Statistical Topic Models for Multi-Label Document Classification

Tim Rubin

In collaboration with America Holloway, Padhraic Smyth, Mark Steyvers

Automatic text categorization for multi-label documents

- Broad goal is classification of *multi-label* documents
- Multi-label data:
  - Each document can be assigned one or more labels
  - E.g., HEALTH *and* GENETICS *and* CANCER RESEARCH

Automatic text categorization for multi-label documents

- Broad goal is classification of *multi-label* documents
- Multi-label data:
  - Each document can be assigned one or more labels
  - E.g., HEALTH *and* GENETICS *and* CANCER RESEARCH
- Popular approach is binary SVMs

Two Challenges

Two Challenges

- Classification of rare labels

"The skewed distribution of the Yahoo! Directory and other large taxonomies with many extremely rare categories makes the classification performance of SVMs unacceptable. More substantial investigation is thus needed to improve SVMs and other statistical methods for very large-scale applications."

-Liu et al. (2005), *Support Vector Machines with a very large taxonomy*

Two Challenges

- Accounting for label dependencies

"The consensus view in the literature is that it is crucial to take into account label correlations during the classification process .... However as the size of the multi-label data sets grows, most methods struggle with the exponential growth in the number of possible correlations. Consequently these methods are able to be more accurate on small data sets, but are not as applicable to larger data sets."

-Read et al. (2009), *Classifier chains for multi-label classification*

Two Challenges

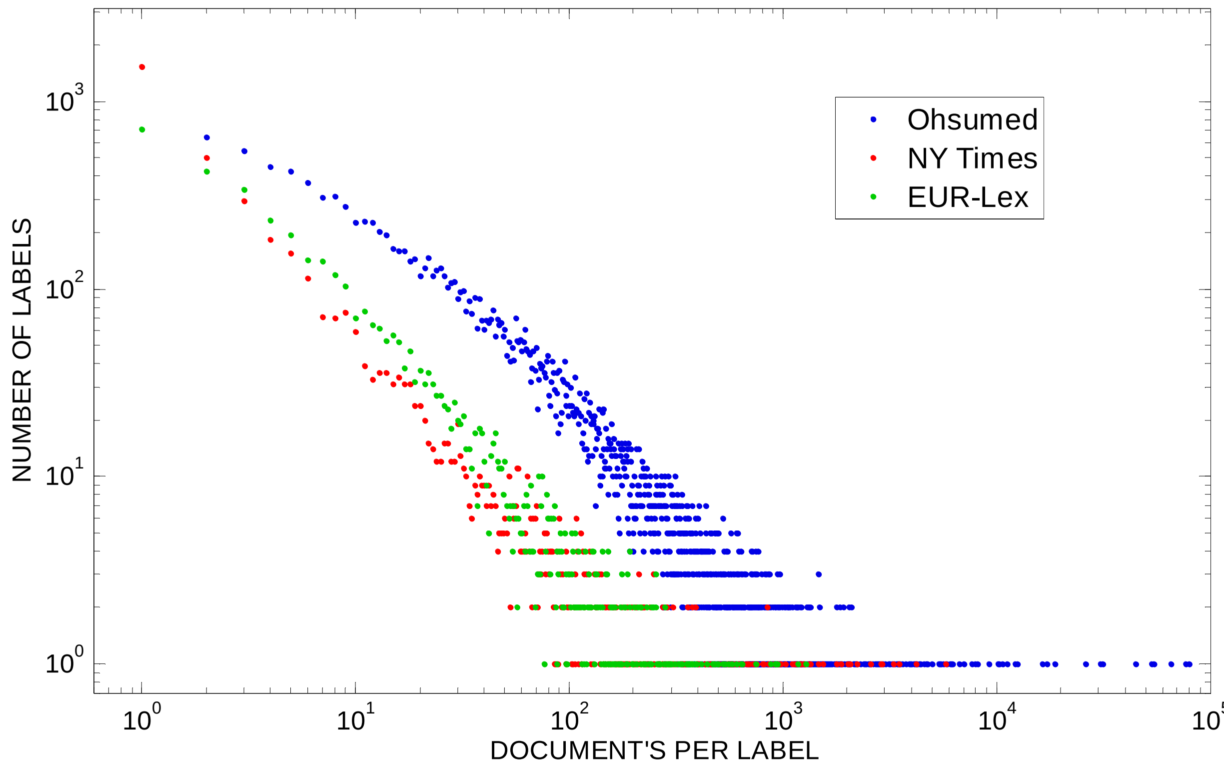
1. Classification of rare labels
2. Accounting for label dependencies

Have significant implications for real-world data

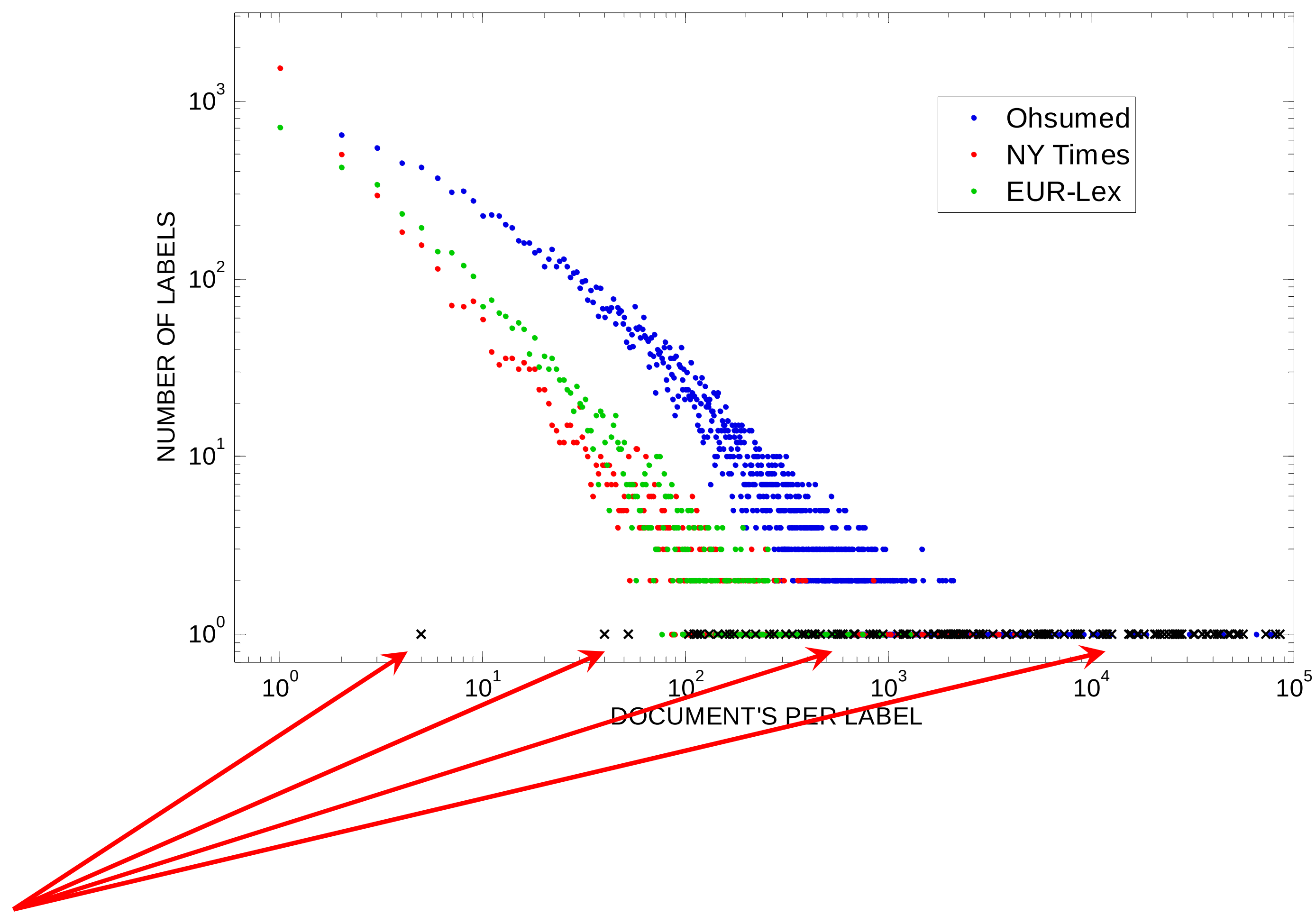
Mismatch between Real World data and Research Datasets

1. Many real world datasets have skewed label-frequency distributions, with many rare labels

Power Law Datasets

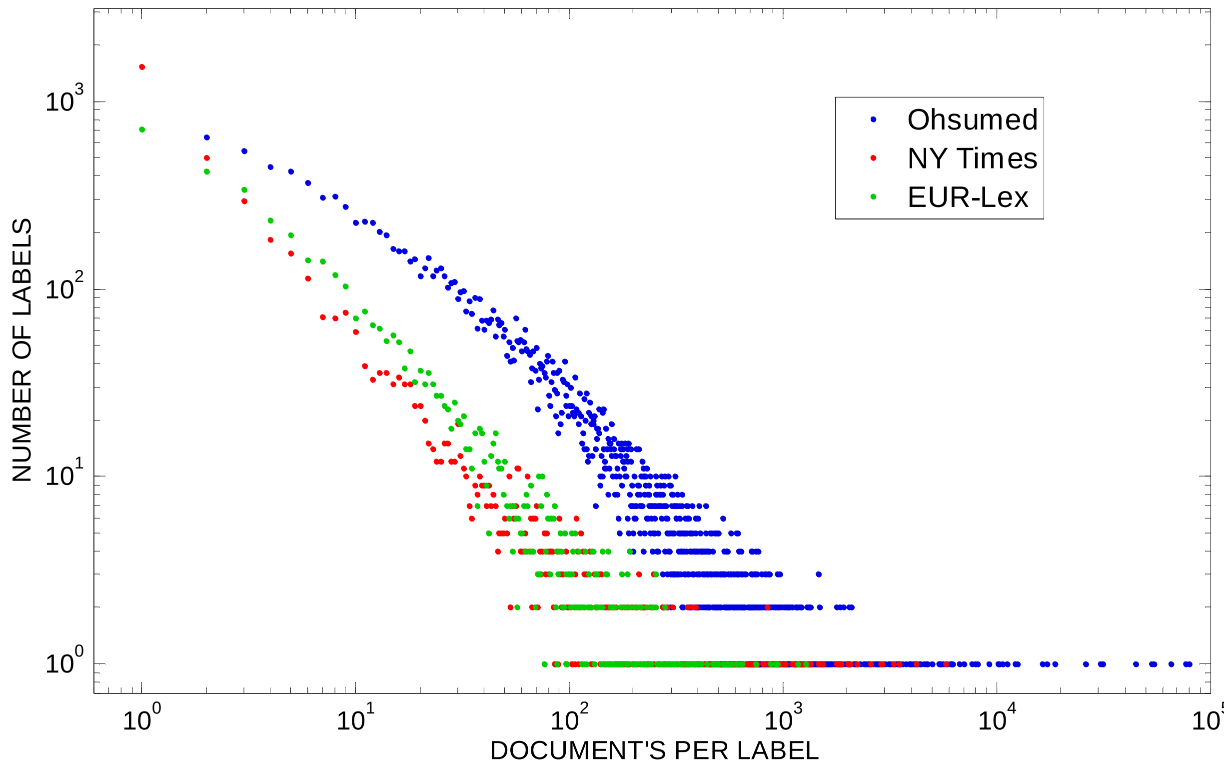


Power Law Datasets

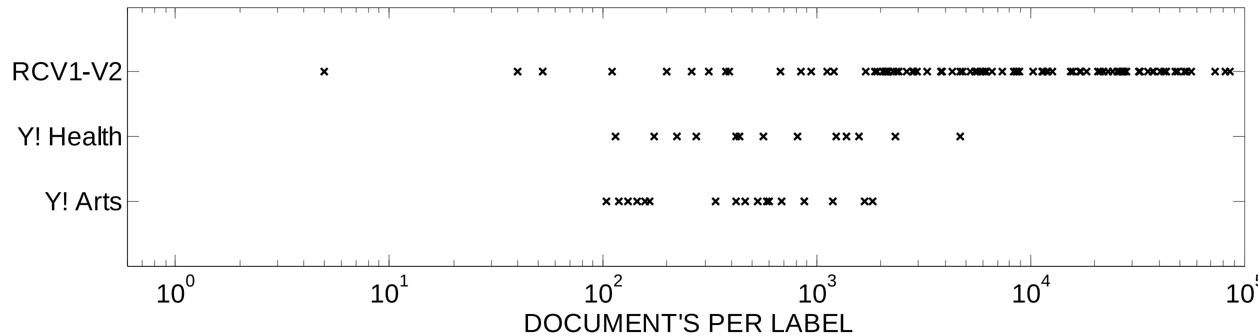


Benchmark Datasets

Power Law Datasets



Benchmark Datasets

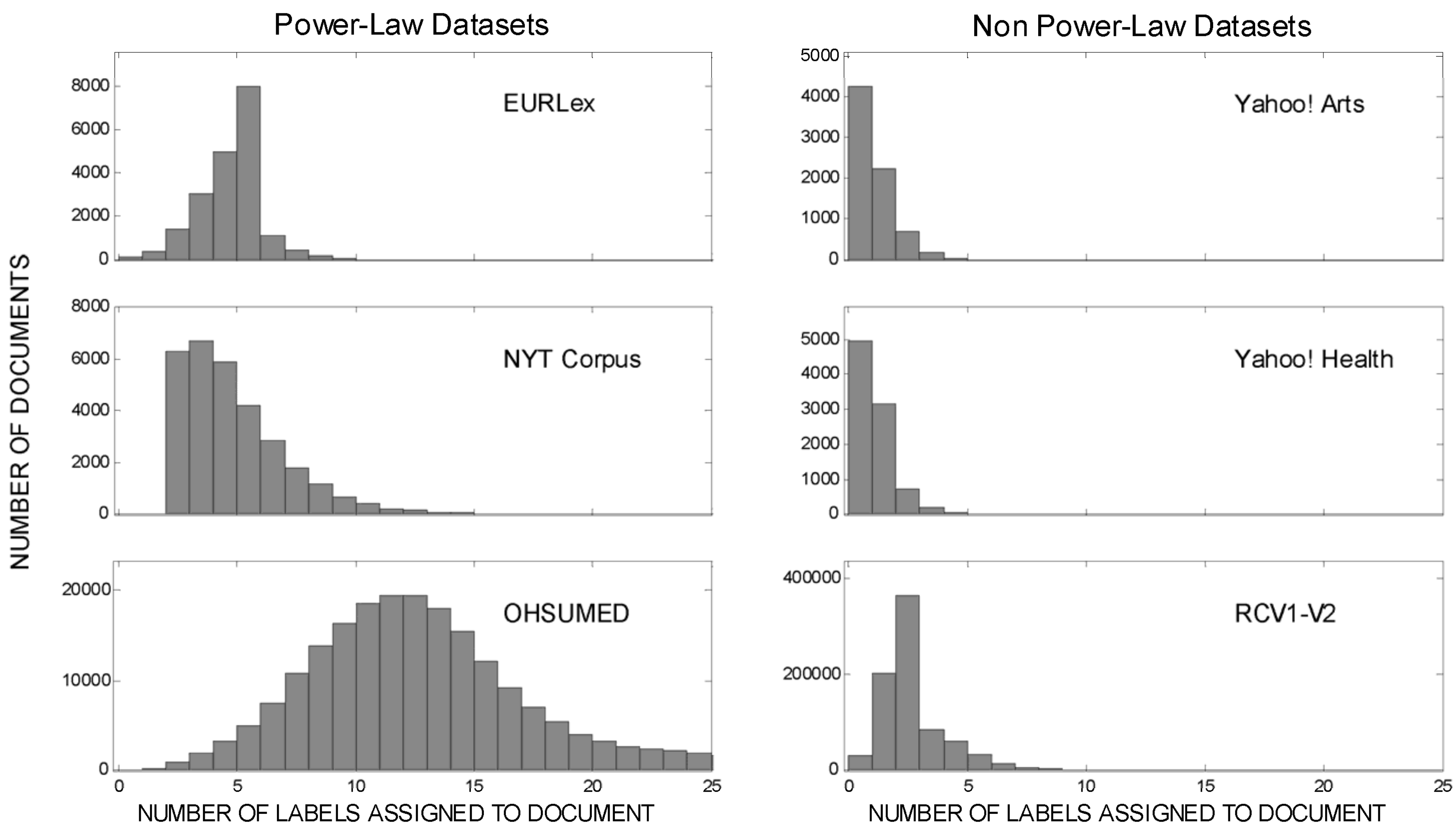


Mismatch between Real World data and Research Datasets

1. Real world datasets have skewed label-frequency distributions, with many rare labels

2. Real world data has many more labels per document than most benchmark data sets

Labels-Per-Document



Looking ahead…

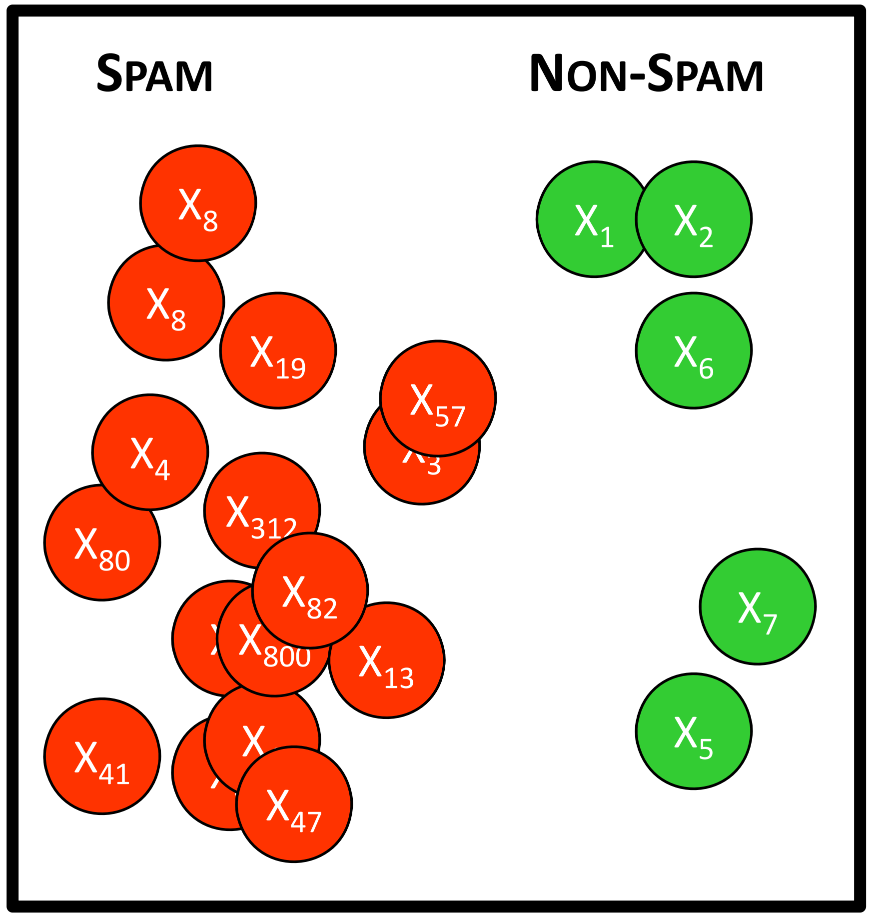
- Present a new probabilistic model that accounts for label dependencies
- Key results:
  - Large improvement over simpler probabilistic models
  - Good performance on *rare* labels
  - Model is highly competitive with discriminative methods on *real-world* data sets

Overview

- Research problem
  - Research goals and challenges
- **Motivate probabilistic methods**
  - Discussion of SVMs for multi-label data
- Present our probabilistic models
- Experimental Results

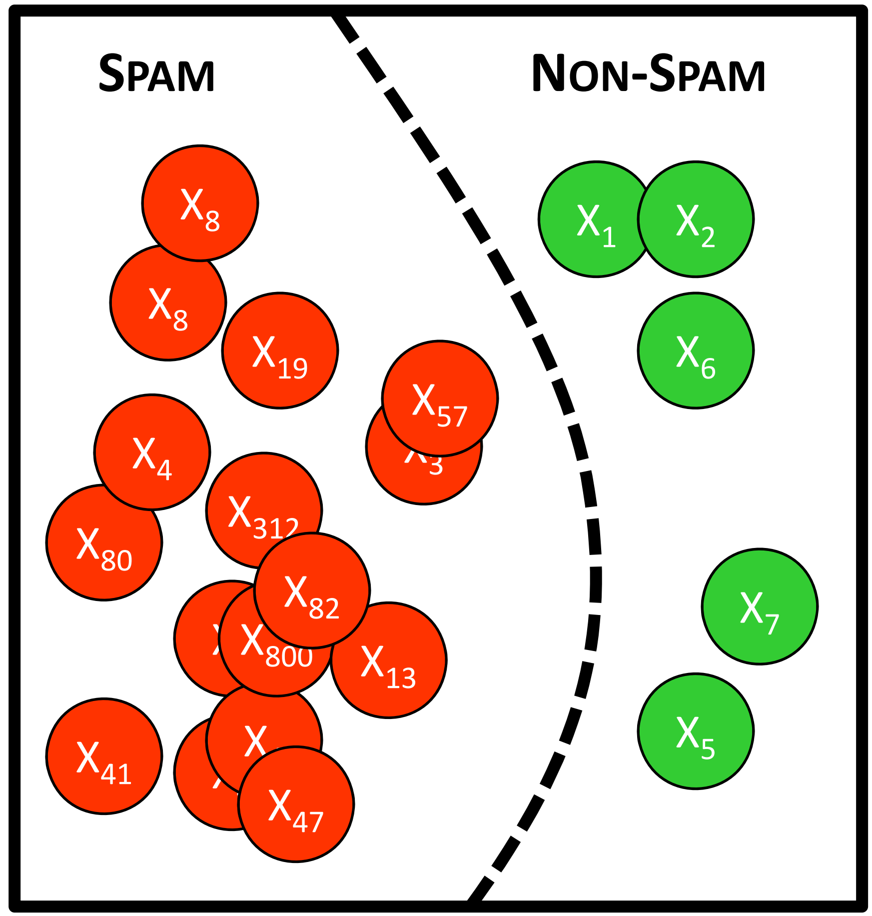
SVMs: Binary data

- SVMs (and *many* other classifiers) are designed for binary data



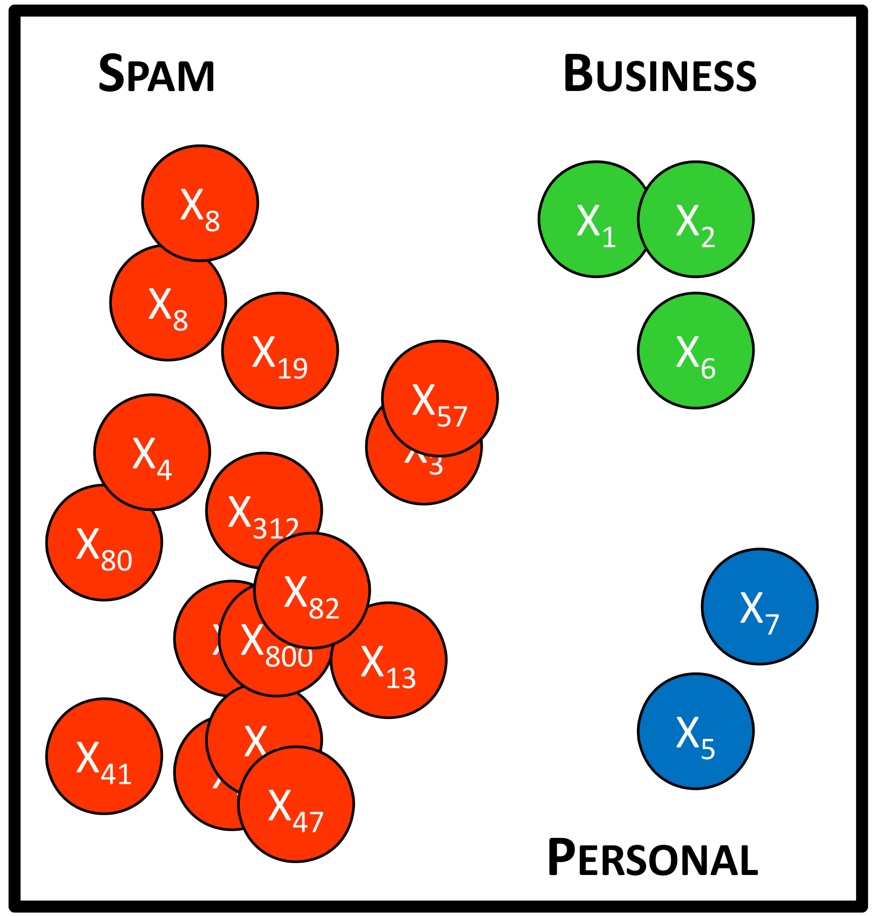
SVMs: Binary data

- SVMs (and *many* other classifiers) are designed for binary data
- **Goal:** Find a separating hyperplane



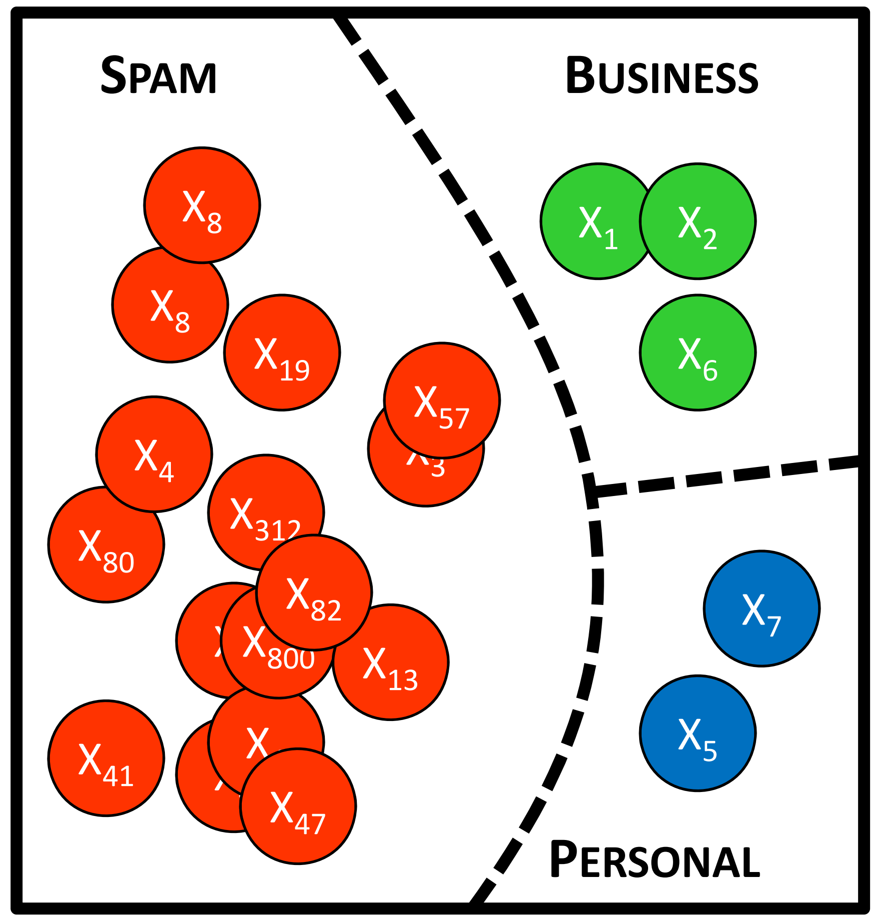
SVMs: Multiple Classes

- Often, we have more than two classes that we wish to predict
- For now, assume ***single label*** per document



SVMs: Multiple Classes

- Often, we have more than two classes that we wish to predict
- For now, assume ***single label*** per document



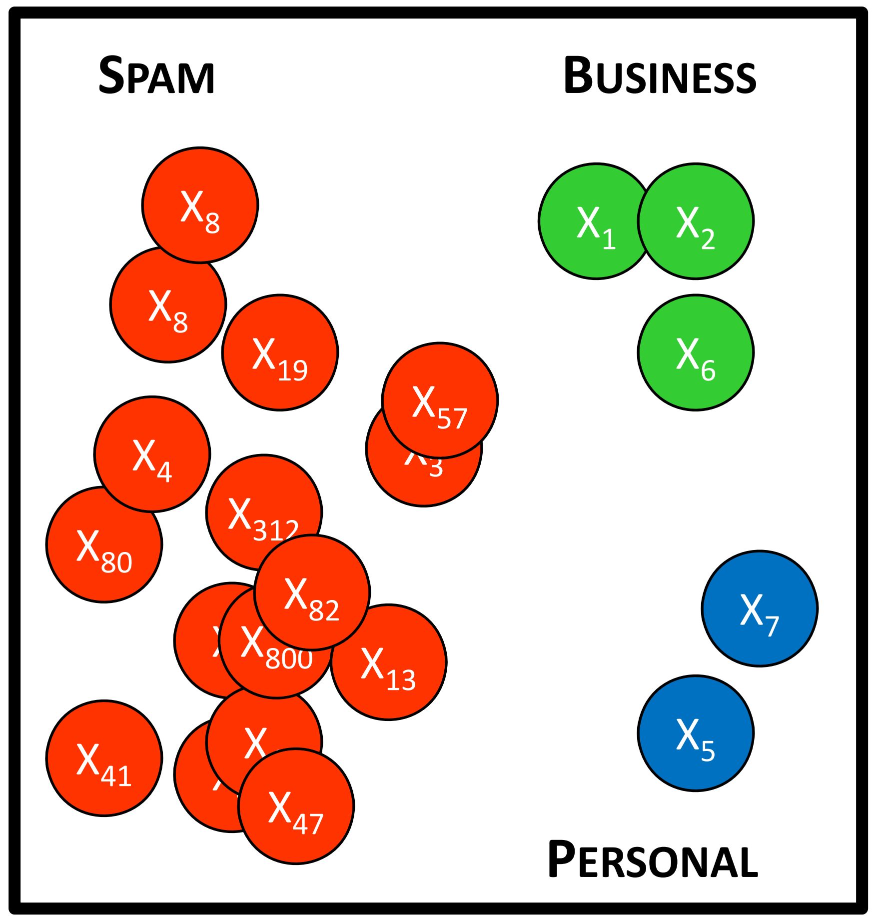
Binary Problem Transformation

- Popular approach:
  - *Transform* the problem into multiple binary classification problems
- "*One-Vs-All*" transformation
  - Note: applicable to any binary classifier

Binary Problem Transformation

"One-Vs-All" Method

- Independently train binary classifier for each of the C classes

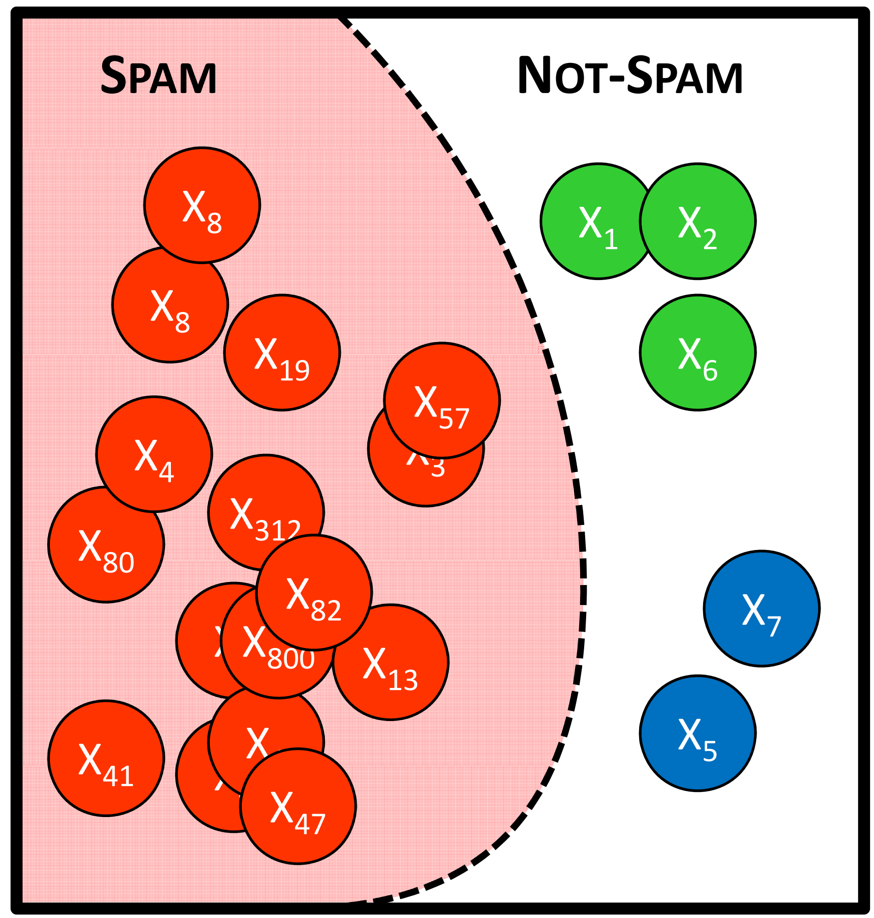


SVMs: Single-Label

"One-Vs-All" Method

- Independently train binary classifier for each of the C classes

Binary Classifier 1: **SPAM**

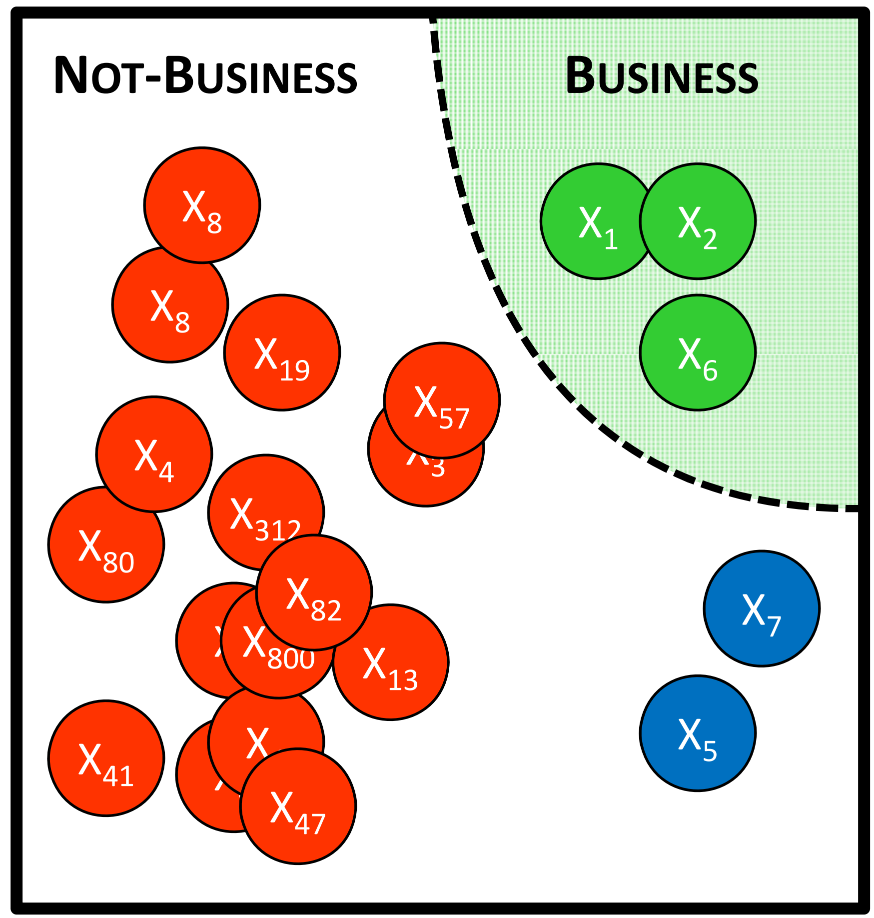


SVMs: Single-Label

"One-Vs-All" Method

- Independently train binary classifier for each of the C classes

Binary Classifier 2: **BUSINESS**

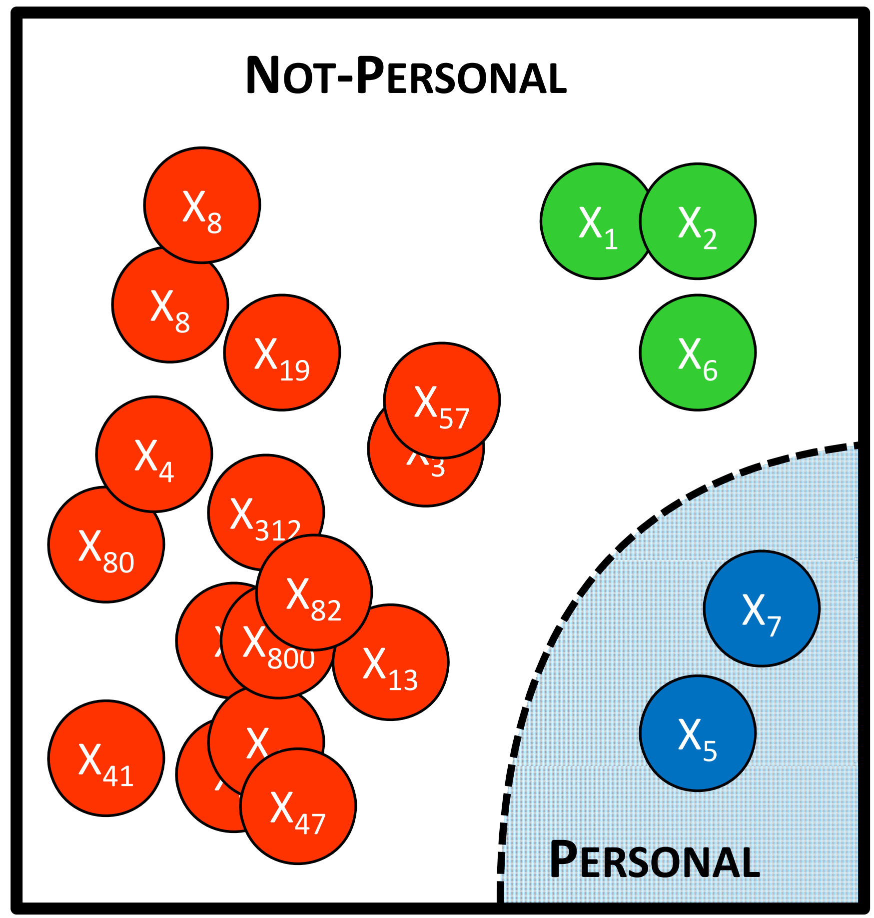


SVMs: Single-Label

"One-Vs-All" Method

- Independently train binary classifier for each of the C classes

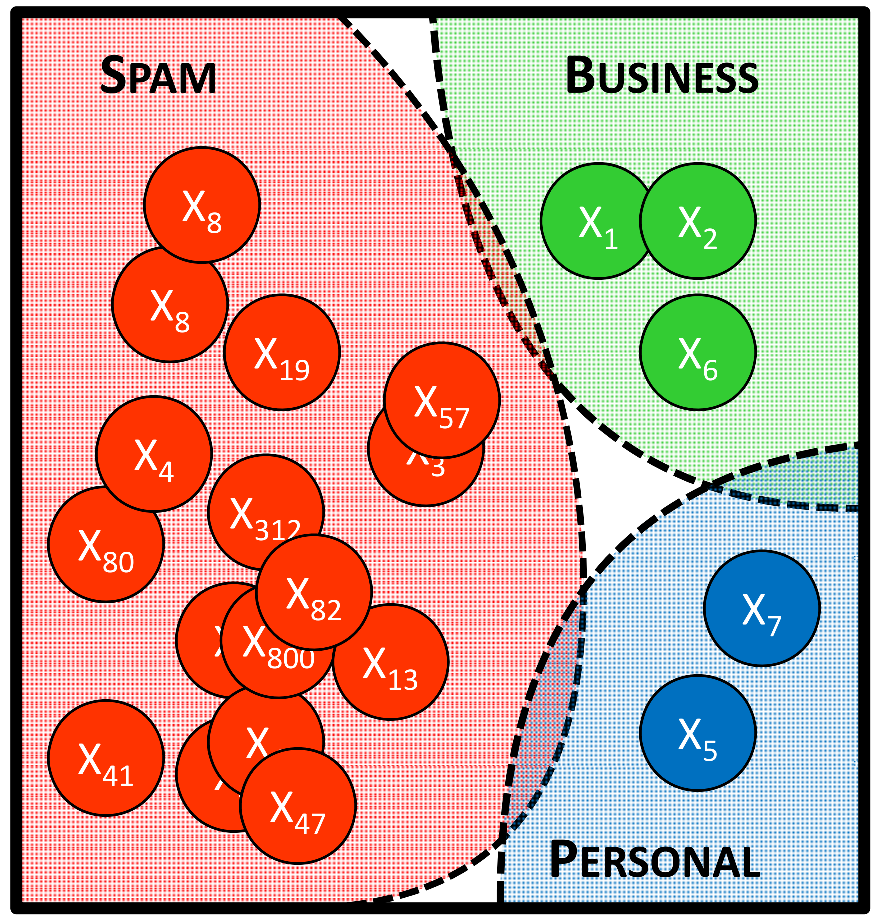
Binary Classifier 3: **PERSONAL**



SVMs: Single-Label

"One-Vs-All" Method

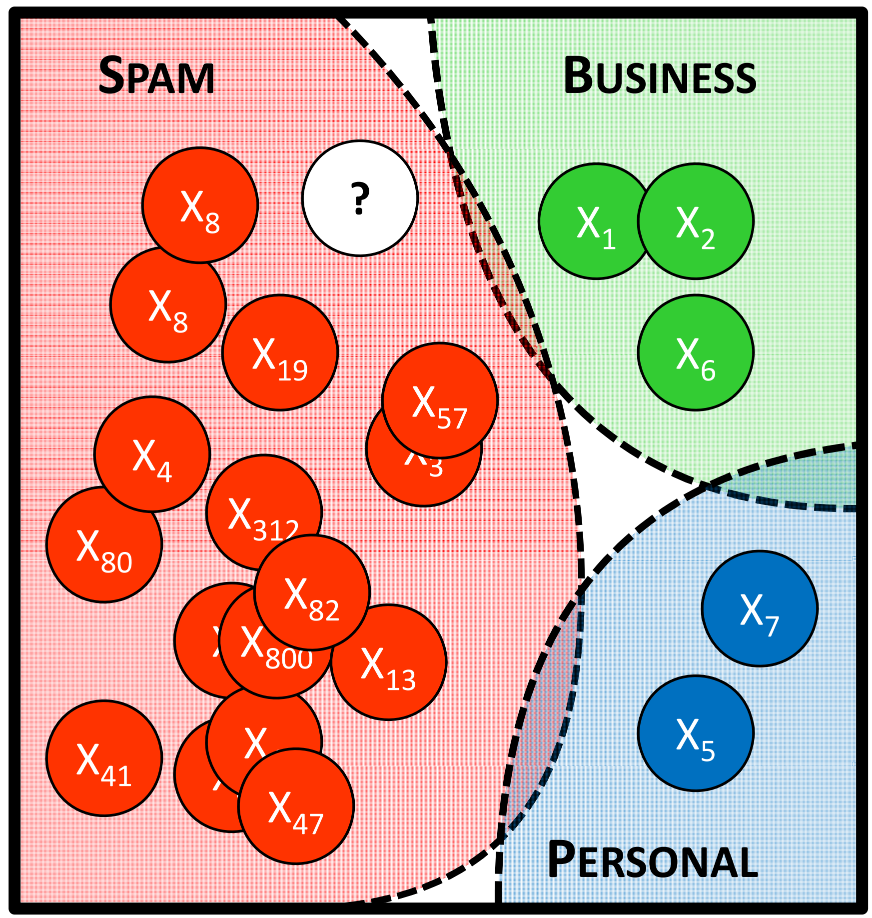
- Independently train binary classifier for each of the C classes



SVMs: Single-Label

"One-Vs-All" Method

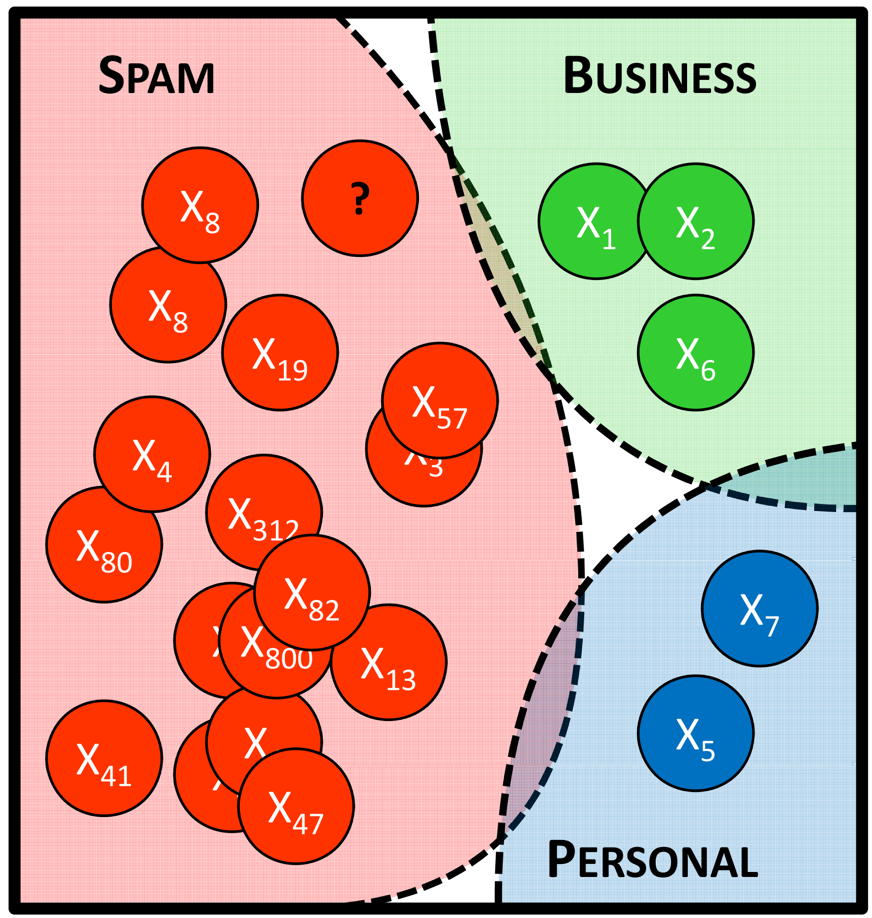
- Independently train binary classifier for each of the C classes



SVMs: Single-Label

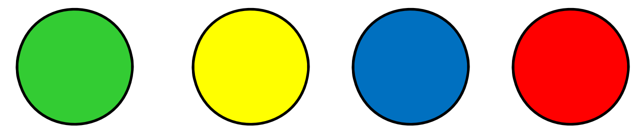
"One-Vs-All" Method

- Independently train binary classifier for each of the C classes

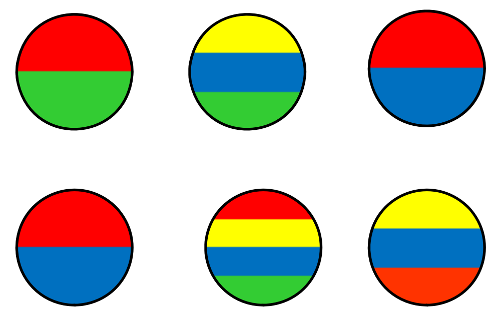


Multi-Label Data

- Previous examples were single-label



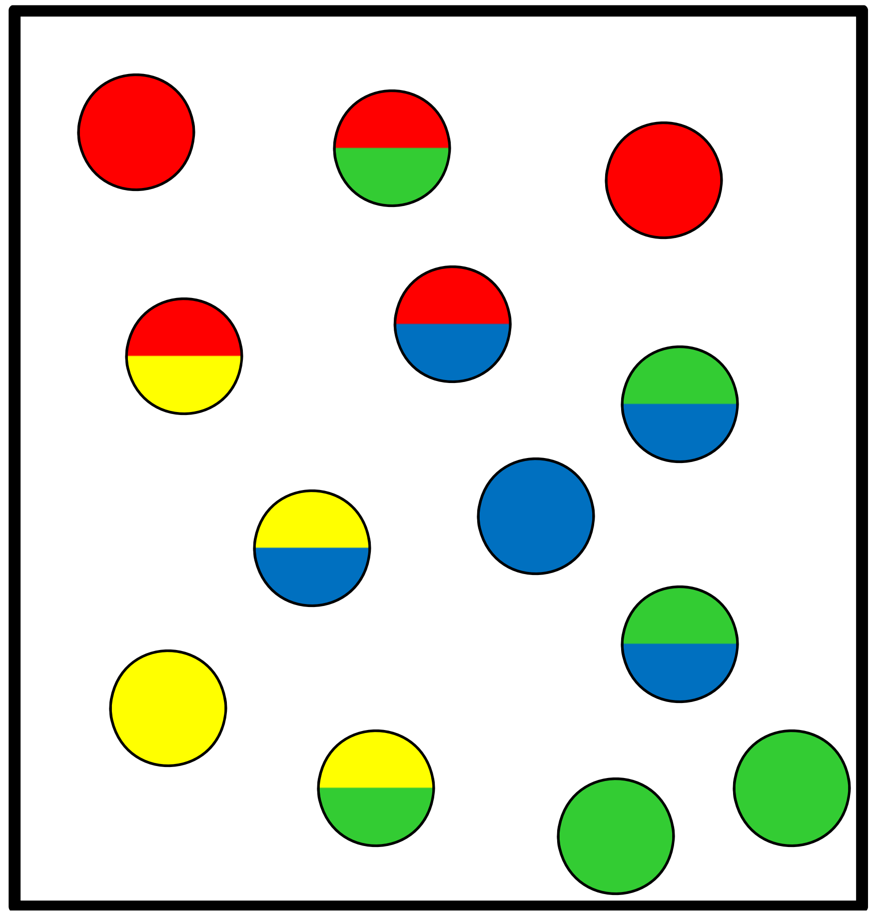
- However, real-world data is often *Multi-Label*
  - Each document can be assigned multiple labels
  - Labels not mutually exclusive



- E.g., A news story might be about **HEALTH** *and* **GENETICS** *and* **CANCER RESEARCH**

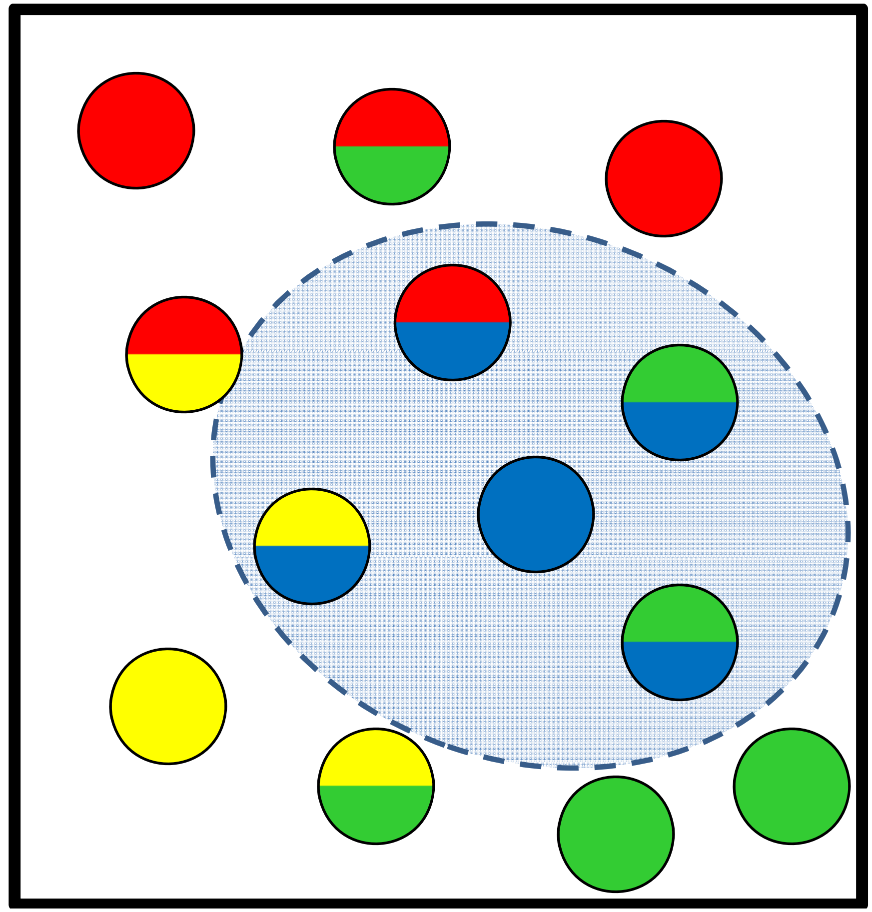
SVMs: Multi-label

- Again, common solution is *to use One-vs-All* transformation



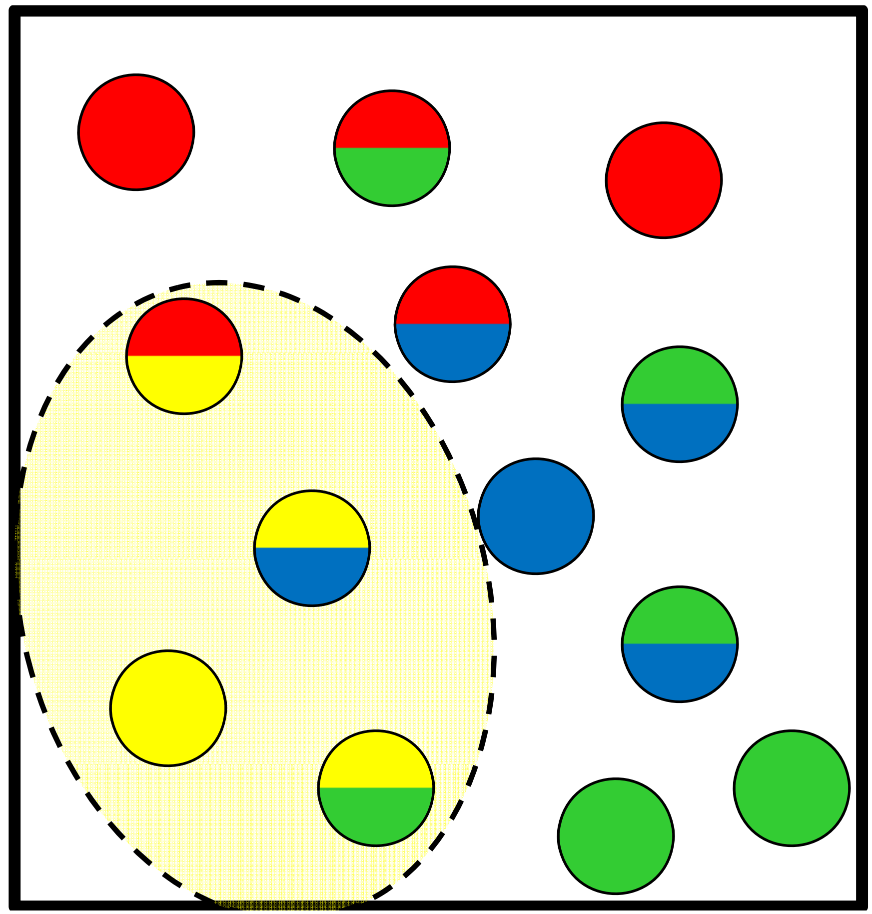
SVMs: Multi-label

- Again, common solution is *to use One-vs-All* transformation



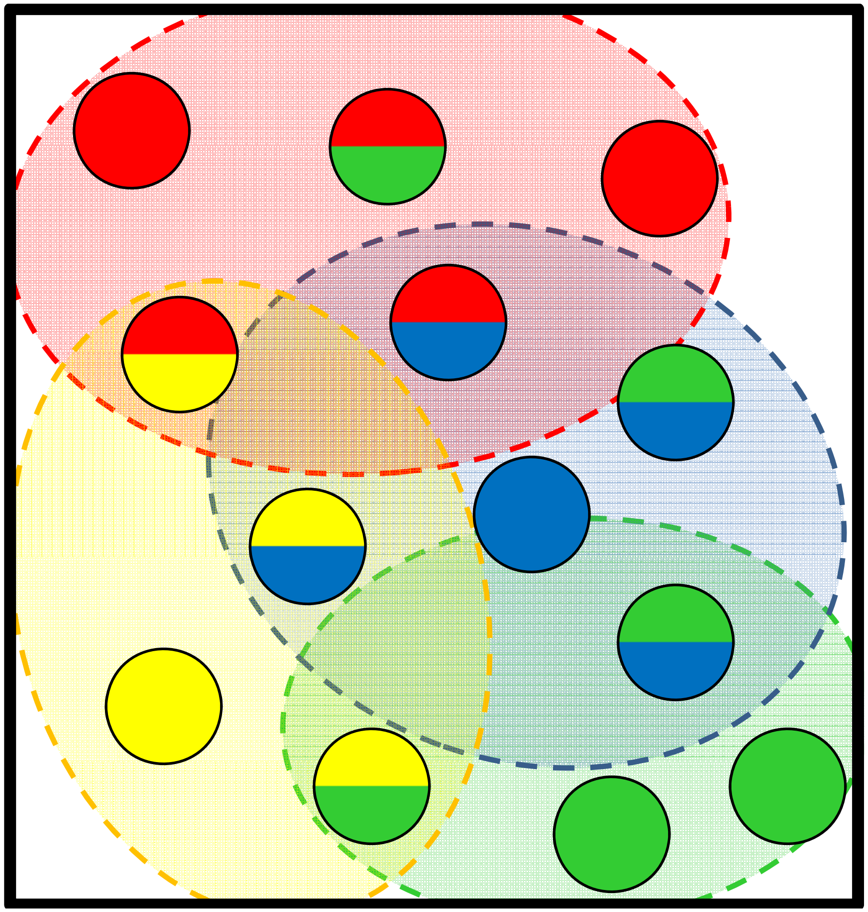
SVMs: Multi-label

- Again, common solution is *to use One-vs-All* transformation



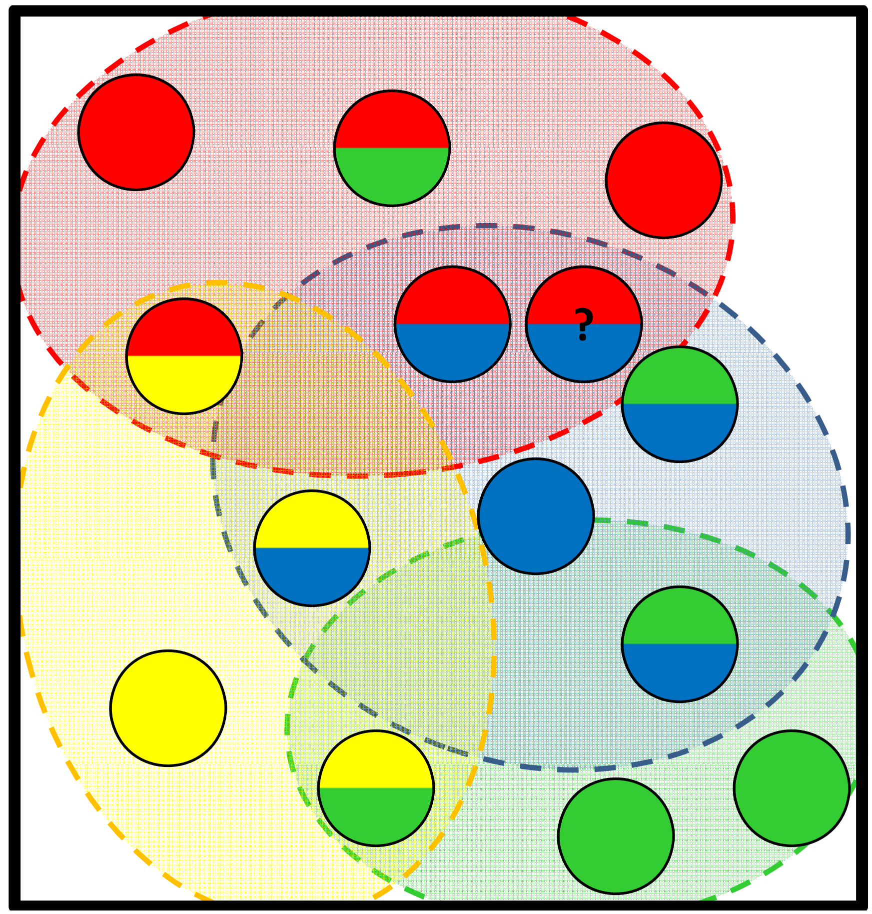
SVMs: Multi-label

- Again, common solution is *to use One-vs-All* transformation
- At test, must make a binary decision for each of the C classes



SVMs: Multi-label

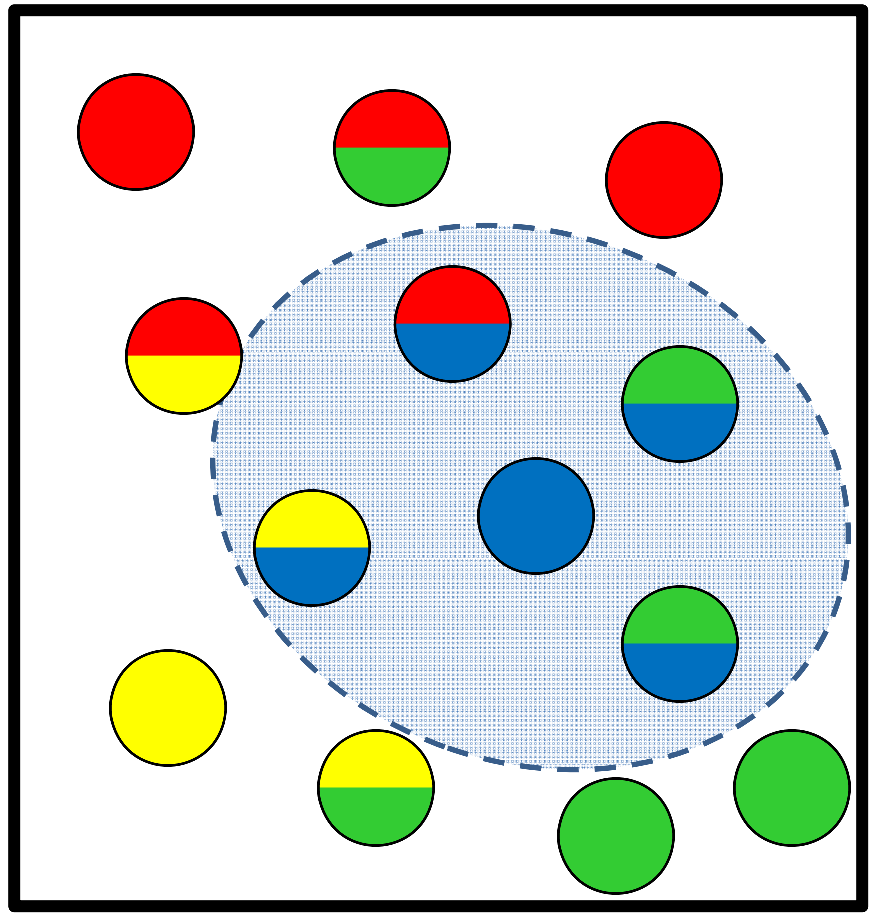
- Again, common solution is *to use One-vs-All* transformation
- At test, must make a binary decision for each of the C classes



Binary SVMs: The Problem

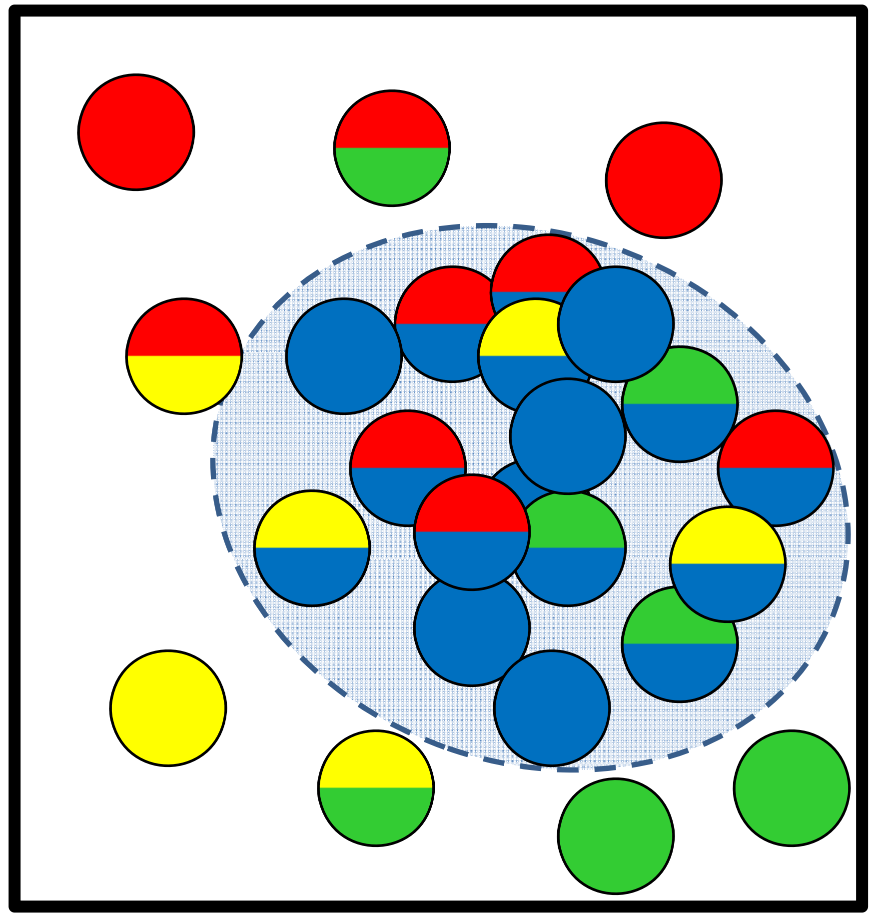
- All data is assumed to be relevant for each label
- Thus, *green, red* and *yellow* data is contaminating our model for *blue*
- Nonetheless, method often works well…

Why is this?



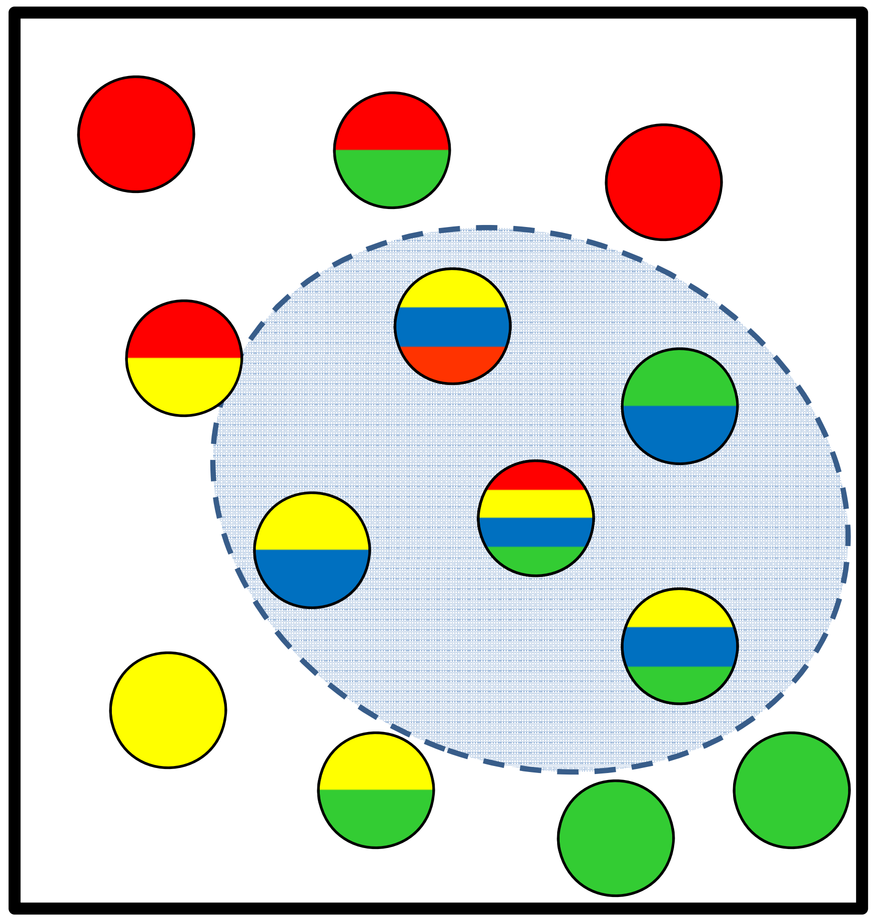
Binary SVMs: The Problem

- Depends on statistics of dataset
- Works fine with lots of training data



Binary SVMs: The Problem

- Depends on statistics of dataset
- Works fine with lots of training data
- However, becomes a problem as:
  1. The number of training documents becomes smaller
  2. The number of labels per document becomes larger



Example – Rare Label

NY Times Article

Labels	Label Freq.
ANTITRUST ACTIONS AND LAWS	19
SUITS AND LITIGATION	67
VIDEO GAMES *	1

Article Excerpt

A flurry of lawsuits, started by a small American software developer, now surrounds the Nintendo Entertainment System...Atari Games argues that Nintendo's high degree of control is tantamount to monopoly, and is suing Nintendo for antitrust violations...

Example – Rare Label

NY Times Article

Labels	**Label Freq.**
Antitrust Actions And Laws	19
Suits and Litigation	67
Video Games *	1

Article Excerpt

A flurry of lawsuits, started by a small American software developer, now surrounds the Nintendo Entertainment System...Atari Games argues that Nintendo's high degree of control is tantamount to monopoly, and is suing Nintendo for antitrust violations...

Example – Rare Label

Models for VIDEO GAMES
SVM
nintendo
mcgowan
futuristic
compatible
illusion
shrewd
inception
truthful
profiles
billionayear
suing
infringement
architecture
handheld
tantamount
payoff
glut
equitable

Example – Rare Label

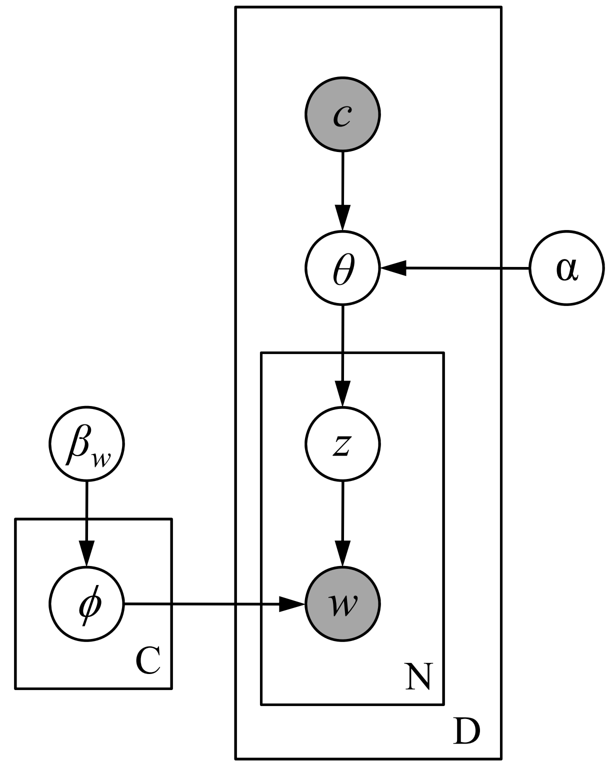
Models for VIDEO GAMES

SVM	**LDA**
nintendo	nintendo
mcgowan	games
futuristic	software
compatible	video
illusion	system
shrewd	game
inception	chip
truthful	control
profiles	market
billionayear	home
suing	computer
infringement	shortage
architecture	say
handheld	buy
tantamount	demand
payoff	developer
glut	generally
equitable	japan

Overview

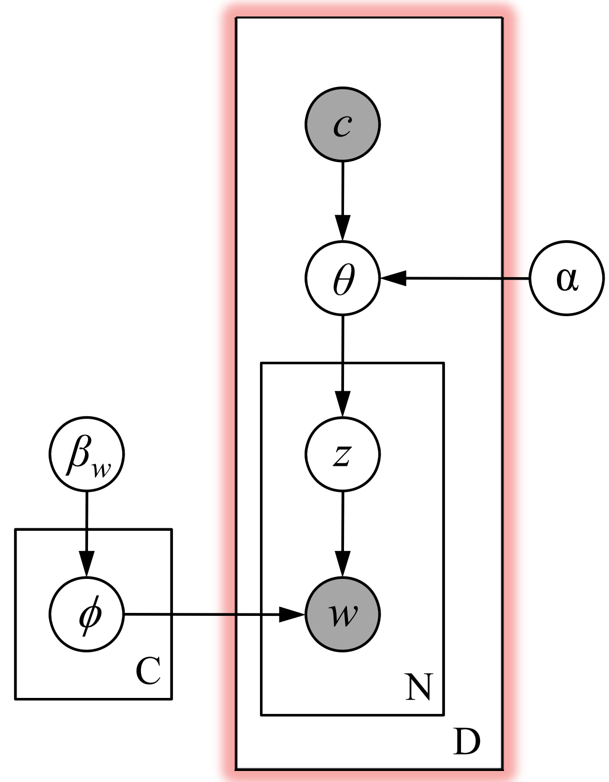
- Introduce the research problem
  - Research goals and challenges
- Motivate the use of probabilistic methods
- **Present our probabilistic models**
- Experimental Results

Flat LDA



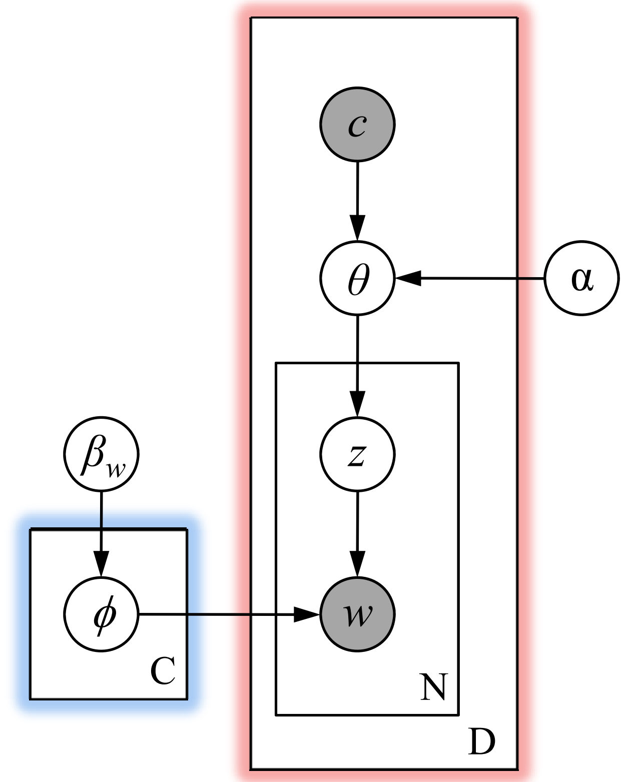
Flat LDA

- **Documents**



Flat LDA

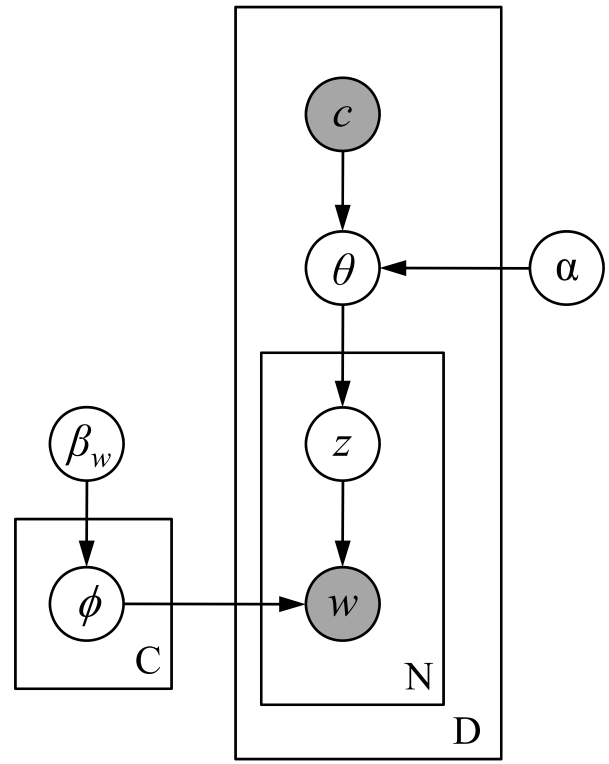
- **Documents**
- **Labels**



Flat LDA

- Documents
- Labels

Mixture Model

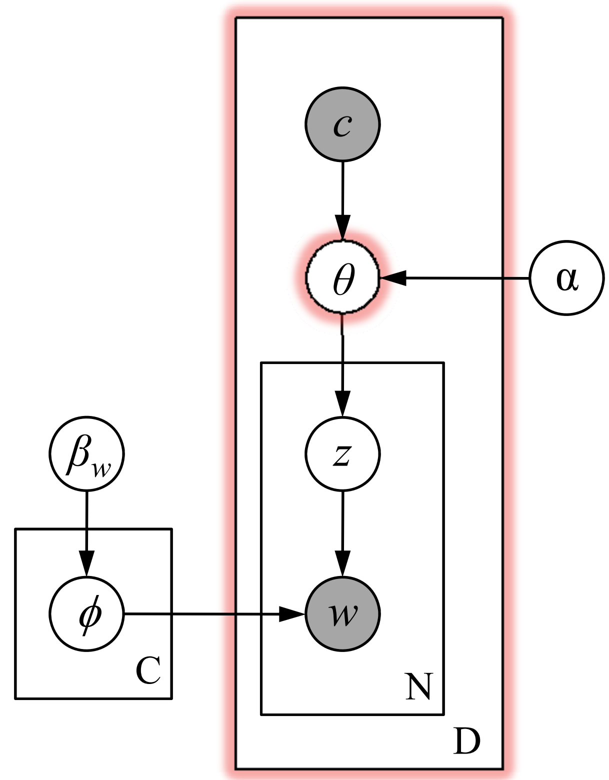


Flat LDA

- **Documents**
- **Labels**

Mixture Model

- Documents are mixtures of labels θ_d

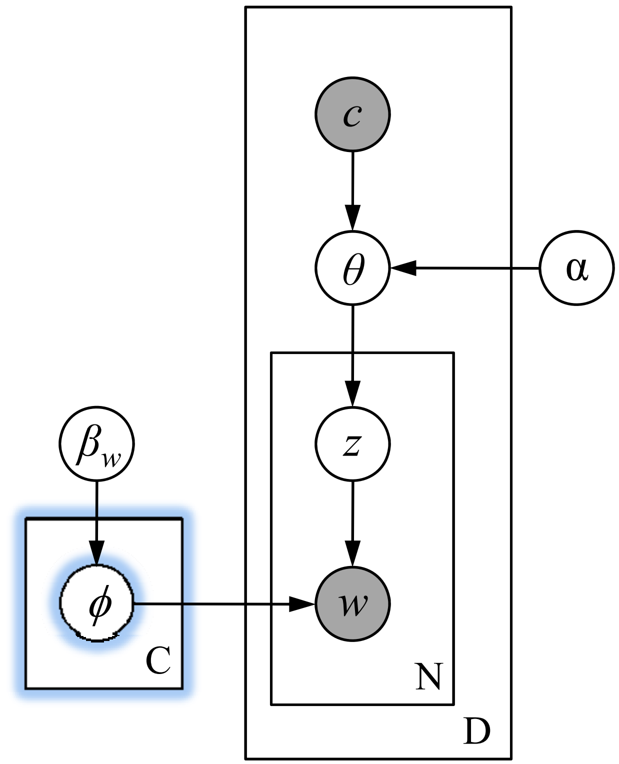


Flat LDA

- **Documents**
- **Labels**

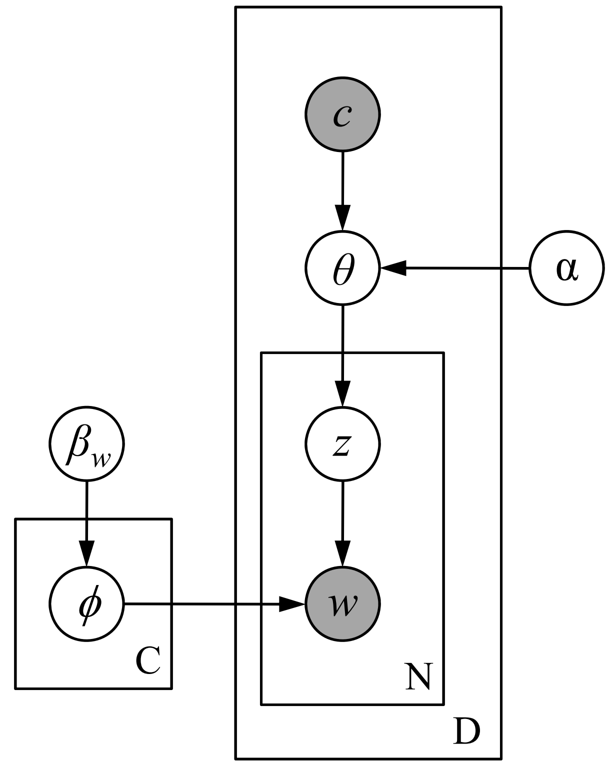
Mixture Model

- Documents are mixtures of labels θ_d
- Labels are probability distributions over words ϕ_c



Flat LDA

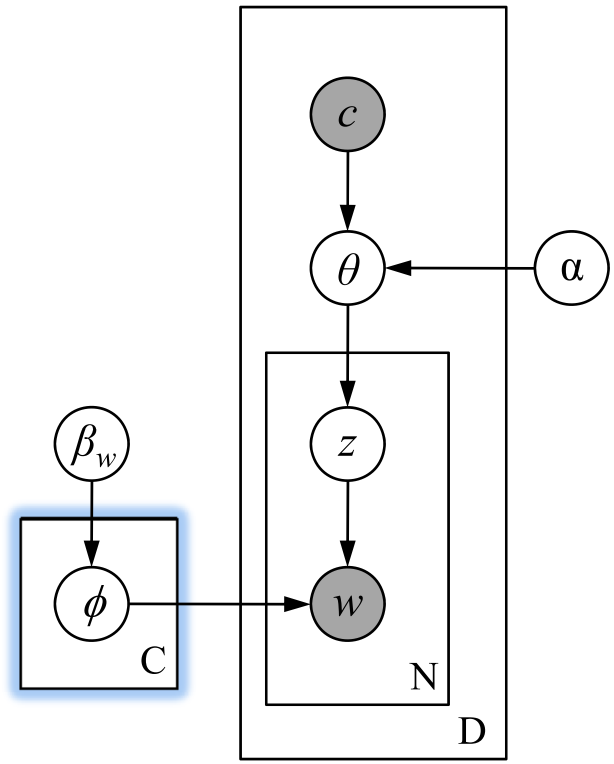
GENERATIVE PROCESS



Flat LDA

GENERATIVE PROCESS

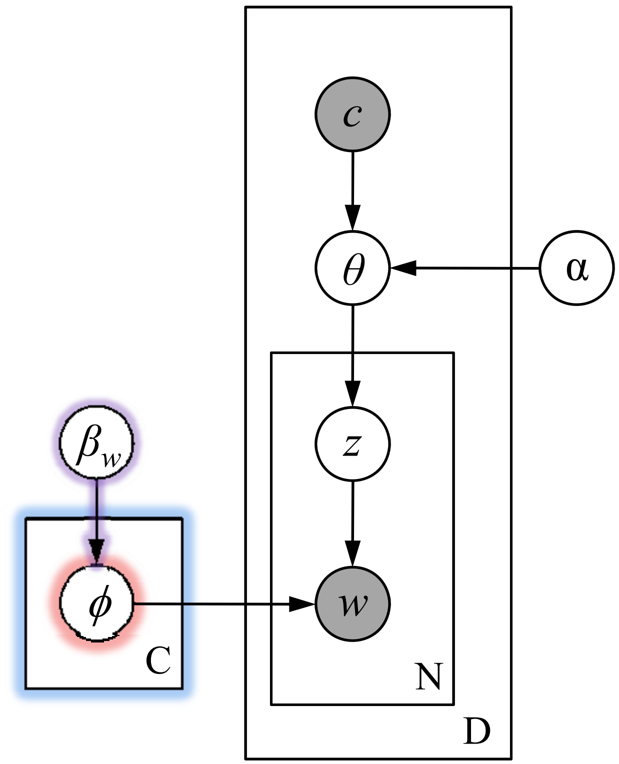
- For each label:



Flat LDA

GENERATIVE PROCESS

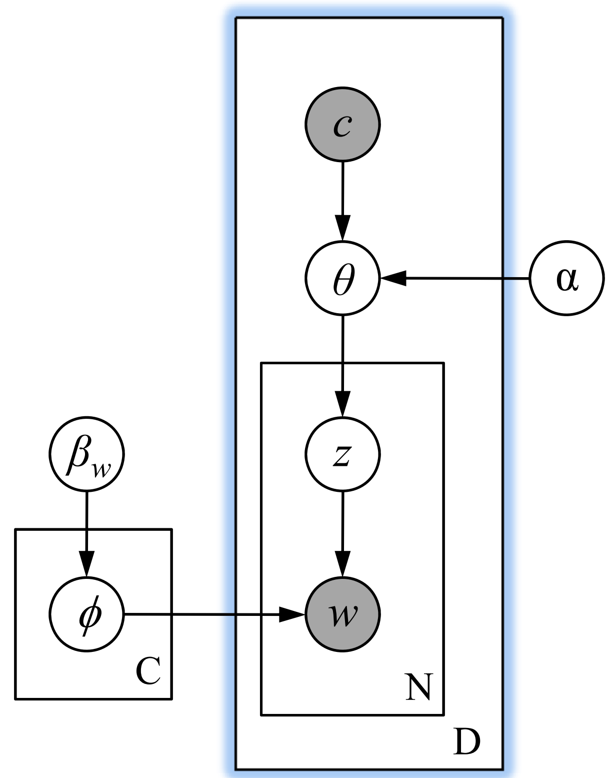
- For each label:
  - Sample a multinomial distribution over words ϕ_c from dirichlet β_w



Flat LDA

GENERATIVE PROCESS

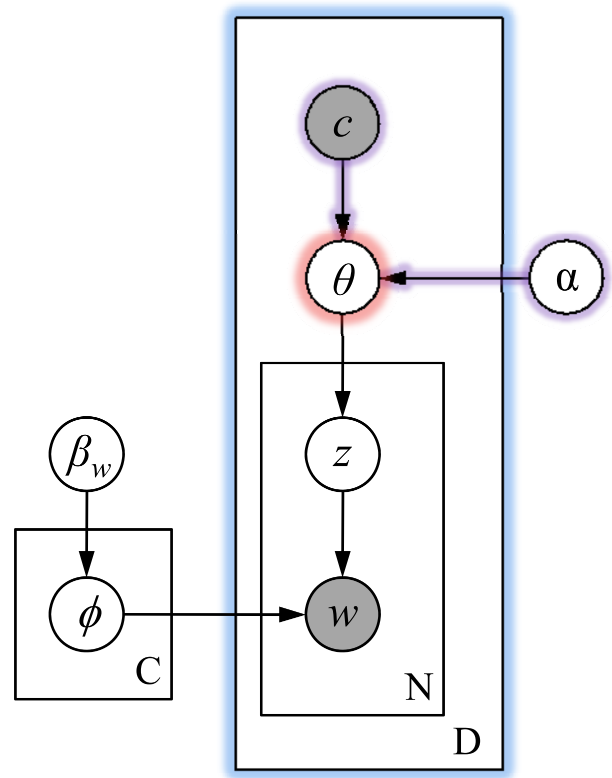
- For each label:
  - Sample a multinomial distribution over words ϕ_c from dirichlet β_w
- For each document:



Flat LDA

GENERATIVE PROCESS

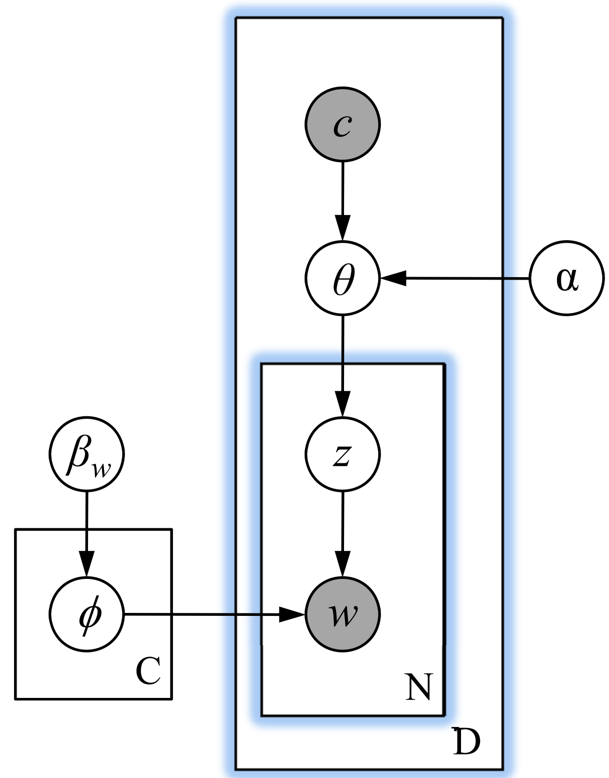
- For each label:
  - Sample a multinomial distribution over words ϕ_c from dirichlet β_w
- For each document:
  - Sample a multinomial distribution θ over the *observed* labels from dirichlet prior α'



Flat LDA

GENERATIVE PROCESS

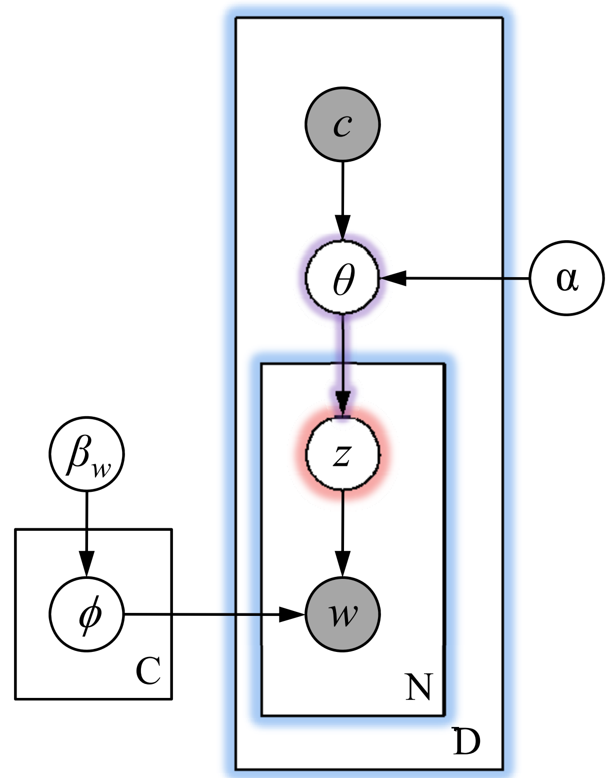
- For each label:
  - Sample a multinomial distribution over words ϕ_c from dirichlet β_w
- For each document:
  - Sample a multinomial distribution θ over the *observed* labels from dirichlet prior α'
- For each word in document:



Flat LDA

GENERATIVE PROCESS

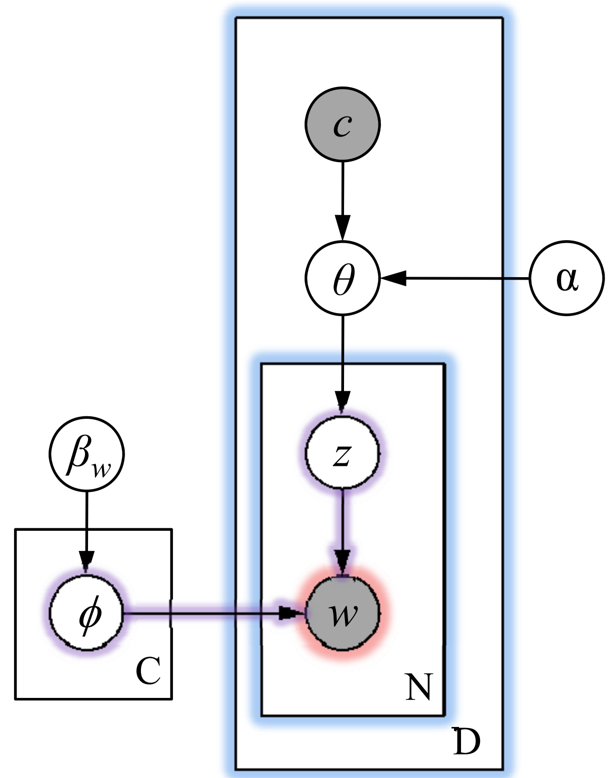
- For each label:
  - Sample a multinomial distribution over words ϕ_c from dirichlet β_w
- For each document:
  - Sample a multinomial distribution θ over the *observed* labels from dirichlet prior α'
- For each word in document:
  - Sample a label z



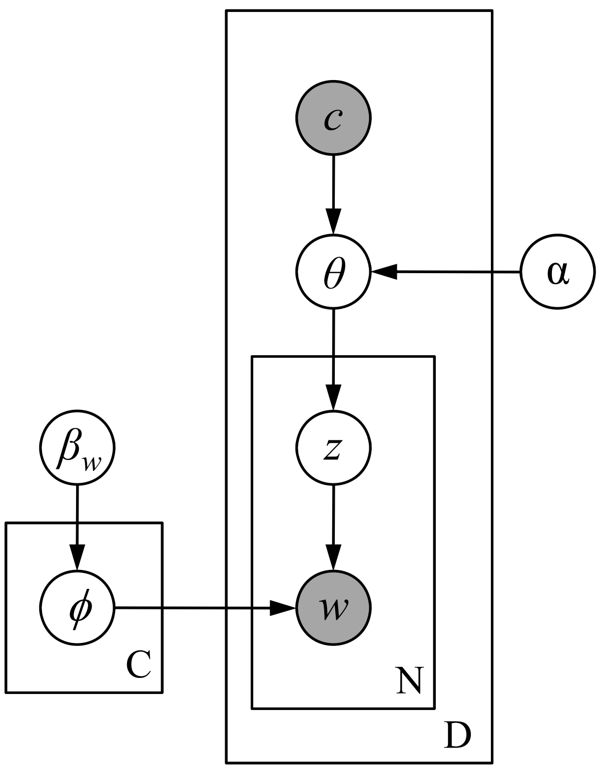
Flat LDA

GENERATIVE PROCESS

- For each label:
  - Sample a multinomial distribution over words ϕ_c from dirichlet β_w
- For each *document*:
  - Sample a multinomial distribution θ over the *observed* labels from dirichlet prior α'
- For each word in document:
  - Sample a label z
  - Generate a word w from ϕ_Z

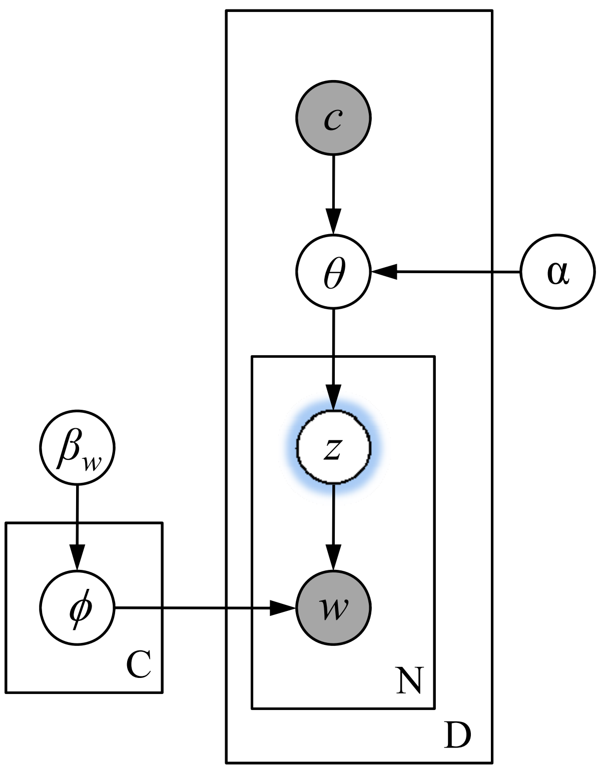


Flat LDA



- **Training:**
  - Learn a distribution over words for each label, ϕ_c

Flat LDA



- **Training:**
  - Learn a distribution over words for each label, ϕ_c
  - Collapsed Gibbs sampler:
    - iteratively updating the z assignments of words to labels

LDA and rare labels

- Labels are learned simultaneously
- Words associated with well-known labels will be "explained away"

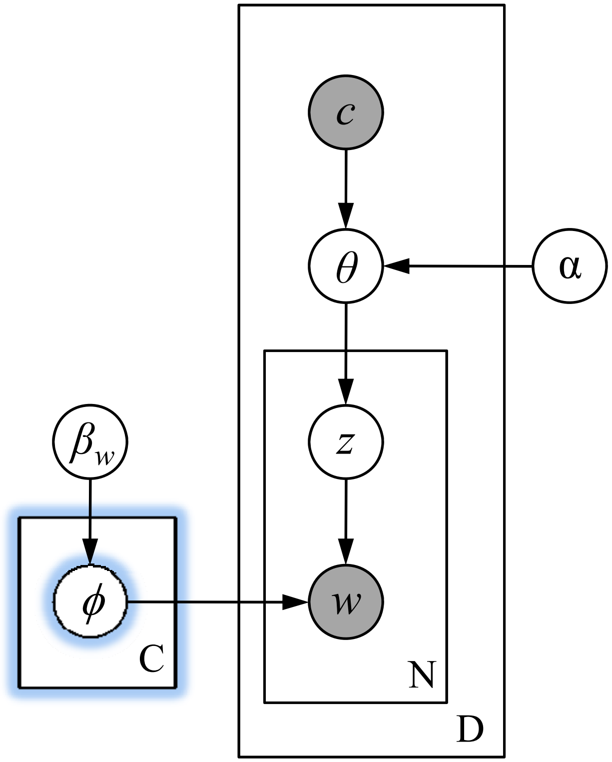
NY Times Article

Labels	Label Freq.
ANTITRUST ACTIONS AND LAWS	19
SUITS AND LITIGATION	67
VIDEO GAMES *	1

Article Excerpt

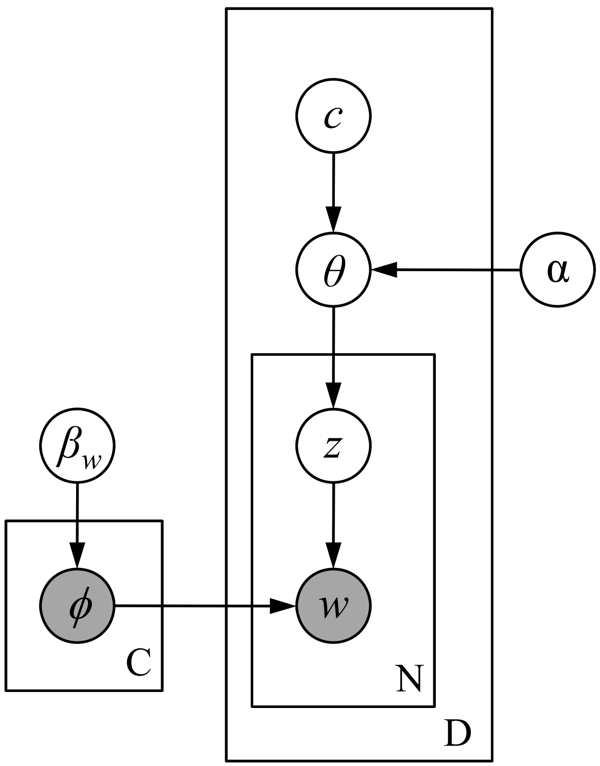
A flurry of lawsuits, started by a small American software developer, now surrounds the Nintendo Entertainment System...Atari Games argues that Nintendo's high degree of control is tantamount to monopoly, and is suing Nintendo for antitrust violations...

Label-Word Distributions



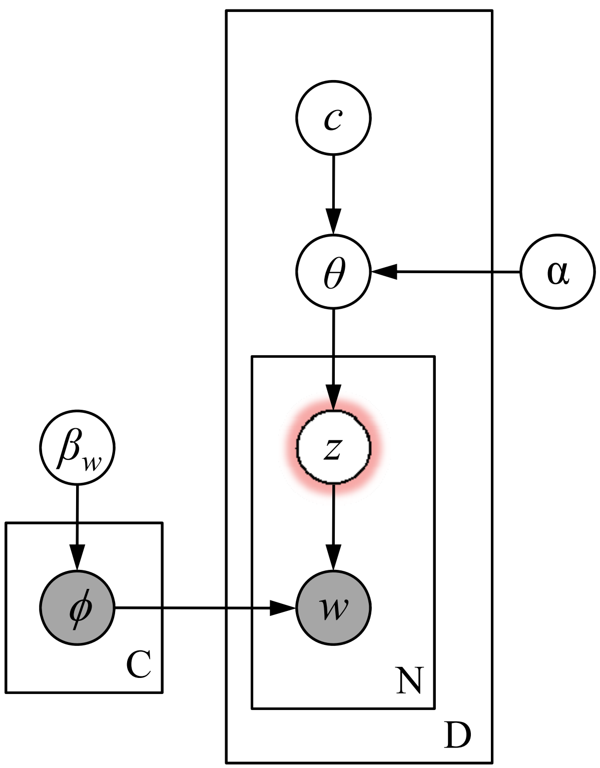
ARMS SALES ABROAD	176	ABORTION	24	ACID RAIN	11	AGNI MISSILE	1
iran	.021	abortion	.098	acid	.070	missile	.032
arms	.019	court	.033	rain	.067	india	.031
reagan	.014	abortions	.028	lakes	.028	technology	.016
house	.014	women	.017	environmental	.026	missiles	.016
president	.014	decision	.016	sulfur	.024	western	.015
north	.012	supreme	.016	study	.023	miles	.014
report	.011	rights	.015	emissions	.021	nuclear	.013
white	.011	judge	.015	plants	.021	indian	.013

Inference for test-Documents



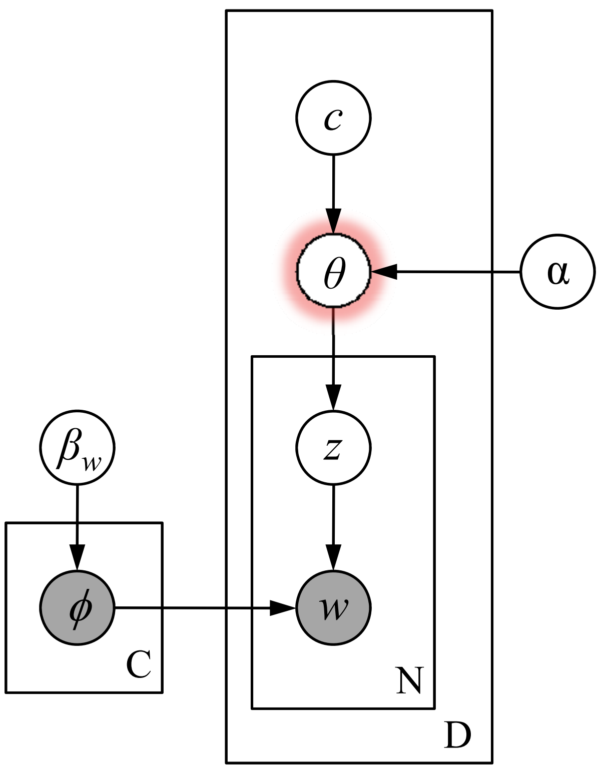
- **Testing:**
  - Treat ϕ as observed

Inference for test-Documents



- **Testing:**
  - Treat ϕ as observed
  - Update z assignments for each test document to learn θ

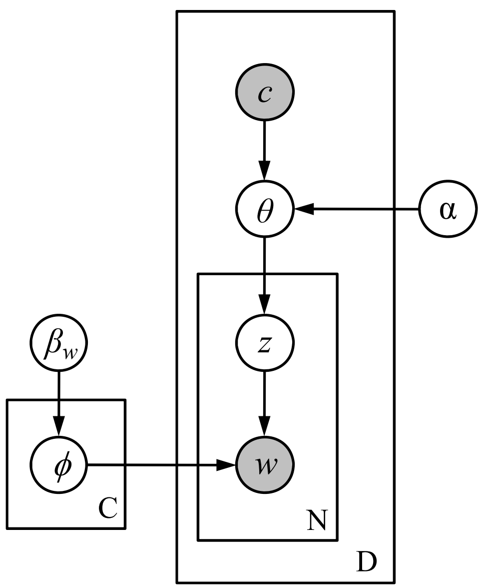
Inference for test-Documents



- **Testing:**
  - Treat ϕ as observed
  - Update z assignments for each test document to learn θ

- **Predictions:**
  - Use θ as a predicted ranking of labels
  - For binary predictions, threshold

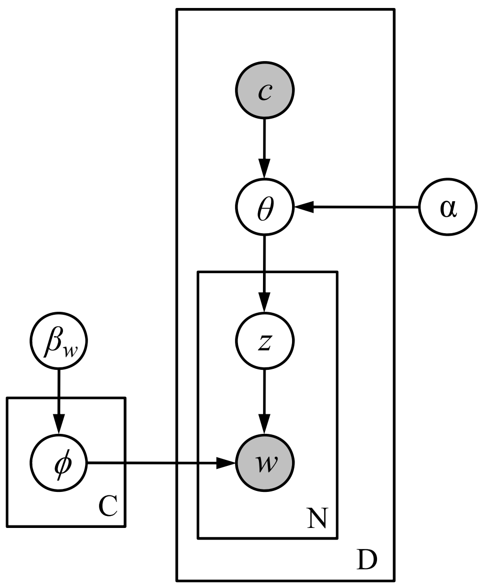
Model Extensions



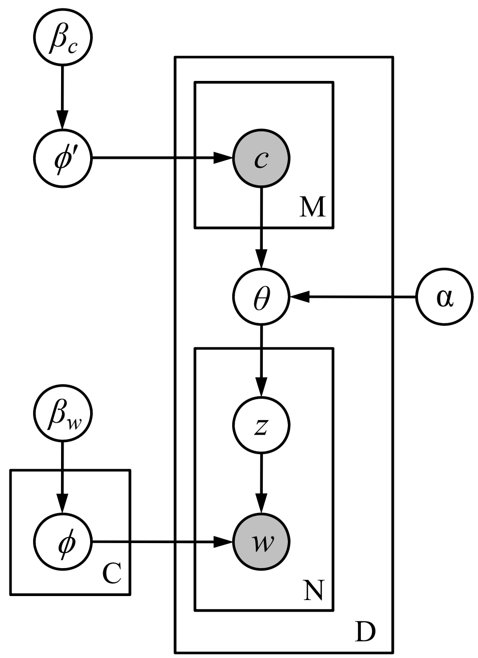
Flat-LDA

Model Extensions

Account for baseline label frequencies



Flat-LDA



Prior-LDA

Prior-LDA

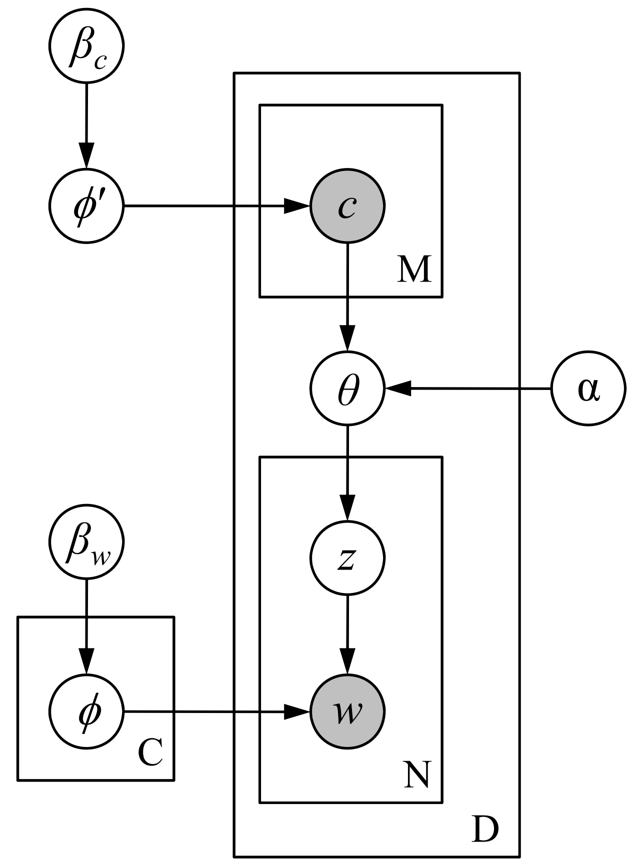
- Extends flat-LDA by incorporating a generative process for labels:
- Two-step generative process for each document

Generating Labels

- Sample labels, *c*, from a multinomial distribution ϕ' that reflects baseline label frequencies

Generating words

- Sample a document mixture over labels, θ, from a Dirichlet prior that is proportional to the number of times each label was sampled
  - α is now a vector of hyperparameters
- Given θ, generate each word as in Flat-LDA



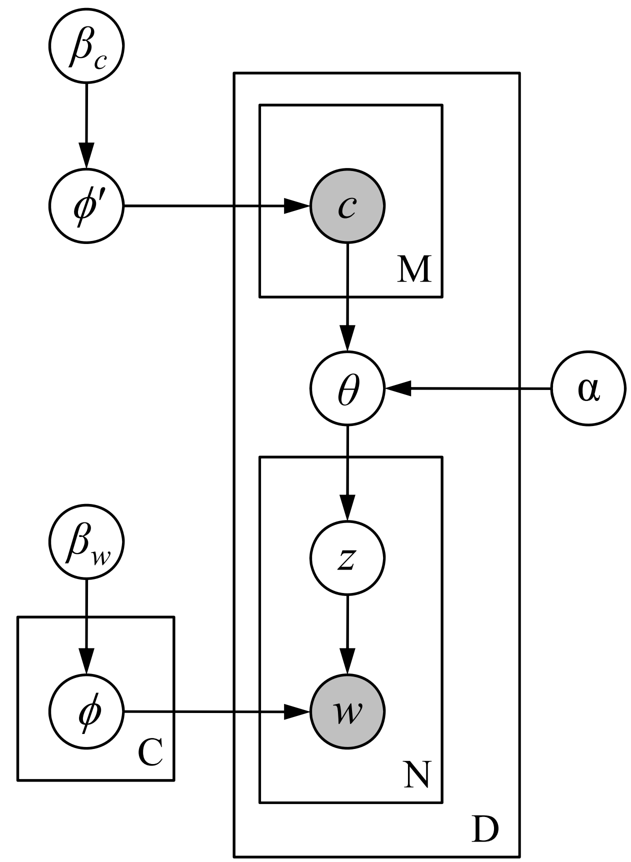
Prior-LDA

Training

- Learning ϕ is equivalent to Flat-LDA
  - Conditionally independent given ***C***
- Estimate ϕ' directly from data

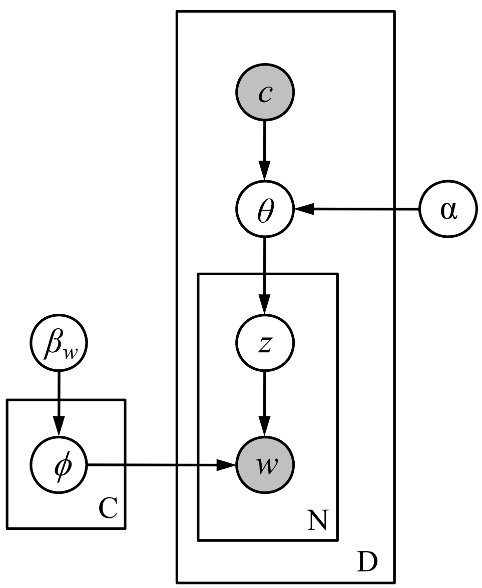
Test Time

- Same as Flat-LDA, except we now use an *informative* dirichlet prior on θ
- Vector of hyperparameters reflects baseline label freqs.

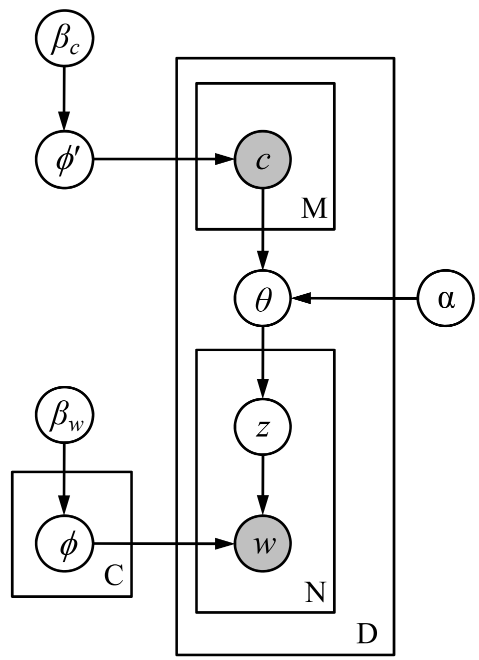


Model Extensions

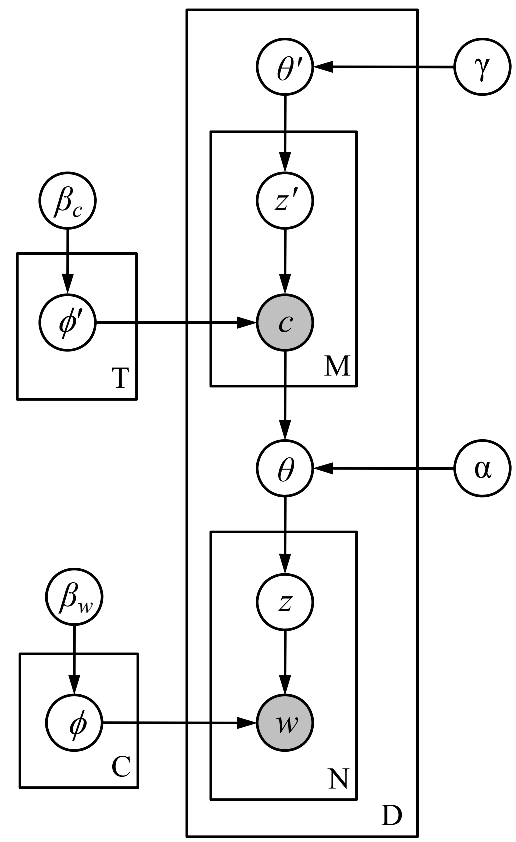
Account for label dependencies



Flat-LDA



Prior-LDA

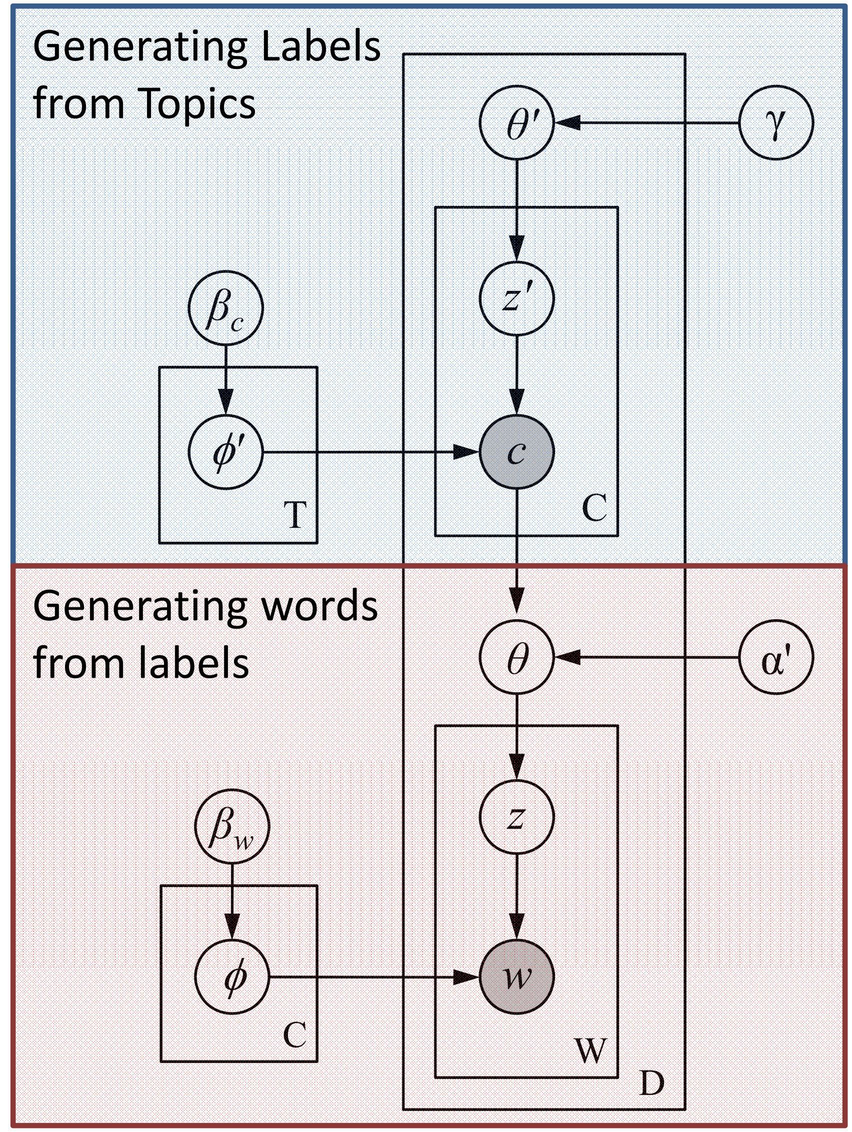


Dependency-LDA

Dependency-LDA

- Each document is a mixture of both
  - Topics θ'
  - Labels θ
- Topics are distributions over labels ϕ'
- Labels are distribution over words ϕ

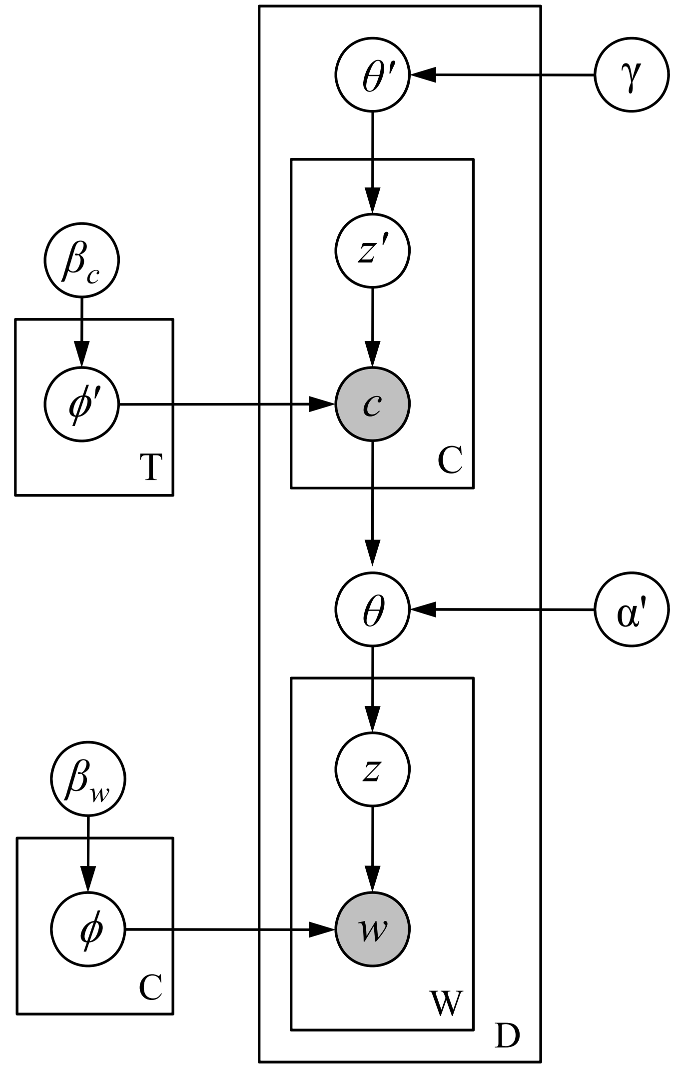
- Three stage generative process for each document
  1. Generate labels from topics
  2. Sample θ conditioned on these labels
  3. Generate words as before



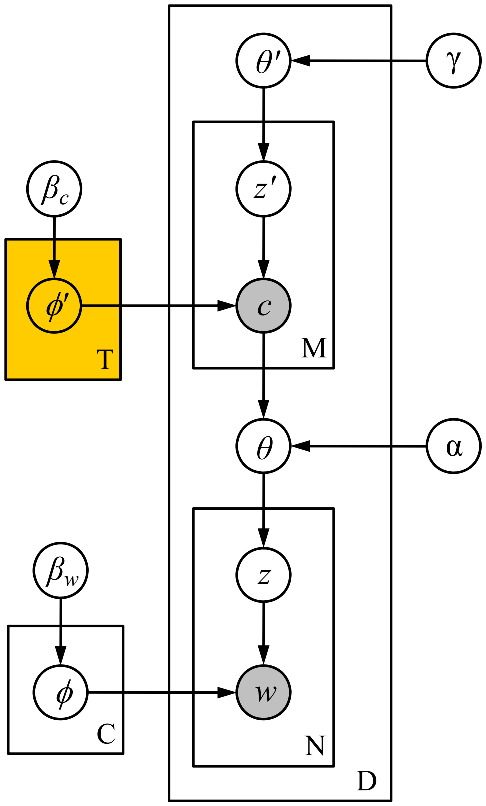
Inference

Training

- Independently train topic-word ϕ' and label-word ϕ distributions
  - *Conditionally independent given* $\boldsymbol{c}$



Topic-Label Distributions

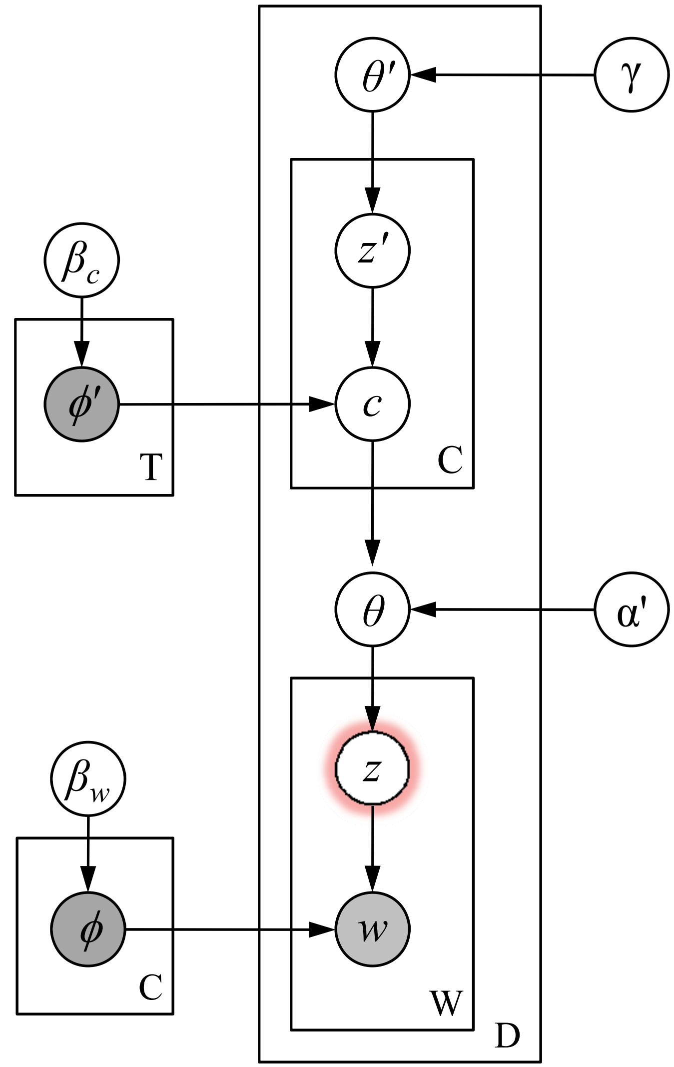


"Consumer Safety"	.017	"Warfare And Disputes"	.024
CANCER	.078	ARMAMENT, DEFENSE AND MILITARY...	.162
HAZARDOUS AND TOXIC SUBSTANCES	.039	INTERNATIONAL RELATIONS	.133
PESTICIDES AND PESTS	.021	UNITED STATES INTERNATIONAL RELA...	.132
RESEARCH	.021	CIVIL WAR AND GUERRILLA WARFARE	.098
SURGERY AND SURGEONS	.021	MILITARY ACTION	.053
TESTS AND TESTING	.021	CHEMICAL WARFARE	.029
FOOD	.018	REFUGEES AND EXPATRIATES	.019
RECALLS AND BANS OF PRODUCTS	.018	INDEPENDENCE MOVEMENTS	.013
CONSUMER PROTECTION	.016	BOUNDARIES AND TERRITORIAL ISSUES	.011
HEALTH, PERSONAL	.016	KURDS	.010

Inference

Test Time

- Fix ϕ and ϕ' distributions



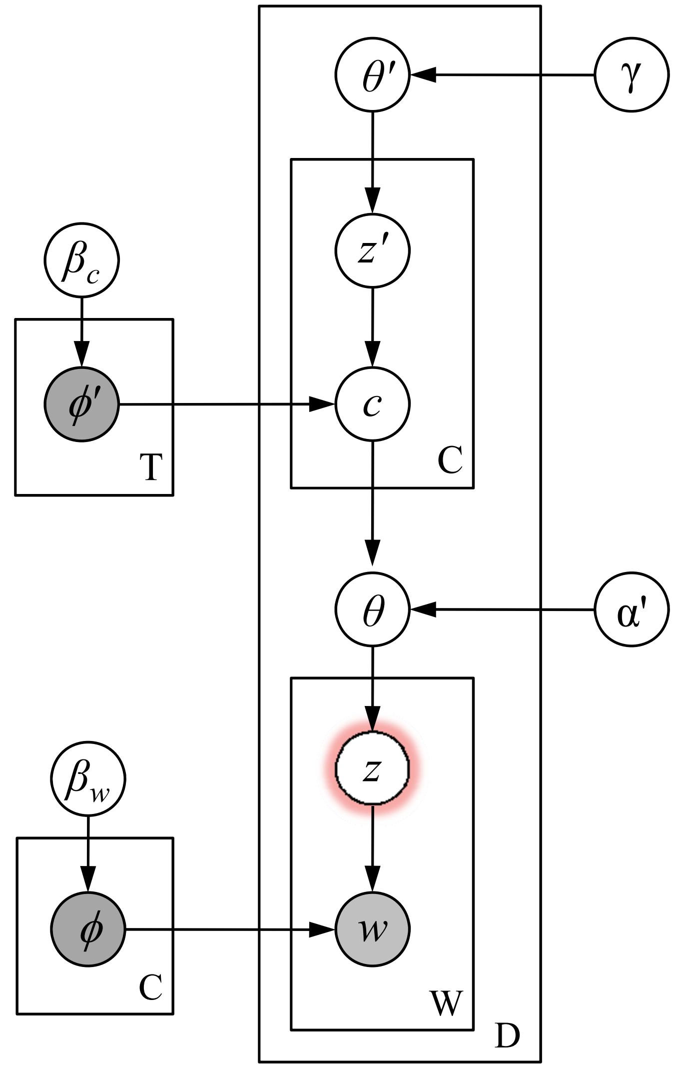
Inference

Test Time

- Fix ϕ and ϕ' distributions

Update Sequence

1. Update all z assignments



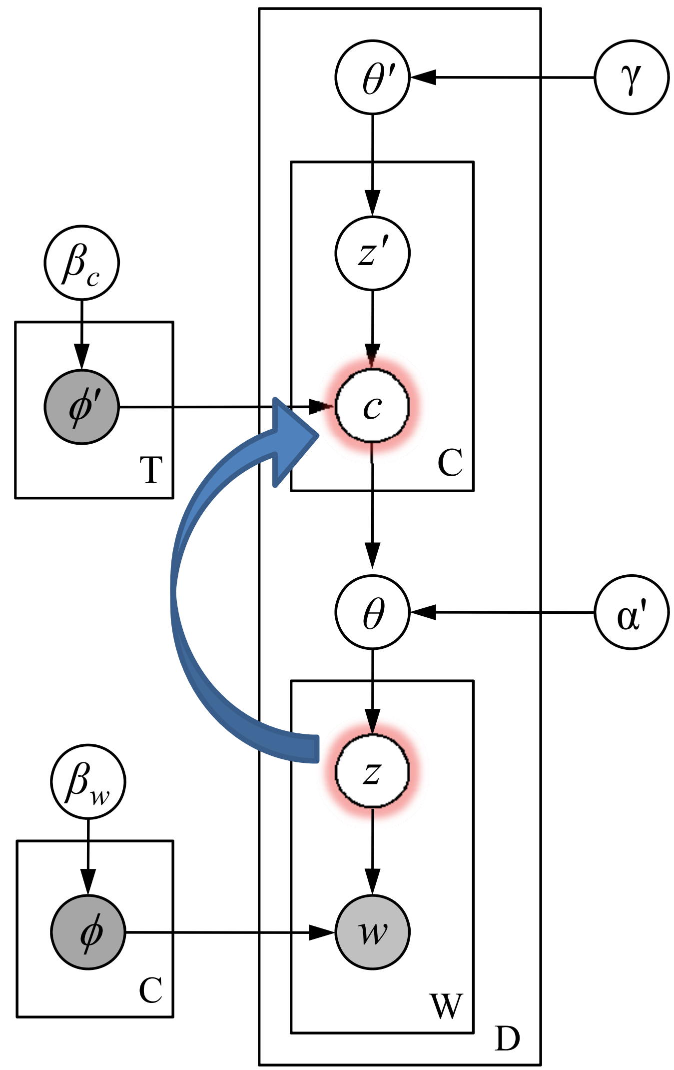
Inference

Test Time

- Fix ϕ and ϕ' distributions

Update Sequence

1. Update all z assignments
2. Set c equal to z



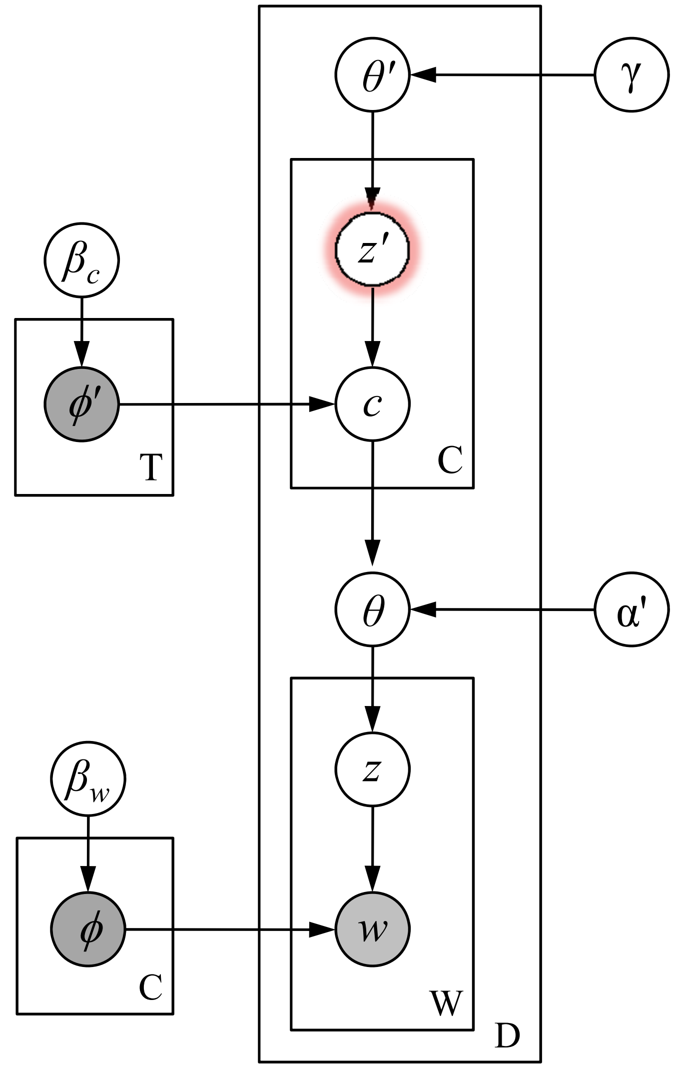
Inference

Test Time

- Fix ϕ and ϕ' distributions

Update Sequence

1. Update all z assignments
2. Set c equal to z
3. Update all z' assignments



Inference

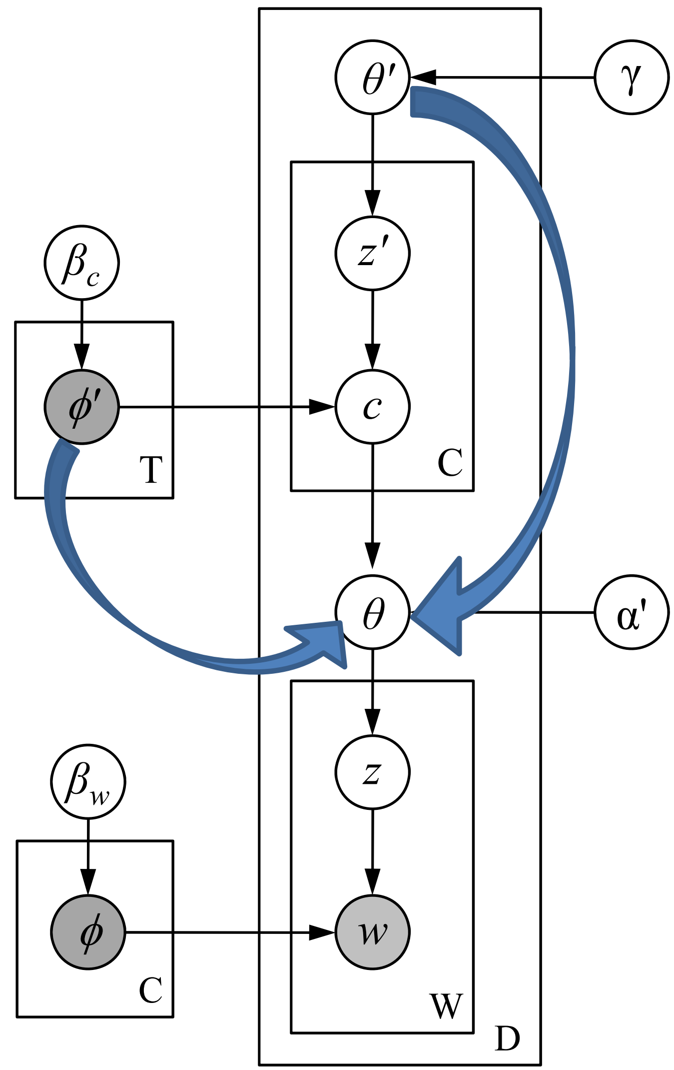
Test Time

- Fix ϕ and ϕ' distributions

Update Sequence

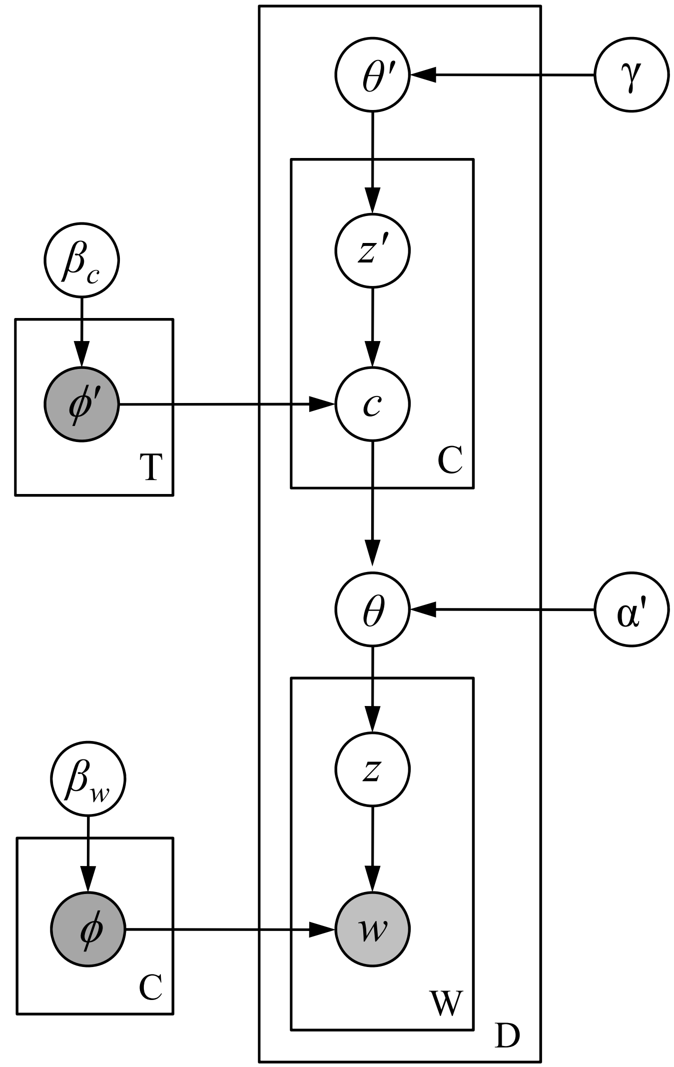
1. Update all z assignments
2. Set c equal to z
3. Update all z' assignments
4. Compute hyperparameter vector α':

$$\alpha'^{(d)} = \eta \cdot \phi'^{(\cdot)} \cdot \theta'^{(d)}$$



Inference

- Key here is that we learn a *document-specific* prior on θ
  - Reflects label dependencies



Example – Predictions

New York Times Article

LEAD: The special Senate and House committees investigating the Iran-contra affair decided today to hold joint hearings, and set a timetable for granting limited immunity from prosecution to the two central witnesses.The extraordinary agreement, which also calls for merging the committee staffs and for sharing evidence, is expected to speed the inquiry...

True Document Labels	Label Freq.
IMMUNITY FROM PROSECUTION	4
ARMS SALES ABROAD	176
ARMAMENT, DEFENSE AND MILITARY FORCES	409
UNITED STATES INTERNATIONAL RELATIONS	630

Example – Predictions

New York Times Article

LEAD: The special Senate and House committees investigating the Iran-contra affair decided today to hold joint hearings, and set a timetable for granting limited immunity from prosecution to the two central witnesses.The extraordinary agreement, which also calls for merging the committee staffs and for sharing evidence, is expected to speed the inquiry...

True Document Labels	Label Freq.
IMMUNITY FROM PROSECUTION	4
ARMS SALES ABROAD	176
ARMAMENT, DEFENSE AND MILITARY FORCES	409
UNITED STATES INTERNATIONAL RELATIONS	630

Binary SVMs	
1	CONGRESSIONAL INVESTIGATIONS
2	**ARMS SALES ABROAD**
3	**ARMAMENT, DEFENSE AND MILITARY FORCES**
4	**UNITED STATES INTERNATIONAL RELATIONS**
5	CIVIL WAR AND GUERRILLA WARFARE
6	ETHICS
7	DISCLOSURE OF INFORMATION
8	FOREIGN AID
9	UNITED STATES ARMAMENT AND DEFENSE
10	LAW AND LEGISLATION

Example – Predictions

New York Times Article

LEAD: The special Senate and House committees investigating the Iran-contra affair decided today to hold joint hearings, and set a timetable for granting limited immunity from prosecution to the two central witnesses.The extraordinary agreement, which also calls for merging the committee staffs and for sharing evidence, is expected to speed the inquiry...

True Document Labels	Label Freq.
IMMUNITY FROM PROSECUTION	4
ARMS SALES ABROAD	176
ARMAMENT, DEFENSE AND MILITARY FORCES	409
UNITED STATES INTERNATIONAL RELATIONS	630

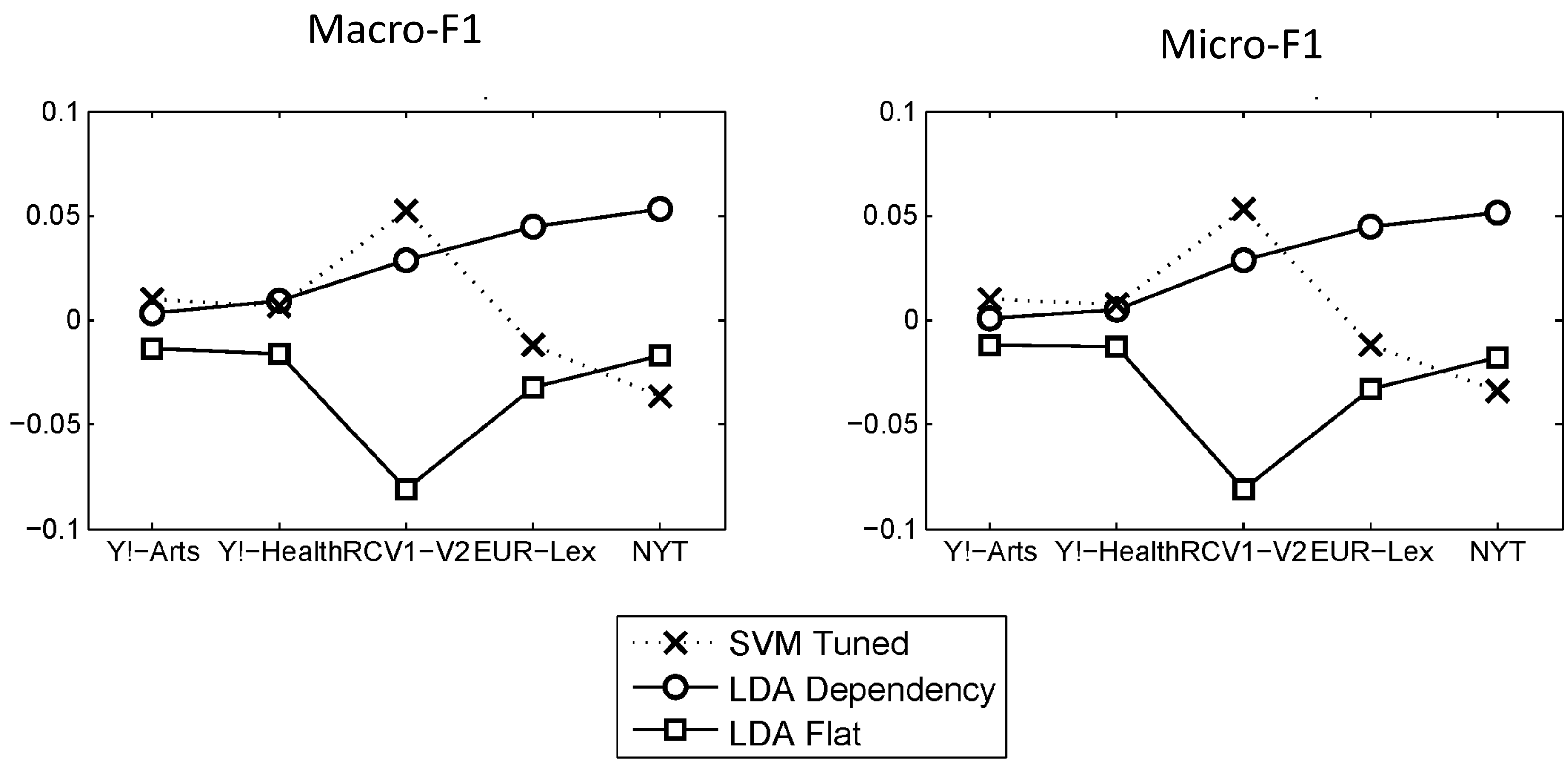
	Binary SVMs	LDA - Flat	p	LDA - Prior	p	LDA - Dependencies	p
1	CONGRESSIONAL INVESTIGATIONS	**ARMS SALES ABROAD**	.204	**ARMS SALES ABROAD**	.261	**ARMS SALES ABROAD**	.291
2	**ARMS SALES ABROAD**	CONGRESSIONAL INVESTIGATIONS	.182	CONGRESSIONAL INVESTIGATIONS	.237	CONGRESSIONAL INVESTIGATIONS	.234
3	**ARMAMENT, DEFENSE AND MILITARY FORCES**	LAW AND LEGISLATION	.059	LAW AND LEGISLATION	.102	**UNITED STATES INTERNATIONAL RELATIONS**	.110
4	**UNITED STATES INTERNATIONAL RELATIONS**	**IMMUNITY FROM PROSECUTION**	.042	**IMMUNITY FROM PROSECUTION**	.062	LAW AND LEGISLATION	.100
5	CIVIL WAR AND GUERRILLA WARFARE	ETHICS	.004	ETHICS	.045	**IMMUNITY FROM PROSECUTION**	.063
6	ETHICS	MIDGETMAN (MISSILE)	.003	**UNITED STATES INTERNATIONAL RELATIONS**	.024	**ARMAMENT, DEFENSE AND MILITARY FORCES**	.049
7	DISCLOSURE OF INFORMATION	VETOES (US)	.003	TRIALS	.018	CIVIL WAR AND GUERRILLA WARFARE	.014
8	FOREIGN AID	UNITED STATES ARMAMENT AND DEFENSE	.003	UNITED STATES ARMAMENT AND DEFENSE	.010	DECISIONS AND VERDICTS	.007
9	UNITED STATES ARMAMENT AND DEFENSE	CONGRESSIONAL COMMITTEES	.003	INTERNATIONAL RELATIONS	.008	FOREIGN AID	.007
10	LAW AND LEGISLATION	B-2 AIRPLANE	.003	FINANCES	.007	TRIALS	.005

Overview

- Introduce the research problem
  - Research goals and challenges
- Motivate the use of probabilistic methods
- Present our probabilistic models
- **Experimental Results**

Full set of experiments

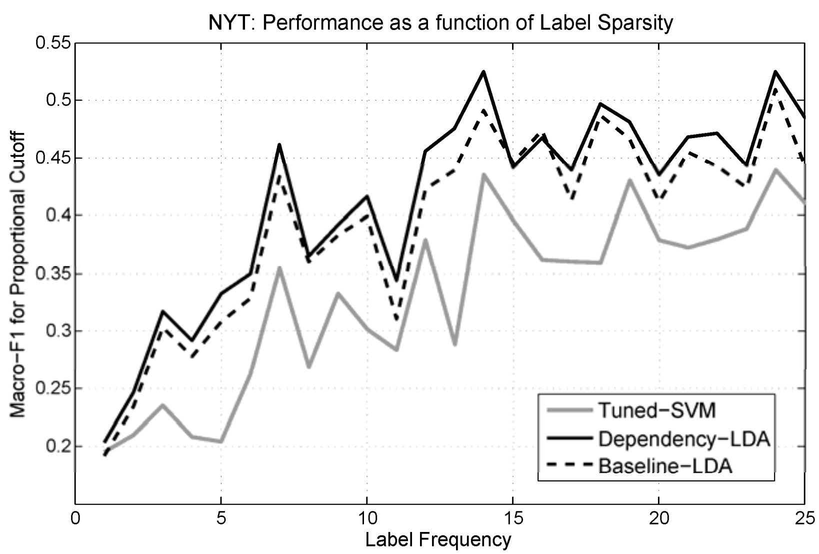
- 5 datasets
- 5 models
  - 2 SVM methods
  - 3 LDA-based models
- 2 prediction tasks
  - 13 evaluation metrics



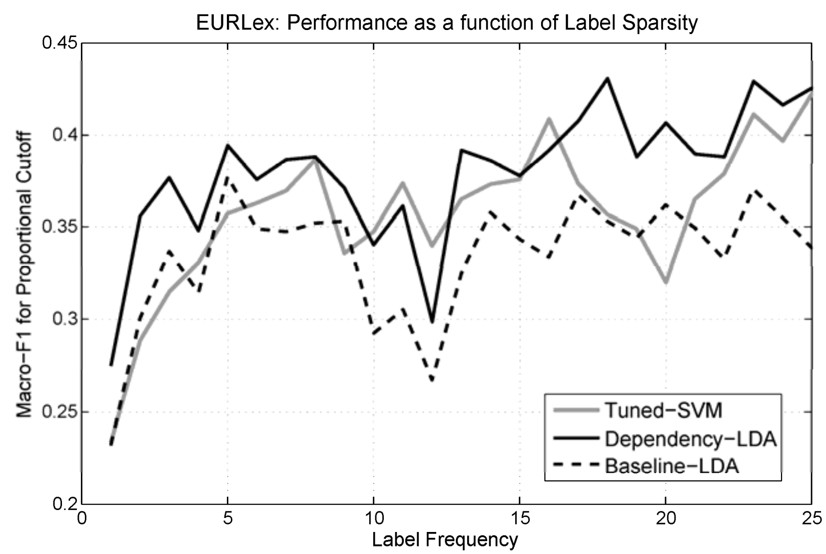
Dataset results, sorting in terms of *increasing* #labels/document

F1 Performance on Rare Labels

NYT



EUR-Lex



Dependency-LDA vs. SVMs

DATASET	MEDIAN FREQ.	MODEL
NYT	3	LDA $_{\text{Dependency}}$
		SVM $_{\text{Tuned}}$
EURLex	6	LDA $_{\text{Dependency}}$
		SVM $_{\text{Tuned}}$
Y! *Arts*	530	LDA $_{\text{Dependency}}$
		SVM $_{\text{Tuned}}$
Y! *Health*	500	LDA $_{\text{Dependency}}$
		SVM $_{\text{Tuned}}$
RCV1-V2	7,410	LDA $_{\text{Dependency}}$
		SVM $_{\text{Tuned}}$

Dependency-LDA vs. SVMs

DATASET	MEDIAN FREQ.	MODEL
NYT	3	$\text{LDA}_{\text{Dependency}}$
		$\text{SVM}_{\text{Tuned}}$
EURLex	6	$\text{LDA}_{\text{Dependency}}$
		$\text{SVM}_{\text{Tuned}}$
Y! *Arts*	530	$\text{LDA}_{\text{Dependency}}$
		$\text{SVM}_{\text{Tuned}}$
Y! *Health*	500	$\text{LDA}_{\text{Dependency}}$
		$\text{SVM}_{\text{Tuned}}$
RCV1-V2	7,410	$\text{LDA}_{\text{Dependency}}$
		$\text{SVM}_{\text{Tuned}}$

Power Law

Benchmark

Dependency-LDA vs. SVMs

Evaluation Metrics

			DOCUMENT-PIVOTED				LABEL-PIVOTED				
DATASET	**MEDIAN FREQ.**	**MODEL**	AUC_{ROC}	Avg-Prec	$F1_{MACRO}$	$F1_{MICRO}$	AUC_{ROC}	Avg-Prec	$F1_{MACRO}$	$F1_{MICRO}$	**TOTAL BOLD**
NYT	3	$LDA_{Dependency}$									
		SVM_{Tuned}									
EURLex	6	$LDA_{Dependency}$									
		SVM_{Tuned}									
Y! *Arts*	530	$LDA_{Dependency}$									
		SVM_{Tuned}									
Y! *Health*	500	$LDA_{Dependency}$									
		SVM_{Tuned}									
RCV1-V2	7,410	$LDA_{Dependency}$									
		SVM_{Tuned}									

Dependency-LDA vs. SVMs

Evaluation Metrics

			DOCUMENT-PIVOTED				LABEL-PIVOTED				
DATASET	**MEDIAN FREQ.**	**MODEL**	AUC_{ROC}	Avg-Prec	F1_{MACRO}	F1_{MICRO}	AUC_{ROC}	Avg-Prec	F1_{MACRO}	F1_{MICRO}	**TOTAL BOLD**
NYT	3	$\text{LDA}_{\text{Dependency}}$									
		$\text{SVM}_{\text{Tuned}}$									
EURLex	6	$\text{LDA}_{\text{Dependency}}$									
		$\text{SVM}_{\text{Tuned}}$									
Y! *Arts*	530	$\text{LDA}_{\text{Dependency}}$									
		$\text{SVM}_{\text{Tuned}}$									
Y! *Health*	500	$\text{LDA}_{\text{Dependency}}$									
		$\text{SVM}_{\text{Tuned}}$									
RCV1-V2	7,410	$\text{LDA}_{\text{Dependency}}$									
		$\text{SVM}_{\text{Tuned}}$									

Rankings

Binary Preds.

Dependency-LDA vs. SVMs

Evaluation Metrics

DATASET	MEDIAN FREQ.	MODEL	DOCUMENT-PIVOTED AUC $_{ROC}$	Avg-Prec	F1 $_{MACRO}$	F1 $_{MICRO}$	LABEL-PIVOTED AUC $_{ROC}$	Avg-Prec	F1 $_{MACRO}$	F1 $_{MICRO}$	TOTAL BOLD
NYT	3	LDA $_{Dependency}$	**.991**	**.631**	**.542**	**.539**	**.958**	**.383**	**.325**	**.541**	**8**
		SVM $_{Tuned}$	.965	.492	.453	.453	**.959**	.309	.270	.487	1
EURLex	6	LDA $_{Dependency}$	**.982**	**.511**	**.458**	**.461**	**.958**	**.472**	**.382**	**.467**	**8**
		SVM $_{Tuned}$	.967	.430	.402	.405	**.960**	.466	.373	**.471**	2
Y! *Arts*	530	LDA $_{Dependency}$	**.855**	**.630**	.454	.416	**.755**	**.341**	**.367**	**.451**	**6**
		SVM $_{Tuned}$	.833	.625	**.461**	**.425**	**.757**	.330	.355	**.454**	**4**
Y! *Health*	500	LDA $_{Dependency}$	**.926**	**.805**	**.619**	**.577**	**.850**	**.569**	.562	.646	**6**
		SVM $_{Tuned}$	.898	.788	**.617**	**.580**	**.849**	**.570**	**.571**	**.656**	**6**
RCV1-V2	7,410	LDA $_{Dependency}$	**.987**	.873	.743	.733	.971	.559	.539	.762	1
		SVM $_{Tuned}$	**.988**	**.896**	**.767**	**.757**	**.981**	**.608**	**.579**	**.787**	**8**

Dependency-LDA vs. SVMs

SCALE OF DIFFERENCES

DATASET	DOCUMENT-PIVOTED AUC ROC	DOCUMENT-PIVOTED Avg-Prec	DOCUMENT-PIVOTED F1 MACRO	DOCUMENT-PIVOTED F1 MICRO	LABEL-PIVOTED AUC ROC	LABEL-PIVOTED Avg-Prec	LABEL-PIVOTED F1 MACRO	LABEL-PIVOTED F1 MICRO	AVG.
NYT	.026	.139	.090	.085	-.001	.074	.055	.054	.065
EURLex	.015	.081	.057	.056	-.002	.006	.009	-.004	.027
Y! *Arts*	.022	.005	-.007	-.009	-.002	.010	.012	-.004	.003
Y! *Health*	.027	.017	.002	-.003	.001	-.001	-.009	-.010	.003
RCV1-V2	-.001	-.023	-.024	-.024	-.010	-.049	-.041	-.025	-.025

GREEN: LDA > SVM

RED: SVM>LDA

Dependency-LDA vs. SVMs

SCALE OF DIFFERENCES

DATASET	DOCUMENT-PIVOTED AUC $_{ROC}$	DOCUMENT-PIVOTED Avg-Prec	DOCUMENT-PIVOTED F1 $_{MACRO}$	DOCUMENT-PIVOTED F1 $_{MICRO}$	LABEL-PIVOTED AUC $_{ROC}$	LABEL-PIVOTED Avg-Prec	LABEL-PIVOTED F1 $_{MACRO}$	LABEL-PIVOTED F1 $_{MICRO}$	AVG.
NYT	.026	.139	.090	.085	-.001	.074	.055	.054	.065
EURLex	.015	.081	.057	.056	-.002	.006	.009	-.004	.027
Y! *Arts*	.022	.005	-.007	-.009	-.002	.010	.012	-.004	.003
Y! *Health*	.027	.017	.002	-.003	.001	-.001	-.009	-.010	.003
RCV1-V2	-.001	-.023	-.024	-.024	-.010	-.049	-.041	-.025	-.025

GREEN: LDA > SVM

RED: SVM>LDA

Comparison with Published Values

Comparison with Published Values

		Avg-Prec	Rnk-Loss	One-Err	Is-Err	Margin
--	MLNB	1.1	22.9	100.0	99.6	1644.0
1 epoch	BR	26.9	40.4	48.7	98.6	3230.7
2 epochs	BR	31.6	35.5	41.5	98.2	3050.1
5 epochs	BR	35.9	31.0	37.3	97.2	2842.6
1 epoch	MMP	29.3	3.9	75.9	98.8	597.6
2 epochs	MMP	39.5	4.4	54.4	97.5	694.1
5 epochs	MMP	47.3	4.7	40.2	96.0	761.2
1 epoch	DMLPP	46.7	2.8	35.5	97.9	433.9
2 epochs	DMLPP	**52.3**	2.5	**29.5**	**96.6**	397.1
	$LDA_{Dependency}$	**51.1**	**1.8**	32.0	97.2	**269.2**

Mencia, Furnkranz (2010)

Future Work

- Hybrid Discriminative / Generative extensions to the model
- Alternative methods for combining strength of both discriminative and generative approaches
- Any suggestions…?

Special Thanks

- America Holloway
- Padhraic Smyth
- Mark Steyvers
- All MADLabbers