al, "Variational Inference: A Review for Statisticians"

Q When to use VI vs MCMC? Which one is more accurate? Which one is faster? What does that imply?

A MCMC is guaranteed to get to the conditional you want. But much more slowly.

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation. And we can do that in such a way that it doesn't involve having to calculate $p(\mathbf{x})$. See Blei et al, "Variational Inference: A Review for Statisticians"

Q When to use VI vs MCMC? Which one is more accurate? Which one is faster? What does that imply?

A MCMC is guaranteed to get to the conditional you want. But much more slowly. So, use VI for larger data sets where we don't care as much about precision in the posteriors:

Variational Inference: more details

The key idea is to make the problem one of optimization of the approximation. And we can do that in such a way that it doesn't involve having to calculate $p(\mathbf{x})$. See Blei et al, "Variational Inference: A Review for Statisticians"

Q When to use VI vs MCMC? Which one is more accurate? Which one is faster? What does that imply?

A MCMC is guaranteed to get to the conditional you want. But much more slowly. So, use VI for larger data sets where we don't care as much about precision in the posteriors: e.g. topic modeling.

Evaluating Topic Models

Evaluating Topic Models

Q Given everything we've discussed thus far, what do you think makes for a 'good' topic model?

Evaluating Topic Models

Q Given everything we've discussed thus far, what do you think makes for a 'good' topic model?

A From perspective of extrinsic evaluation,

Evaluating Topic Models

Q Given everything we've discussed thus far, what do you think makes for a 'good' topic model?

A From perspective of extrinsic evaluation, might use held-out likelihood, or ability to predict some target of interest.

Evaluating Topic Models

Q Given everything we've discussed thus far, what do you think makes for a 'good' topic model?

A From perspective of extrinsic evaluation, might use held-out likelihood, or ability to predict some target of interest.

But unclear if 'good' models in that technical sense are 'good' from human perspective.

Evaluating Topic Models

Q Given everything we've discussed thus far, what do you think makes for a 'good' topic model?

A From perspective of extrinsic evaluation, might use held-out likelihood, or ability to predict some target of interest.

But unclear if 'good' models in that technical sense are 'good' from human perspective. Enter: Chang et al "Reading Tea Leaves: How Humans Interpret Topic Models".

Evaluating Topic Models

Q Given everything we've discussed thus far, what do you think makes for a 'good' topic model?

A From perspective of extrinsic evaluation, might use held-out likelihood, or ability to predict some target of interest.

But unclear if 'good' models in that technical sense are 'good' from human perspective. Enter: Chang et al "Reading Tea Leaves: How Humans Interpret Topic Models".

Basic idea: ask humans to evaluate different topic models in terms of coherence and relevance.

Details

Details

1 Word Intrusion.

Details

1. Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one ‘intruder’. If they all (independently) pick same intruder, the topic is coherent.

Details

1 Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one 'intruder'. If they all (independently) pick same intruder, the topic is coherent.

e.g. `children, schools, haddock, standards, education`.

Details

1 Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one ‘intruder’. If they all (independently) pick same intruder, the topic is coherent.

e.g. `children, schools, haddock, standards, education`.

2 Topic Intrusion.

Details

1 Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one 'intruder'. If they all (independently) pick same intruder, the topic is coherent.

e.g. `children, schools, haddock, standards, education`.

2 Topic Intrusion. Give crowdworkers three real (high probability) topics for a document (title and snippet), and one intruder topic (in terms of top words). If they all pick same intruder, topics are relevant.

Details

1. Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one 'intruder'. If they all (independently) pick same intruder, the topic is coherent.

e.g. `children, schools, haddock, standards, education`.

2. Topic Intrusion. Give crowdworkers three real (high probability) topics for a document (title and snippet), and one intruder topic (in terms of top words). If they all pick same intruder, topics are relevant.

Details

1. Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one 'intruder'. If they all (independently) pick same intruder, the topic is coherent.

e.g. `children, schools, haddock, standards, education.`

2. Topic Intrusion. Give crowdworkers three real (high probability) topics for a document (title and snippet), and one intruder topic (in terms of top words). If they all pick same intruder, topics are relevant.

Fit various topic models (type, k) to NYT and Wikipedia. Compare 'best fitting' technical model with what humans think of topic coherence and relevance.

Details

1. Word Intrusion. Give crowdworkers five real (high probability) words from topic, and one 'intruder'. If they all (independently) pick same intruder, the topic is coherent.

e.g. `children, schools, haddock, standards, education`.

2. Topic Intrusion. Give crowdworkers three real (high probability) topics for a document (title and snippet), and one intruder topic (in terms of top words). If they all pick same intruder, topics are relevant.

Fit various topic models (type, k) to NYT and Wikipedia. Compare 'best fitting' technical model with what humans think of topic coherence and relevance.

→ "Traditional metrics are, indeed, *negatively* correlated with the measures of topic quality developed in this paper."

Topic Intrusion Example

6 / 10

DOUGLAS_HOFSTADTER

Douglas Richard Hofstadter (born February 15, 1945 in New York, New York) is an American academic whose research focuses on consciousness, thinking and creativity. He is best known for ", first published in

Show entire excerpt

student	school	study	education	research	university	science	learn
human	life	scientific	science	scientist	experiment	work	idea
play	role	good	actor	star	career	show	performance
write	work	book	publish	life	friend	influence	father