

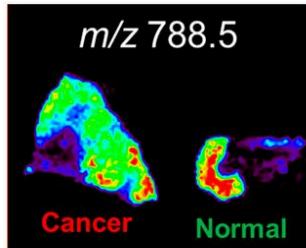
# Machine Learning & Data Mining

## CS/CNS/EE 155

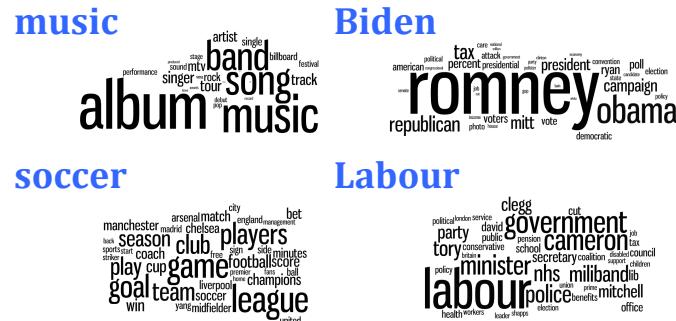
Lecture 13:  
Recent Applications

# Today: Three Recent Applications

# Lasso Cancer Detection



# Personalization via twitter

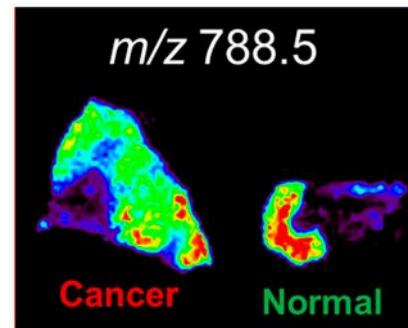


# Learning Visual Style

Slide material borrowed from Rob Tibshirani, Khalid El-Arini, and Julian McAuley

Image Sources: <http://www.pnas.org/content/111/7/2436>  
<https://dl.acm.org/citation.cfm?id=2487596>  
<http://www.cs.cornell.edu/~andreas/iccv15.pdf>

# Lasso Cancer Detection



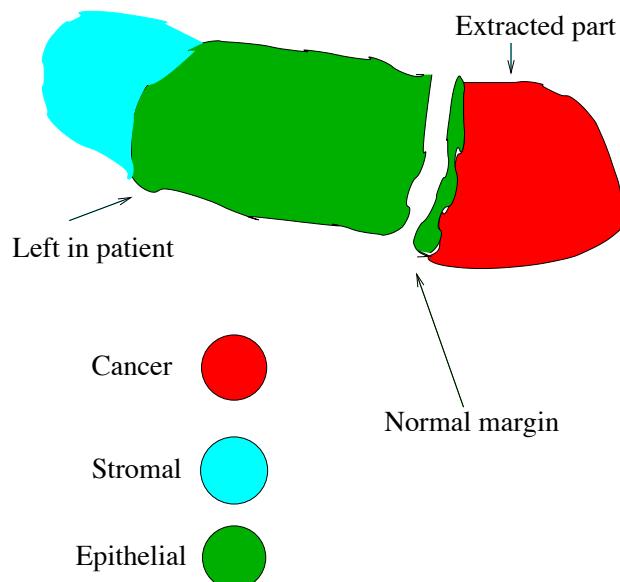
# “Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging”

**Proceedings of the National Academy of Sciences (2014)**

Livia S. Eberlin, Robert Tibshirani, Jialing Zhang, Teri Longacre, Gerald Berry, David B. Bingham, Jeffrey Norton, Richard N. Zare, and George A. Poulsides

<http://www.pnas.org/content/111/7/2436>

<http://statweb.stanford.edu/~tibs/ftp/canc.pdf>

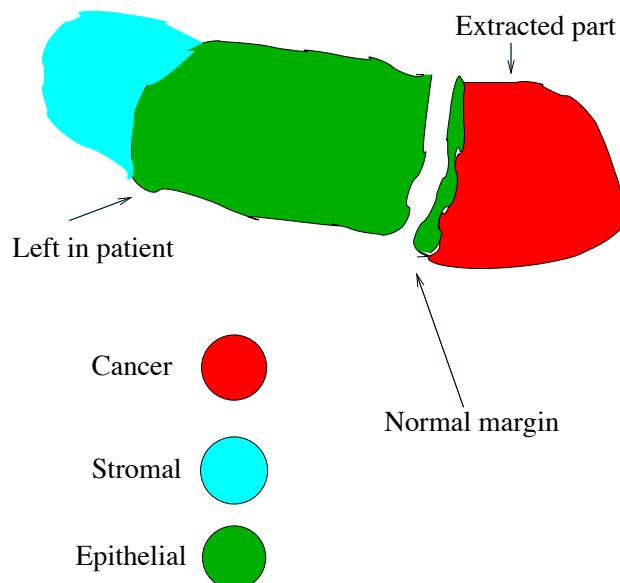


## Gastric (Stomach) Cancer

1. Surgeon removes tissue
2. Pathologist examines tissue
  - Under microscope
3. If no margin, GOTO Step 1.

# Drawbacks

- **Expensive:** requires a pathologist
- **Slow:** examination can take up to an hour
- **Unreliable:** 20%-30% can't predict on the spot



## Gastric (Stomach) Cancer

1. Surgeon removes tissue
2. Pathologist examines tissue
  - Under microscope
3. If no margin, GOTO Step 1.

# Machine Learning to the Rescue!

(actually just statistics)

- Lasso originated from statistics community.
  - **But we machine learners love it!**

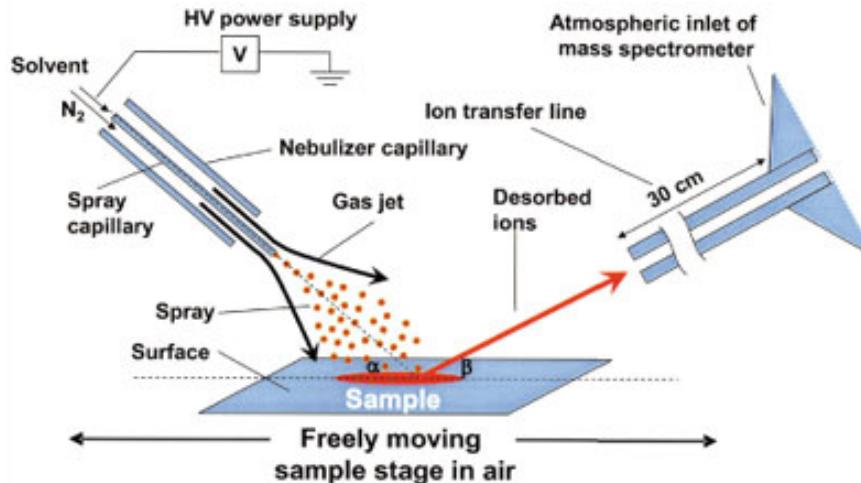
Basic Lasso:

$$\operatorname{argmin}_{w,b} \lambda |w| + \sum_{i=1}^N (y_i, f(x \mid w, b))$$

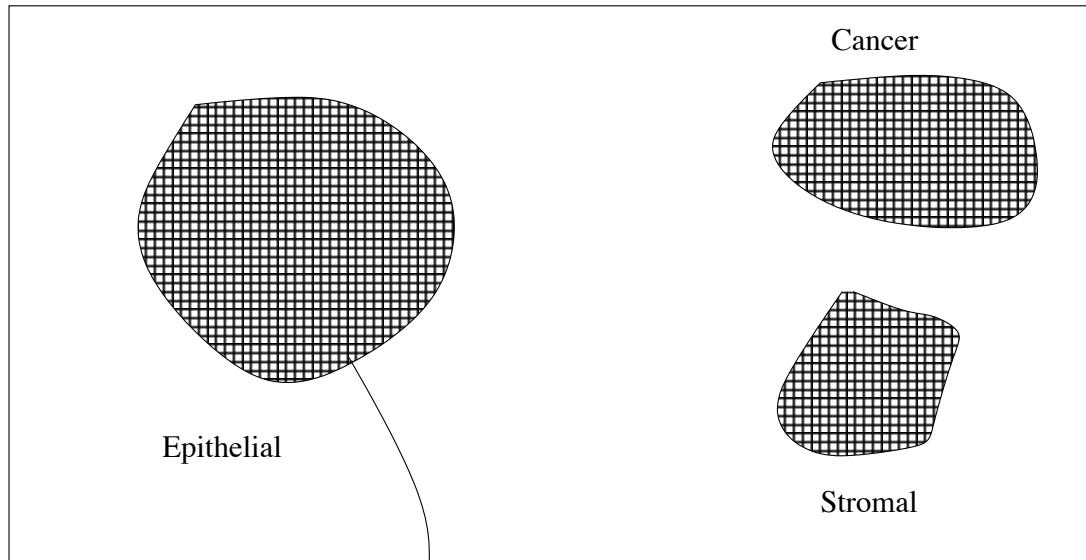
- Train a model to predict cancerous regions!
  - $Y = \{C, E, S\}$
  - What is  $X$ ?
  - What is loss function?

# Mass Spectrometry Imaging

- DESI-MSI (Desorption Electrospray Ionization)

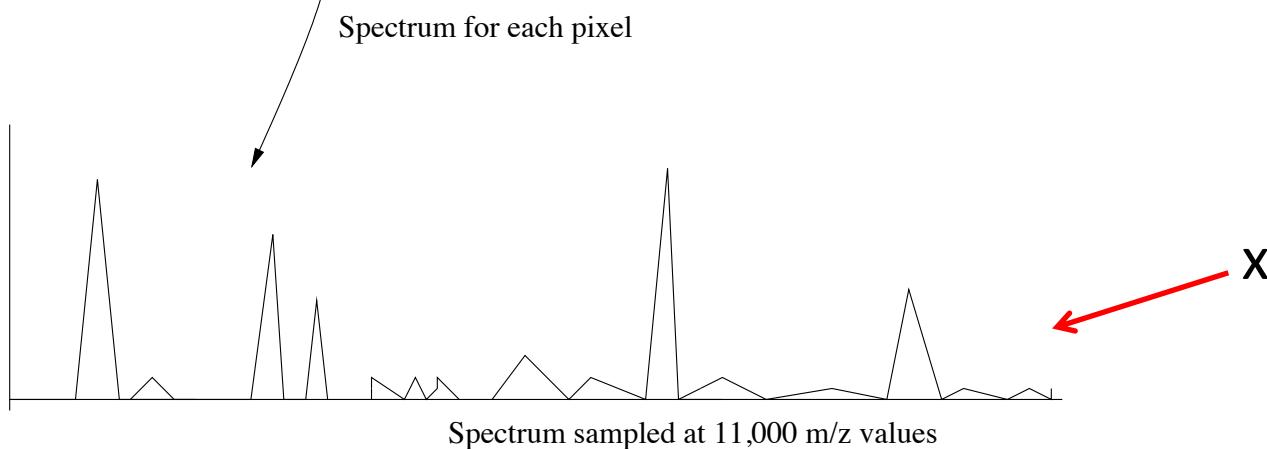


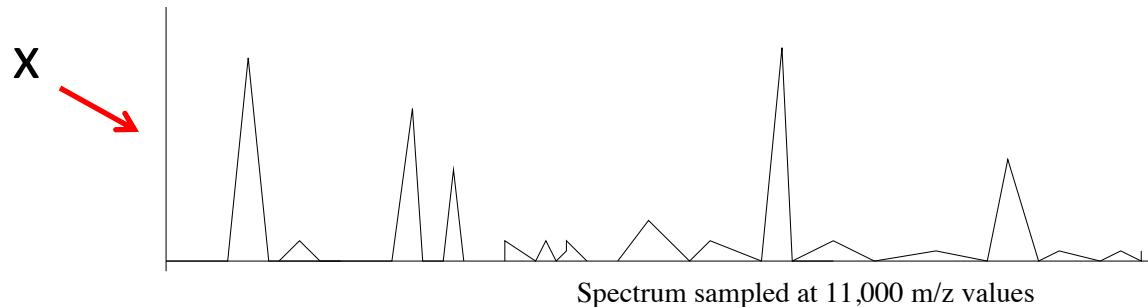
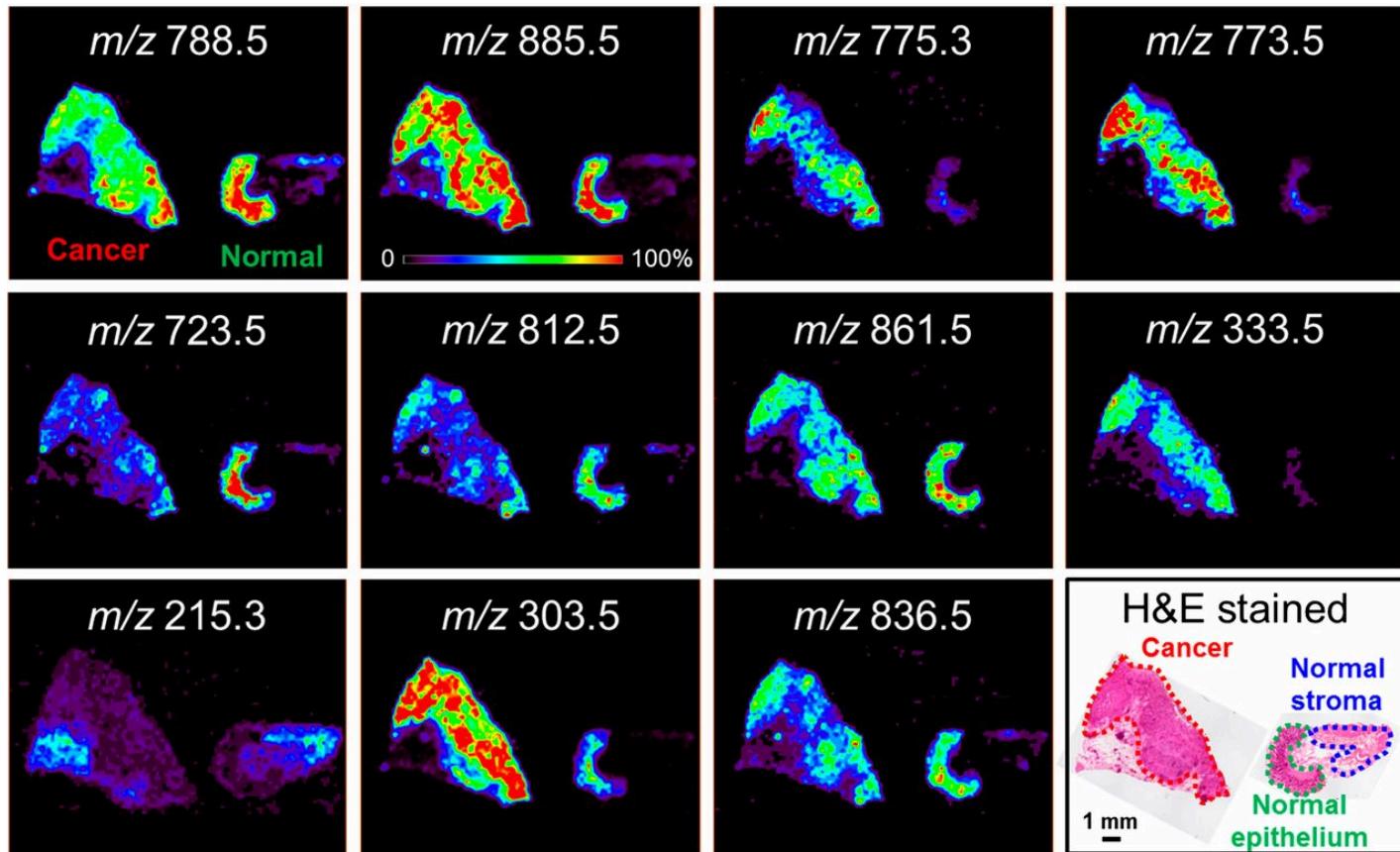
- Effectively runs in real-time (used to generate x)  
[http://en.wikipedia.org/wiki/Desorption\\_electrospray\\_ionization](http://en.wikipedia.org/wiki/Desorption_electrospray_ionization)



Each pixel is data point

x via spectroscopy  
y via cell-type label





Each pixel has  
11K features.  
Visualizing a  
few features.

# Recap: Multiclass Logistic Regression

**Binary LR:**  $P(y|x, w, b) = \frac{e^{y(w^T x - b)}}{e^{y(w^T x - b)} + e^{-y(w^T x - b)}}$   $y \in \{-1, +1\}$

**“Log Linear” Property:**  $P(y|x, w, b) \propto e^{y(w^T x - b)}$   $(w_1, b_1) = (-w_{-1}, -b_{-1})$

**Extension to Multiclass:**  $P(y = k|x, w, b) \propto e^{w_k^T x - b_k}$  Keep a  $(w_k, b_k)$  for each class

**Multiclass LR:**  $P(y = k|x, w, b) = \frac{e^{w_k^T x - b_k}}{\sum_m e^{w_m^T x - b_m}}$

Referred to as Multinomial Log-Likelihood by Tibshirani

<http://statweb.stanford.edu/~tibs/ftp/canc.pdf>

# Lasso Multiclass Logistic Regression

$$\operatorname{argmin}_{w,b} \lambda |w| + \sum_i -\ln P(y_i | x_i, w, b)$$
$$x \in R^D$$
$$y \in \{1, 2, \dots, K\}$$

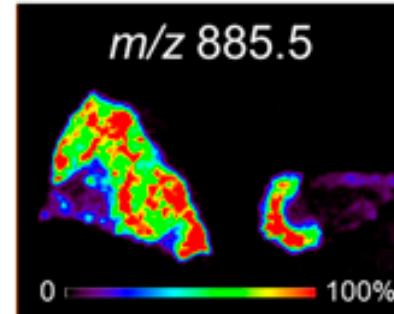
$$|w| = \sum_k |w_k| = \sum_k \sum_d |w_{kd}|$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}$$

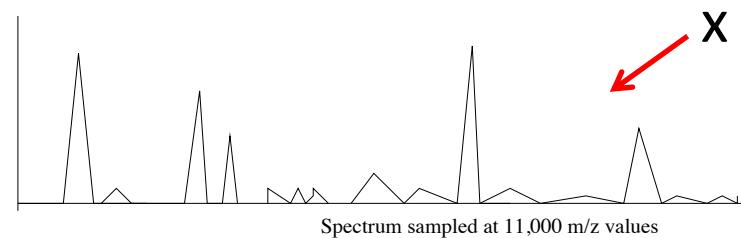
- Probabilistic model
- Sparse weights

# Back to the Problem

- Image Tissue Samples
- Each pixel is an x
  - 11K features via Mass Spec
  - Computable in real time
  - 1 prediction per pixel
- y via lab results
  - ~2 weeks turn-around



Visualization of all pixels for one feature

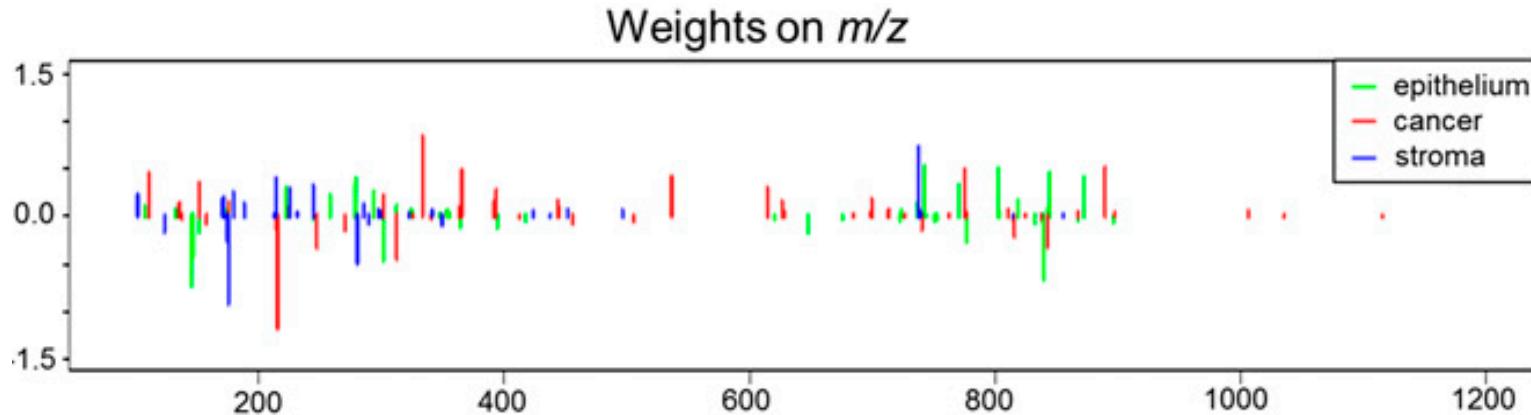


# Learn a Predictive Model

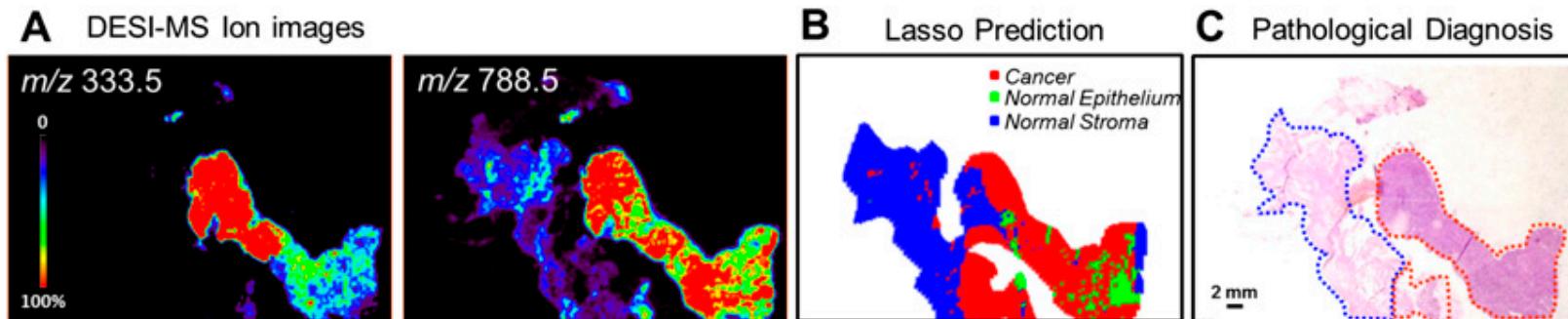
- Training set: 28 tissue samples from 14 patients
  - Cross validation to select  $\lambda$
- Test set: 21 tissue samples from 9 patients
- Test Performance:

Pathology	Predicted				Agreement, %	Overall agreement, %
	Cancer	Epithelium	Stroma	Don't know		
Cancer	5,809	114	2	230	97.0	97.2
Epithelium	134	3,566	118	122	96.8	
Stroma	25	82	2,630	143	96.1	
Cancer		Normal			Agreement, %	Overall agreement, %
Cancer	5,809	116		230	97.0	98.4
Normal	159	6,396		265	99.7	

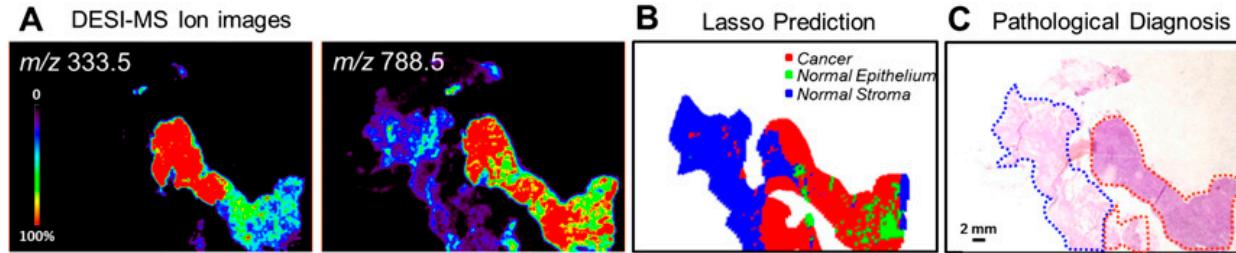
$\geq 0.2$  margin  
in probability



- **Lasso yields sparse weights! (Manual Inspection Feasible!)**
- Many correlated features
  - Lasso tends to focus on one



# Recap: Cancer Detection



- Seems Awesome! What's the catch?
  - Small sample size
    - Tested on 9 patients
  - Machine Learning only part of the solution
    - Need infrastructure investment, etc.
    - Analyze the scientific legitimacy
  - Social/Political/Legal
    - If there is mis-prediction, who is at fault?

# Personalization via twitter

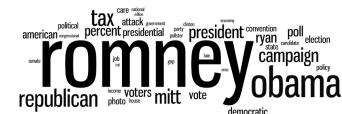
## music



# soccer



# Biden



# Labour



# “Representing Documents Through Their Readers”

## Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (2013)

Khalid El-Arini, Min Xu, Emily Fox, Carlos Guestrin

<https://dl.acm.org/citation.cfm?id=2487596>

The Washington Post

THE HUFFINGTON POST  
THE INTERNET NEWSPAPER: NEWS BLOGS VIDEO COMMUNITY

The New York Times

the guardian



ft.com/frontpage UK All times are London time  
FINANCIAL TIMES



THE DAILY BEAST



Slate

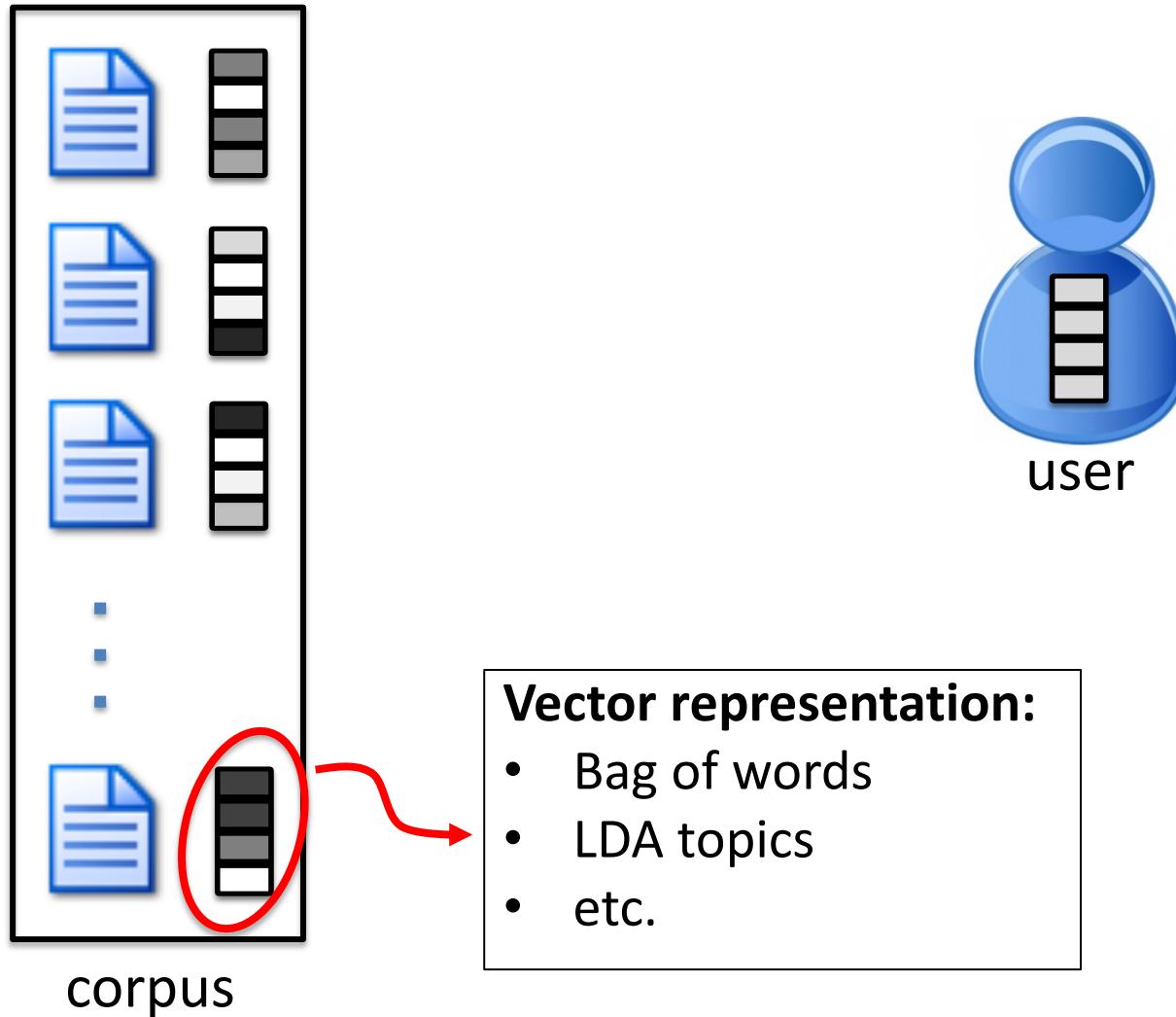


## overloaded by news

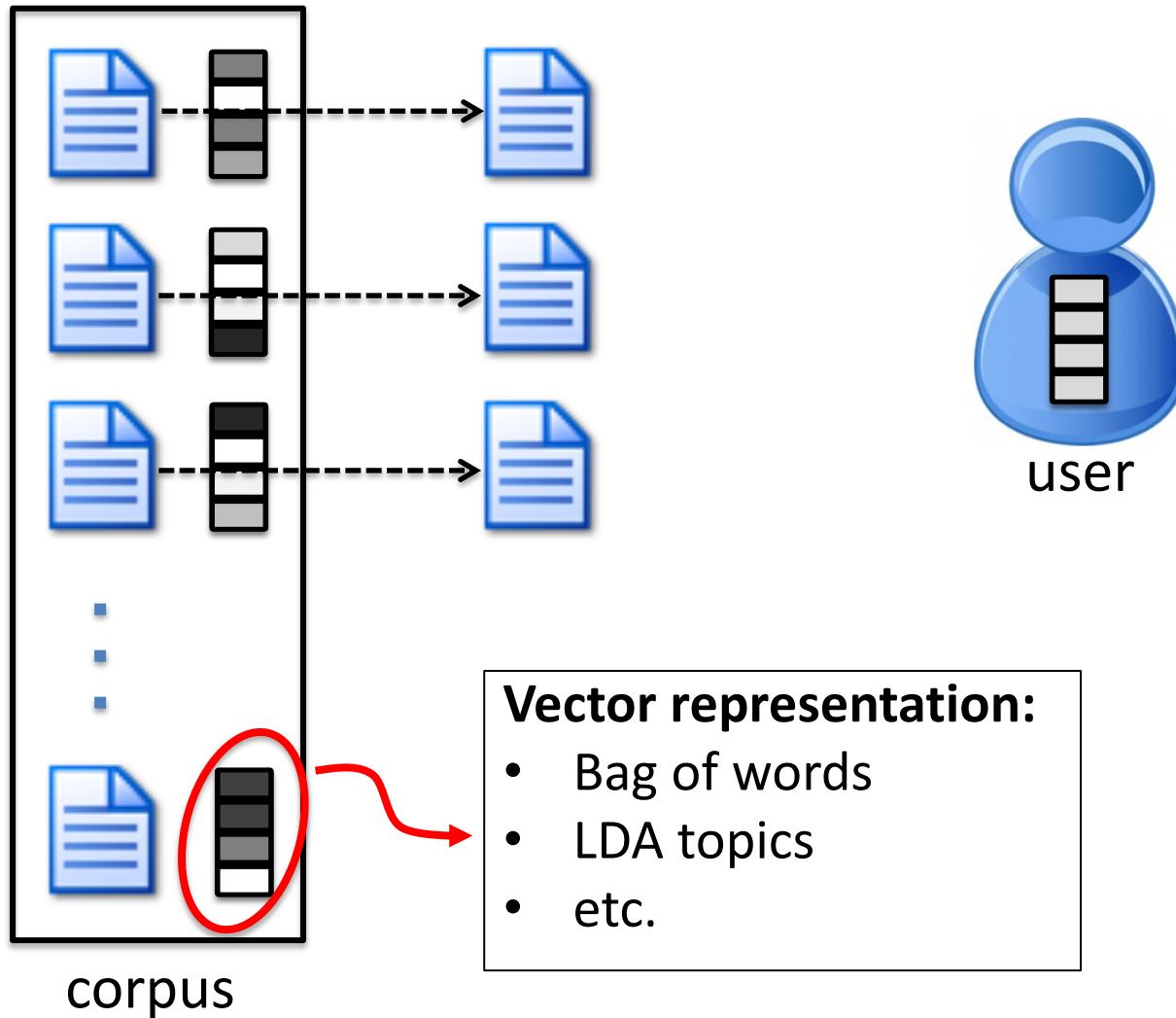
≥ 1 million news articles & blog posts generated every hour\*

\* [www.spinn3r.com](http://www.spinn3r.com)

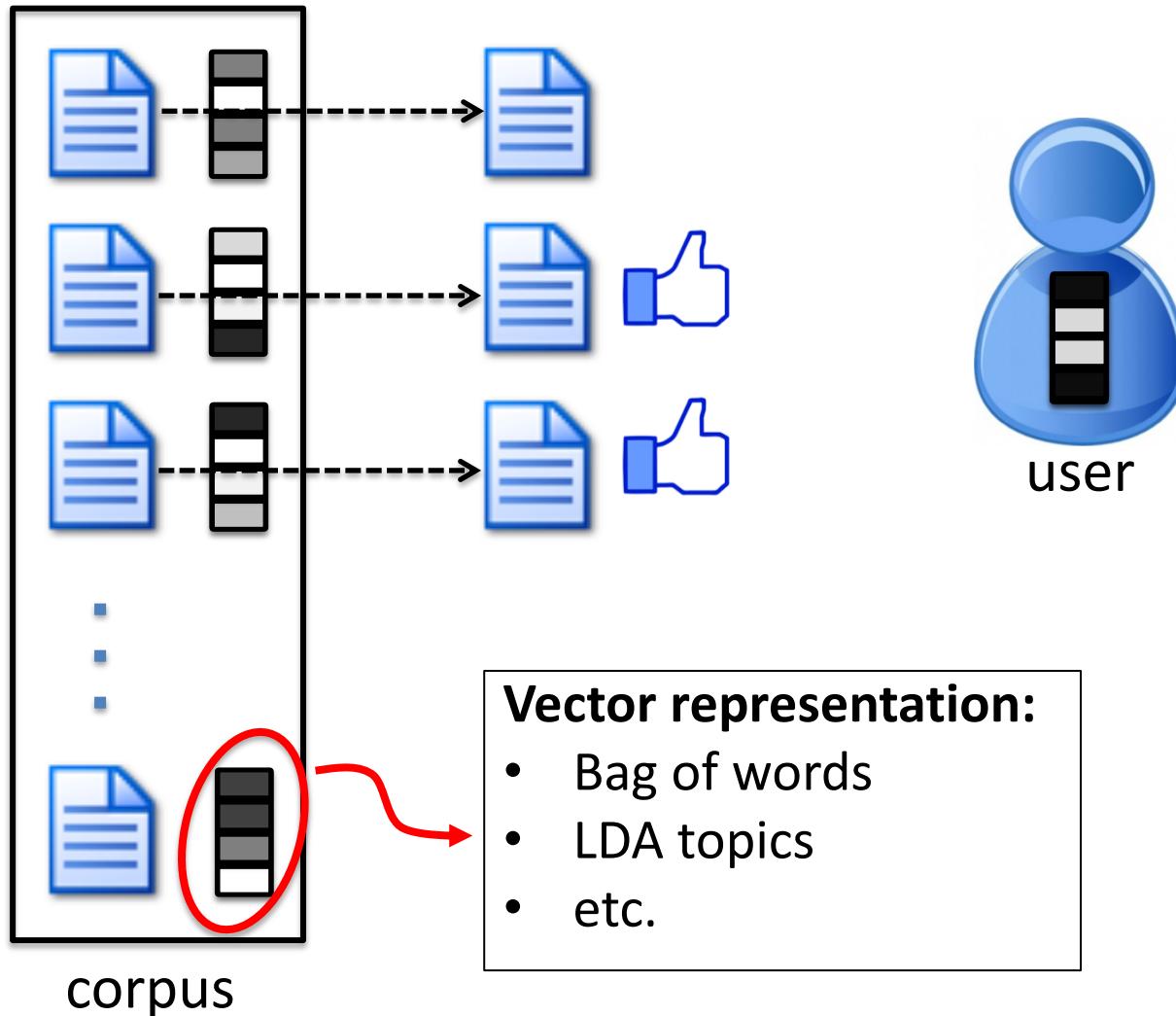
# News Recommendation Engine



# News Recommendation Engine



# News Recommendation Engine



# Challenge

Most common representations don't naturally line up with user interests



Fine-grained representations (bag of words) **too specific**

Haqqani network is considered most ruthless branch of Afghan insurgency  
Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion



High-level topics (e.g., from a topic model)

- too fuzzy and/or vague
- can be inconsistent over time

# Goal

Improve recommendation  
performance through a  
more natural document  
representation

# An Opportunity: News is Now Social

- In 2012, Guardian announced more readers visit site via Facebook than via Google search

## Other Agencies Clamor for Data N.S.A. Compiles

By ERIC LICHTBLAU and MICHAEL S. SCHWARTZ

Published: August 3, 2013

238 Comments

WASHINGTON — The [National Security Agency's](#) dominant role as the nation's spy warehouse has spurred frequent tensions and turf fights with other federal intelligence agencies that want to use its surveillance tools for their own investigations, officials say.

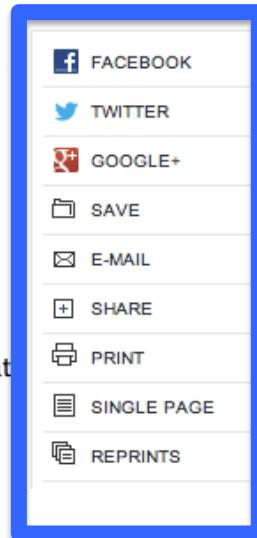
Connect With Us on Twitter



Follow  
@NYTNational for breaking news and headlines.

Twitter List: Reporters and Editors

Agencies working to curb drug trafficking, cyberattacks, money laundering, counterfeiting and even copyright infringement complain that their attempts to exploit the security agency's vast resources have often been turned down because their own



Log in to see what your friends are sharing on [Log In With Facebook](#)  
[nytimes.com](#). [Privacy Policy](#) | [What's This?](#)

**What's Popular Now** [f](#)

Cory Booker for Senator  Michael Ansara, Actor Who Played Cochise and Kang, Dies at 91 



# Substandard Nerd

@substandardnerd

*Gig Going, Festival Attending, Music Loving, Linux Fettling, Perl Hacking, Cycling, Vegan*

The Gdansk of Oxfordshire .



<https://www.youtube.com/user/apusskidu/featured>



**Substandard Nerd** @substandardnerd

13 Jan

Stevie Nicks: the return of Fleetwood Mac

[guardian.co.uk/music/2013/jan...](http://guardian.co.uk/music/2013/jan...)

View summary

# Approach

Learn a document representation based on how readers publicly describe themselves

# Substandard Nerd

@substandardnerd

*Gig Going, Festival Attending, Music Loving, Linux Fettling, Perl Hacking, Cycling, Vegan*

The Gdansk of Oxfordshire .

<https://www.youtube.com/user/apusskidu/featured>



**Substandard Nerd** @substandardnerd

13 Jan

Stevie Nicks: the return of Fleetwood Mac

[guardian.co.uk](https://www.theguardian.com/music/2013/jan/13/stevie-nicks-fleetwood-mac-interview)

[View summary](#)

Culture > Music > Stevie Nicks

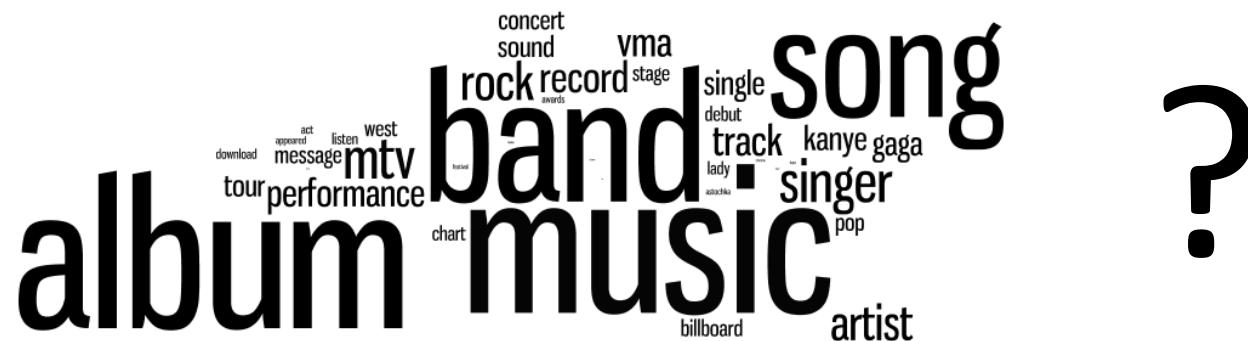
## Stevie Nicks: the return of Fleetwood Mac

Stevie Nicks's tumultuous life as a rock queen led her to addiction, heartbreak and "insanity". As Fleetwood Mac reunite, she tells Caspar Llewellyn Smith why she's going back for more

Using many tweets, can we learn  
that someone who identifies with

via profile badges → **music**

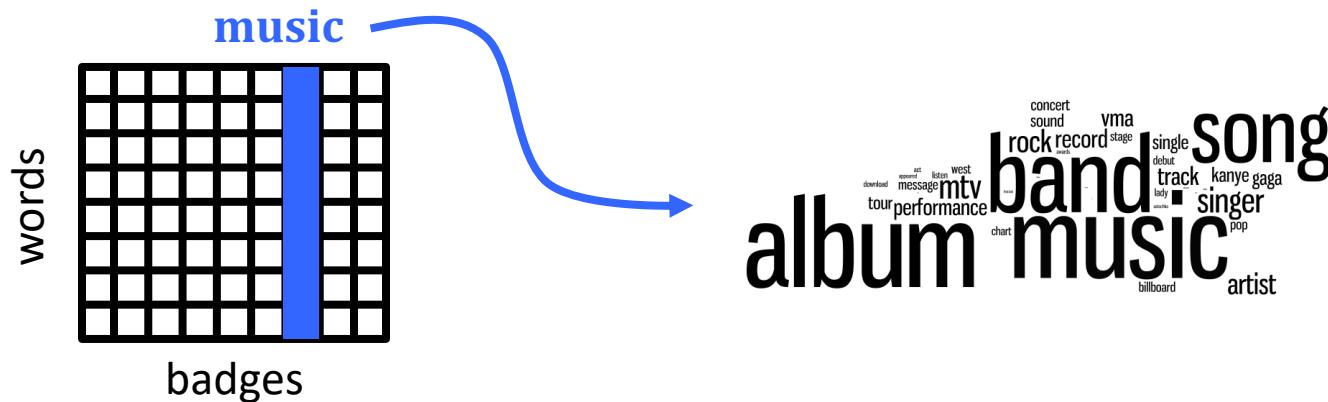
reads articles with these words:



**Given:** training set of tweeted news articles from a specific period of time

3 million articles

## 1. Learn a **badge dictionary** from training set



## 2. Use badge dictionary to encode new articles

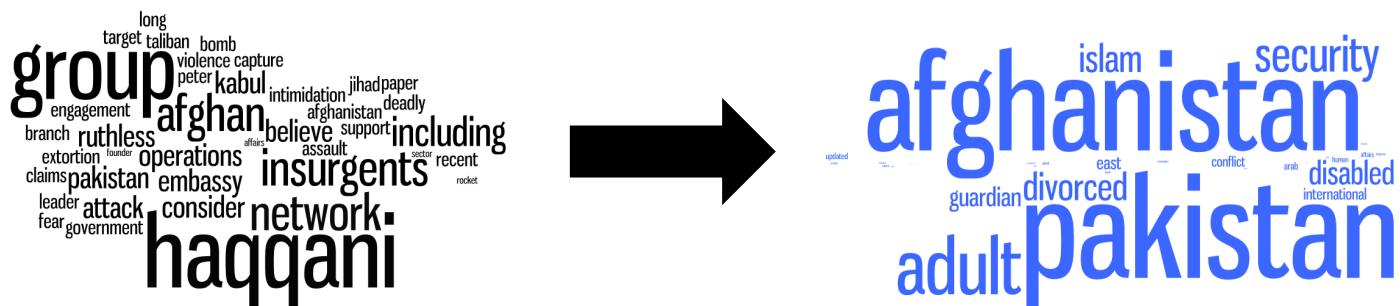
Haqqani network is considered most ruthless branch of Afghan insurgency  
Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion



islam security  
afghanistan east conflict  
guardian divorced arab disabled  
adult pakistan international

# Advantages

- Interpretable
  - Clear labels
  - Correspond to user interests
- Higher-level than words



# Advantages

- Interpretable
  - Clear labels
  - Correspond to user interests

- Higher

Haqqani network is considered most ruthless branch of Afghan insurgency

Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion

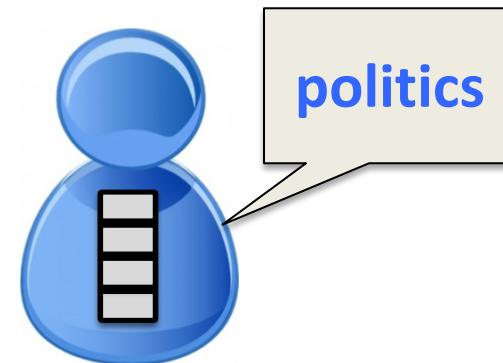
group  
afghan  
pakistan  
haqqani



afghanistan  
pakistan

# Advantages

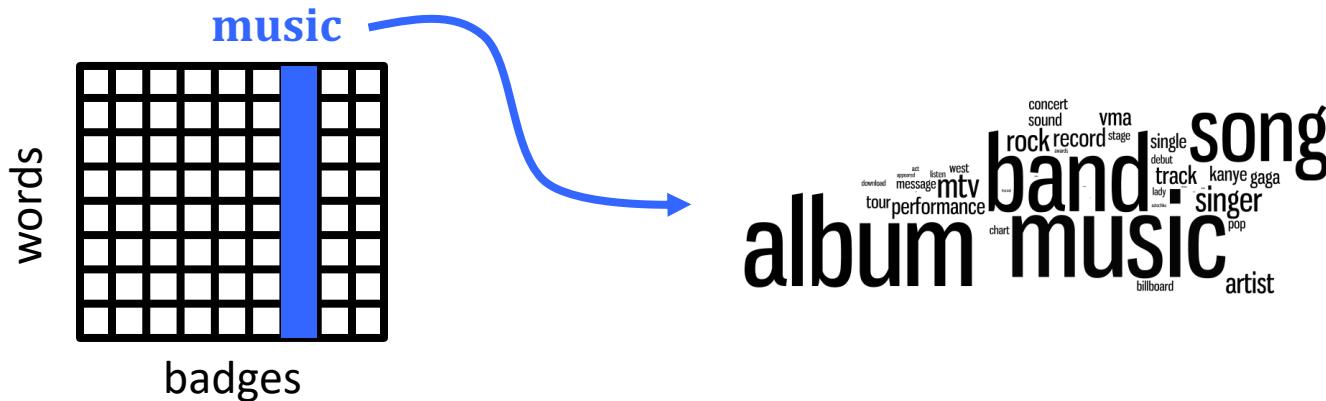
- Interpretable
  - Clear labels
  - Correspond to user interests
- Higher-level than words
- Semantically consistent over time



**Given:** training set of tweeted news articles from a specific period of time

3 million articles

## 1. Learn a **badge dictionary** from training set



## 2. Use badge dictionary to encode new articles

Haqqani network is considered most ruthless branch of Afghan insurgency  
Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion



islam security  
afghanistan east conflict  
guardian divorced arab disabled  
adult pakistan international

# Dictionary Learning

- Training data :

Identifies badges  
in Twitter profile  
of tweeter

$$S = \left\{ (z_i, y_i) \right\}_{i=1}^N$$

Bag-of-words  
representation of  
document

Culture > Music > Stevie Nicks

## Stevie Nicks: the return of Fleetwood Mac

Stevie Nicks's tumultuous life as a rock queen led her to addiction, heartbreak and "insanity". As Fleetwood Mac reunite, she tells Caspar Llewellyn Smith why she's going back for more

## Substandard Nerd

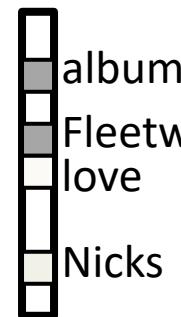
@substandardnerd

Gig Going, Festival Attending, Music Loving, Linux Fettling, Perl Hacking, Cycling, Vegan

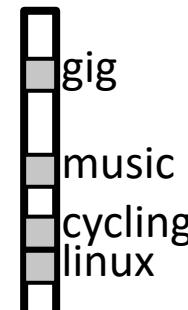
The Gdansk of Oxfordshire

<https://www.youtube.com/user/apusskidu/featured>

$y$



$z$



# Dictionary Learning

$$S = \{(z_i, y_i)\}_{i=1}^N$$

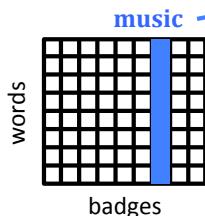
Identifies badges in Twitter profile of tweeter

Bag-of-words representation of document

- Training Objective:

$$\operatorname{argmin}_{B,W} \lambda_B |B| + \lambda_W |W| + \sum_{i=1}^N \|y_i - BW_i\|^2$$

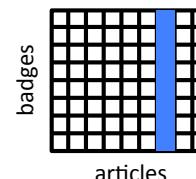
“Dictionary”



music  
band  
album  
song  
concert  
sound  
rock record  
vma  
single  
tour performance  
mtv  
kanye gaga  
track  
singer  
pop  
artist

Haqqani network is considered most ruthless branch of Afghan insurgency  
Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion

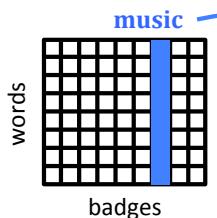
“Encoding”



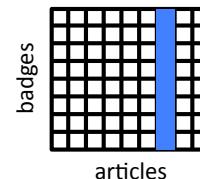
afghanistan  
islam  
adult  
pakistan  
guardian  
divorced  
sex  
cancer  
disabled  
prostitution

$$\operatorname{argmin}_{B,W} \lambda_B |B| + \lambda_W |W| + \sum_{i=1}^N \|y_i - BW_i\|^2$$

“Dictionary”



“Encoding”



Haqqani network is considered most ruthless branch of Afghan insurgency  
Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion

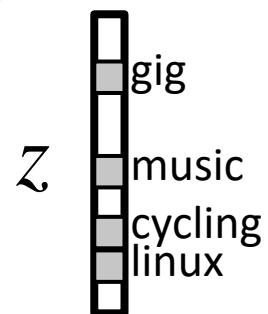


- Not convex! (because of BW term)
- Convex if only optimize B or W (but not both)
- Alternating Optimization (between B and W)
- How to initialize?

Use:  $S = \{(z_i, y_i)\}_{i=1}^N$

Initialize:

$$W_i = \frac{z_i}{|z_i|}$$



$$\underset{B,W}{\operatorname{argmin}} \lambda_B |B| + \lambda_W |W| + \sum_{i=1}^N \|y_i - BW_i\|^2$$

- Suppose Badge s often co-occurs with Badge t
  - $B_s$  &  $B_t$  are correlated
- From perspective of W, B's are features.
  - Lasso tends to focus on one correlated feature
- Graph Guided Fused Lasso:

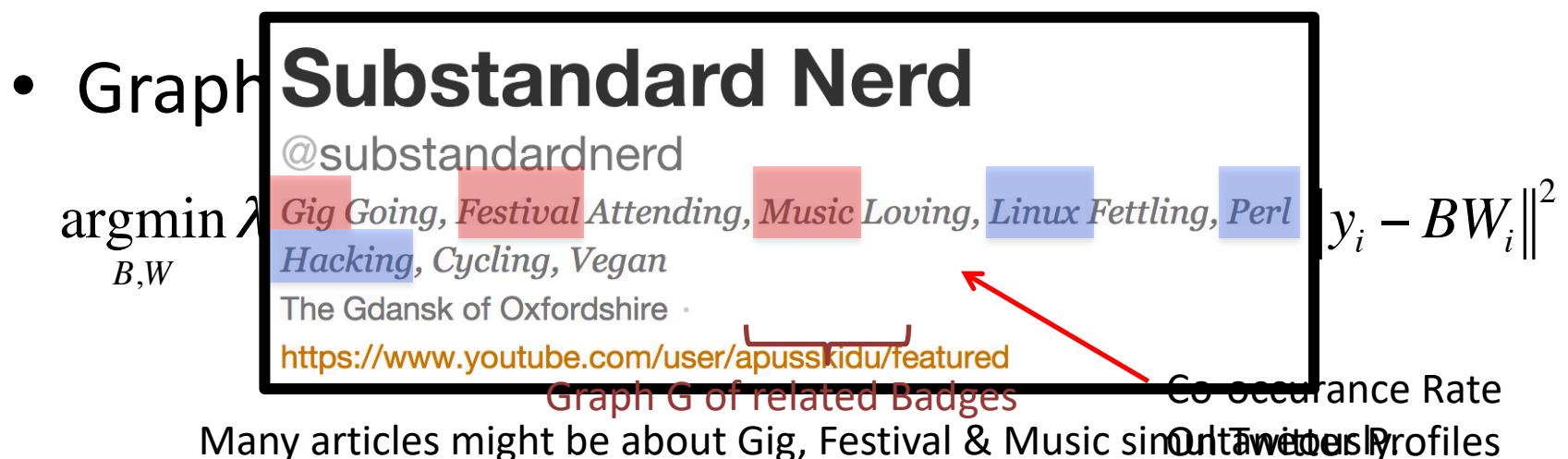
$$\underset{B,W}{\operatorname{argmin}} \lambda_B |B| + \lambda_W |W| + \lambda_G \sum_{i=1}^N \sum_{(s,t) \in E(G)} \omega_{st} |W_{is} - W_{it}| + \sum_{i=1}^N \|y_i - BW_i\|^2$$

↗  
[ Graph G of related Badges [span style="color: red; font-size: 1.5em; margin-left: 20px;"]

Co-occurrence Rate  
 On Twitter Profiles

$$\underset{B,W}{\operatorname{argmin}} \lambda_B |B| + \lambda_W |W| + \sum_{i=1}^N \|y_i - BW_i\|^2$$

- Suppose Badge s often co-occurs with Badge t
  - $B_s$  &  $B_t$  are correlated
- From perspective of W, B's are features.
  - Lasso tends to focus on one correlated feature



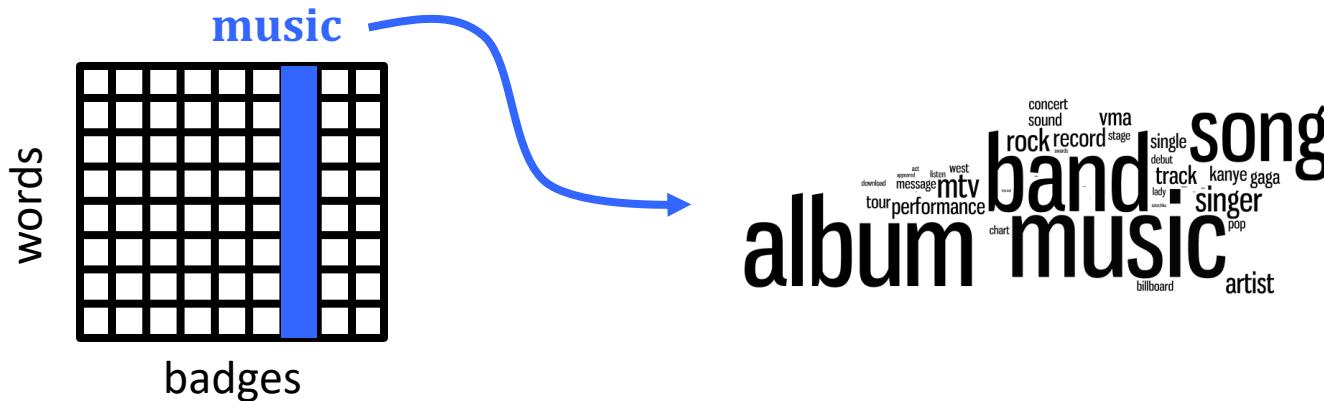
# Encoding New Articles

- Badge Dictionary  $B$  is already learned
- Given a new document  $j$  with word vector  $y_j$ 
  - Learn Badge Encoding  $W_j$ :

$$\operatorname{argmin}_{W_j} \lambda_W |W_j| + \lambda_G \sum_{(s,t) \in G} |W_{js} - W_{jt}| + \|y_j - BW_j\|^2$$

# Recap: Badge Dictionary Learning

- ## 1. Learn a **badge dictionary** from training set



- ## 2. Use badge dictionary to encode new articles

## **Haqqani network is considered most ruthless branch of Afghan insurgency**

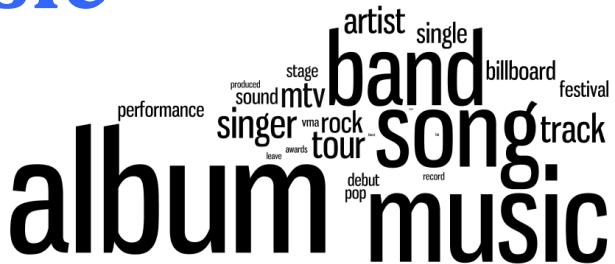
Group that started as part of anti-Soviet jihad has moved into mafia-like violence, intimidation and extortion



afghanistan islam security  
adult guardian divorced east conflict  
pakistan arab home disabled international

# Examining B

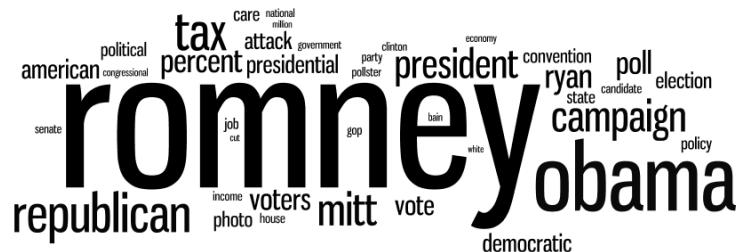
# music



# soccer



# Biden

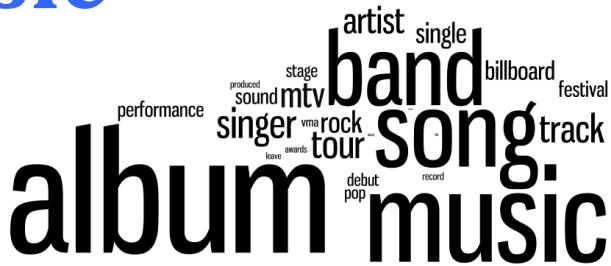


# Labour



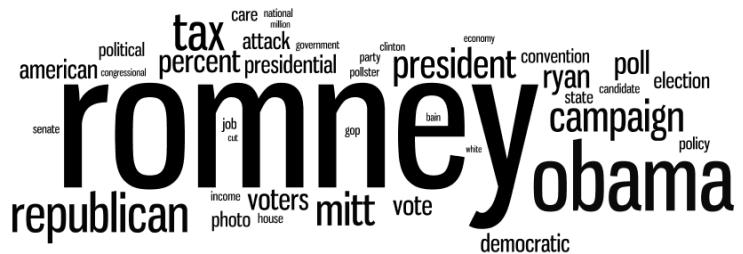
# Badges Over Time

# music



# September 2012

# Biden



# September 2010

concert  
sound  
vma  
stage  
single  
debut  
track  
kanye gaga  
lady  
gaga  
singer  
pop  
billboard  
artist  
chart  
band  
music  
album

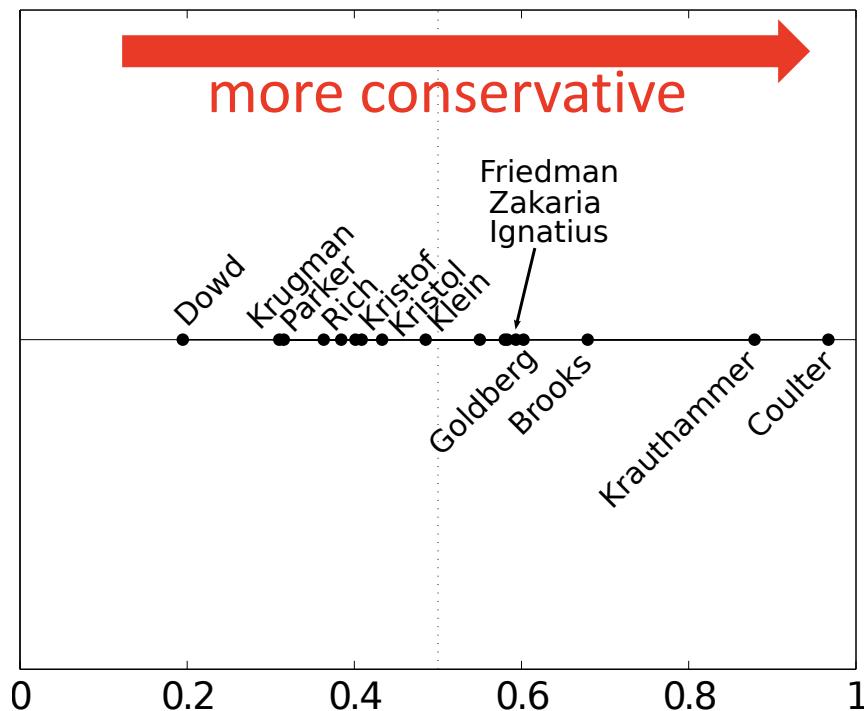
**biden** administration削弱 forgotpasswordform house fundraising thanksgiving whining hostess republican senator president clegg vice screeching president election white gop register joe  
party administration削弱 forgotpasswordform house fundraising thanksgiving whining hostess republican senator president clegg vice screeching president election white gop register joe  
cowboys administration削弱 forgotpasswordform house fundraising thanksgiving whining hostess republican senator president clegg vice screeching president election white gop register joe  
administration削弱 forgotpasswordform house fundraising thanksgiving whining hostess republican senator president clegg vice screeching president election white gop register joe  
party administration削弱 forgotpasswordform house fundraising thanksgiving whining hostess republican senator president clegg vice screeching president election white gop register joe  
dang union est post-season piers voters thanks obama upfront registrationform iran morgan hamas twinkies

**colbert**

# A Spectrum of Pundits

“top conservatives on Twitter”

- Limit badges to **progressive** and **TCOT**
- Predict political alignments of likely readers?

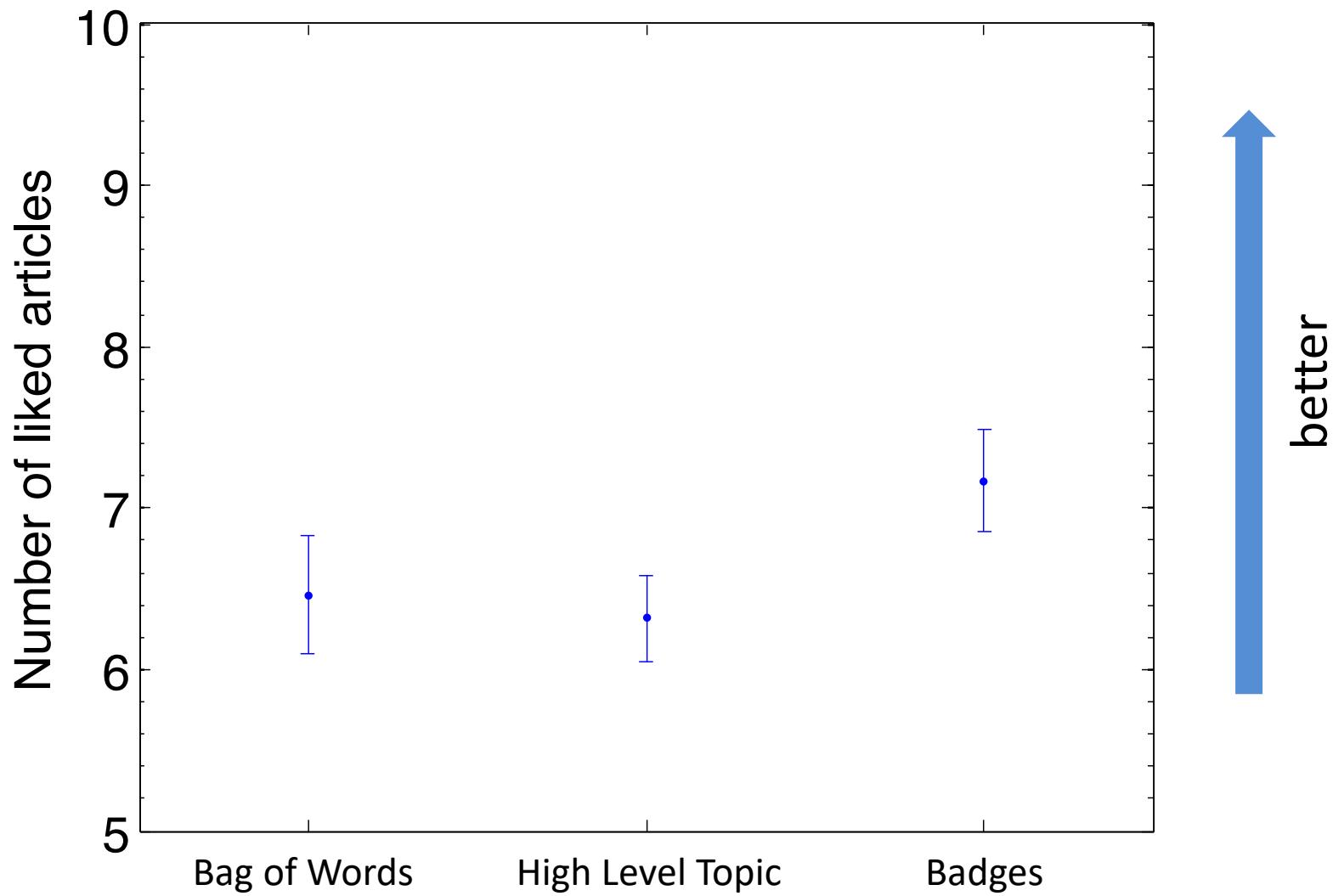


- Took all articles by columnist
- Looked at encoding score
  - progressive vs TCOT
- Average

# User Study

- Which representation best captures user preferences over time?
- Study on Amazon Mechanical Turk with 112 users
  1. Show users random 20 articles from Guardian, from time period 1, and obtain ratings
  2. Pick random representation
    - bag of words, high level topic, Badges
  3. Represent user preferences as mean of liked articles
  4. GOTO next time period
    - Recommend according to preferences
    - GOTO STEP 2

# User Study



# Recap: Personalization via twitter

- Sparse Dictionary Learning
  - Learn a new representation of articles
  - Encode articles using dictionary
  - Better than Bag of Words
  - Better than High Level Topics
- Based on social data
  - Badges on twitter profile & tweeting
  - Semantics not directly evident from text alone

# Learning Visual Style



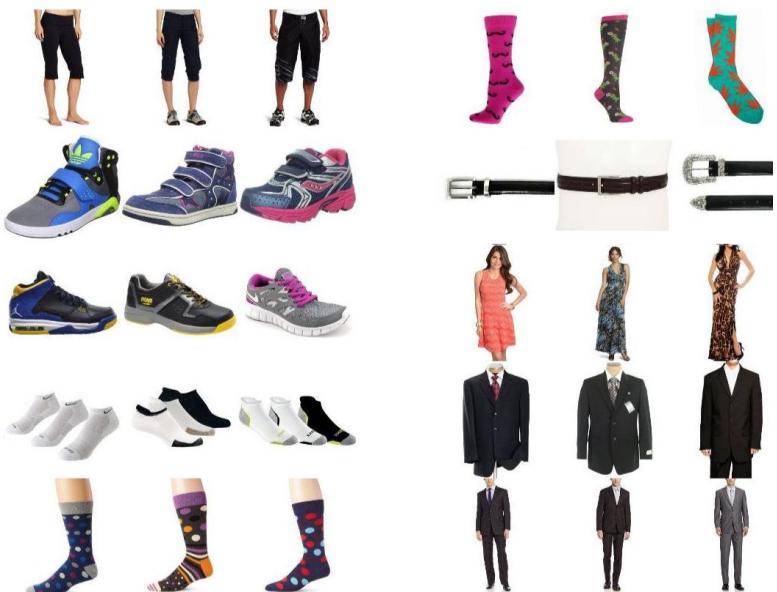
# Learning Visual Clothing Style with Heterogeneous Dyadic Co-occurrences

Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, Serge Belongie, ICCV 2015

## Visually Compatible



## Visually Incompatible



<http://vision.cornell.edu/se3/projects/clothing-style/>

# Training Data

- Ground set of items
  - ~1M items
  - Image of item x
  - Category of item c
    - Coat, belt, pants, socks, etc.
- Pairwise relationships
  - “frequently bought together”
  - Interpret as visually compatible



# Training Goal

(ignoring regularization)

$$\operatorname{argmin}_{\Theta} \sum_{(i,j) \in D} L^+ (\Phi(x_i), \Phi(x_j)) + \sum_{(i,j) \in \tilde{D}} L^- (\Phi(x_i), \Phi(x_j))$$

Penalizes too far

Penalizes too close

Embedding of image

Embedding of image

All Model Parameters

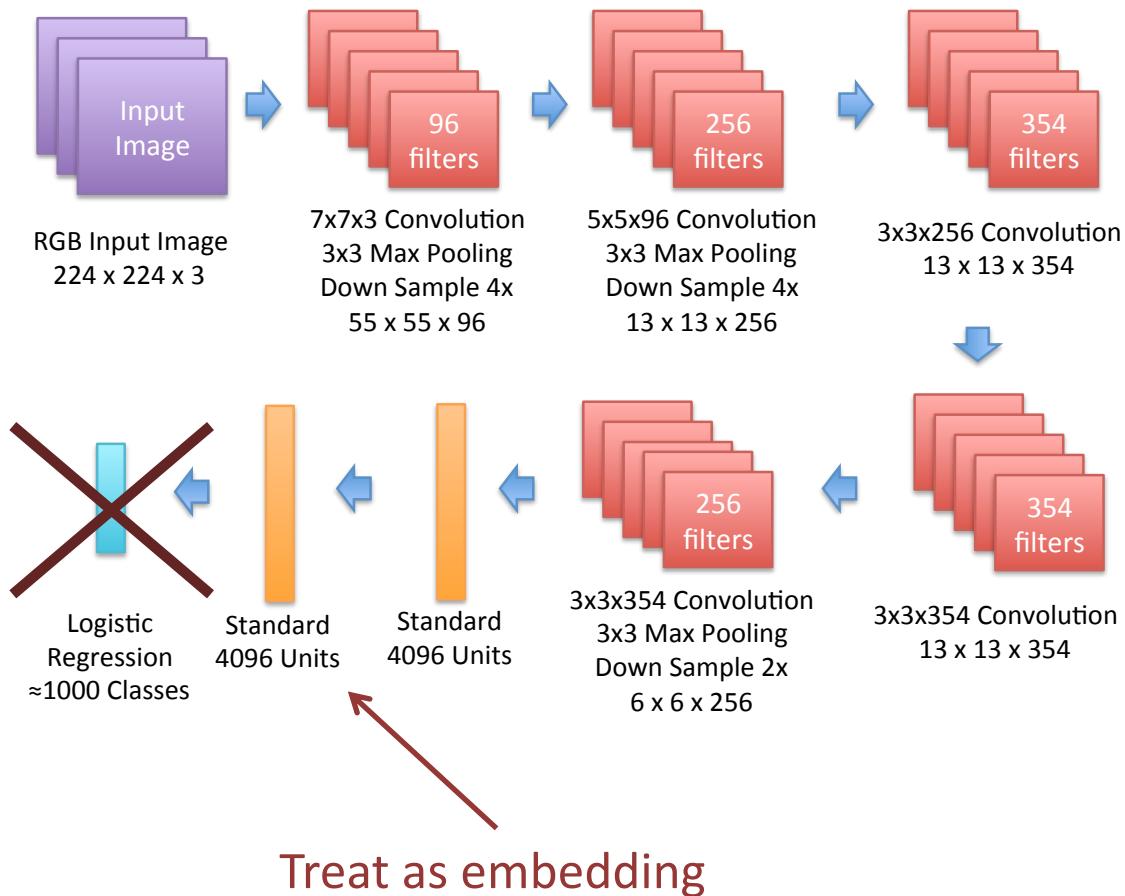
Compatible Pairs

Incompatible Pairs

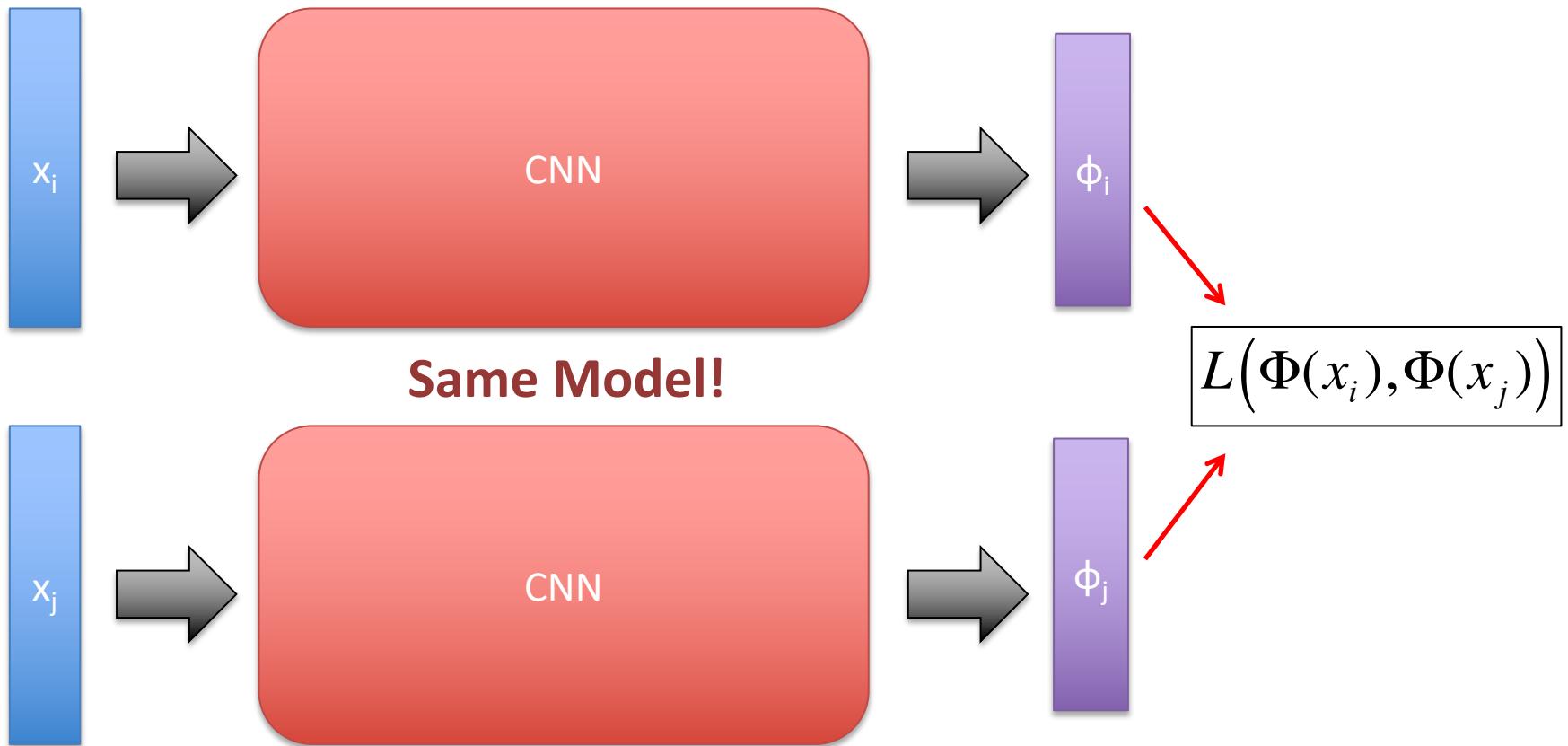
Only pairs in different categories.

The diagram illustrates the components of the training goal equation. At the top, two red arrows point from the text 'Penalizes too far' and 'Penalizes too close' to the terms  $L^+$  and  $L^-$  respectively in the equation. Below these, two red arrows point from the text 'Embedding of image' to the embeddings  $\Phi(x_i)$  and  $\Phi(x_j)$ . A red arrow points from the text 'All Model Parameters' to the  $\Theta$  in the  $\operatorname{argmin}$  term. Another red arrow points from the text 'Compatible Pairs' to the summation term over  $D$ . A red arrow points from the text 'Incompatible Pairs' to the summation term over  $\tilde{D}$ . A blue arrow points from the text 'Only pairs in different categories.' to the label 'Only pairs in different categories.'

# Recall: Convolutional Neural Networks

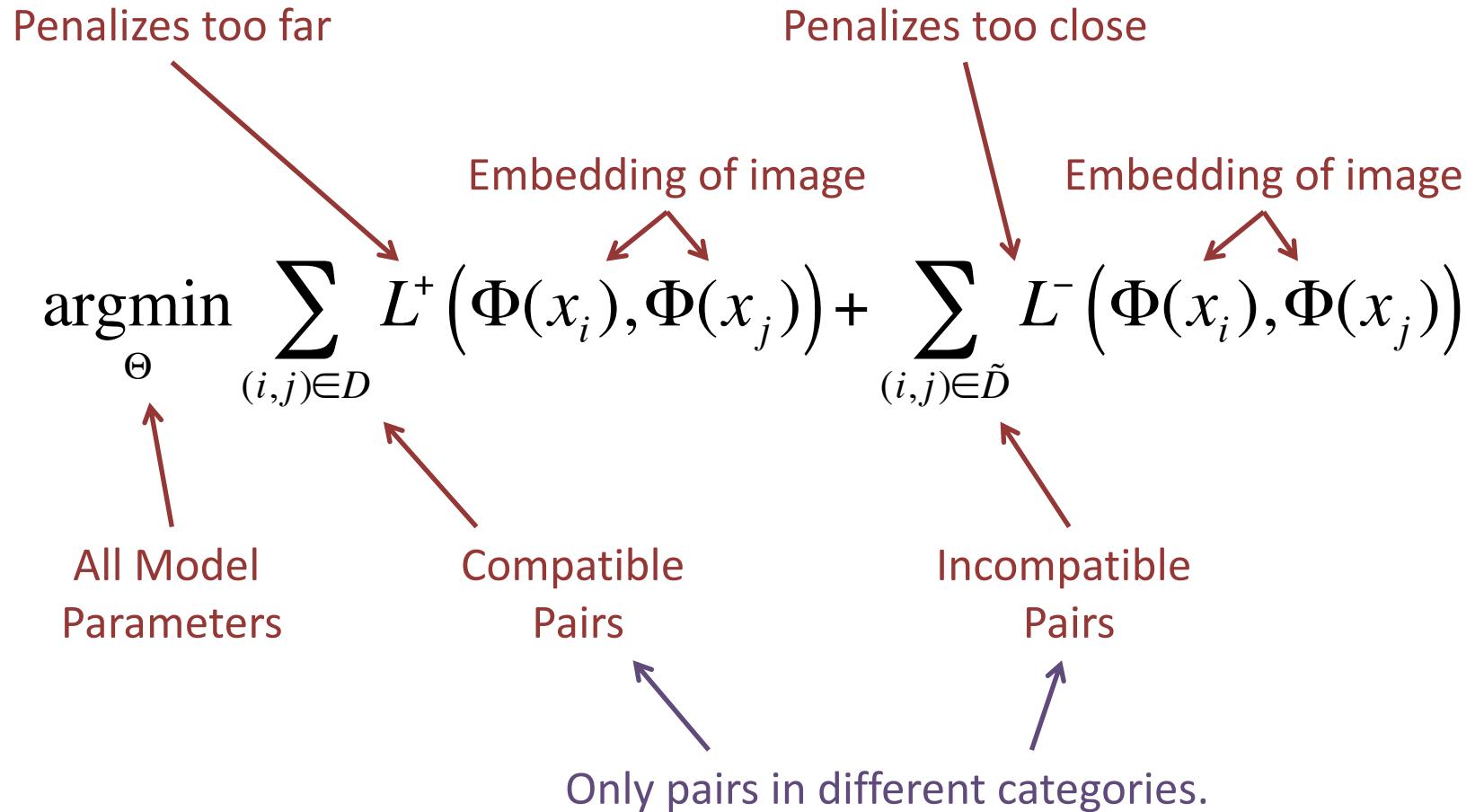


# Siamese Convolutional Neural Networks



More details: <http://www.cs.cornell.edu/~kb/publications/SIG15ProductNet.pdf>

# Recap: Training Goal



**Model Embedding via Siamese Convolutional Neural Network!**

# Training Details

- Want embedding dimension smaller
  - E.g., 128 rather than 4096
- Need to subsample negative pairs
  - Most items are not frequently bought together
  - Negative component can overwhelm objective



<http://www.cs.cornell.edu/~andreas/iccv15.pdf>

# Suggesting Outfits

Upper  
Garment



Footware

# Suggesting Outfits

- Given query item  $i$ 
  - Embedding  $\varphi_i = \Phi(x_i | \Theta)$
  - Category  $c_i$
- For other categories
  - Recommend item with closest embedding  $\varphi$
- **Not robust to label noise!**

<http://www.cs.cornell.edu/~andreas/iccv15.pdf>

# Label Noise

- Amazon category labels are noisy
    - Eg., some pants mis-categorized as shoes
  - Pants are visually very similar

$\Phi( ) \approx \Phi( )$

Pants

Shoes

Mis-categorized!

# Making Robust Suggestions

- Mis-categorizations are rare
  - Instead of predicting closest shoe...
  - Predict closest cluster of shoes!
- Preprocessing: cluster every category
- Given input query (category=pants)
  - Find closest cluster center (category=shoes)
  - Output shoes item close to cluster center

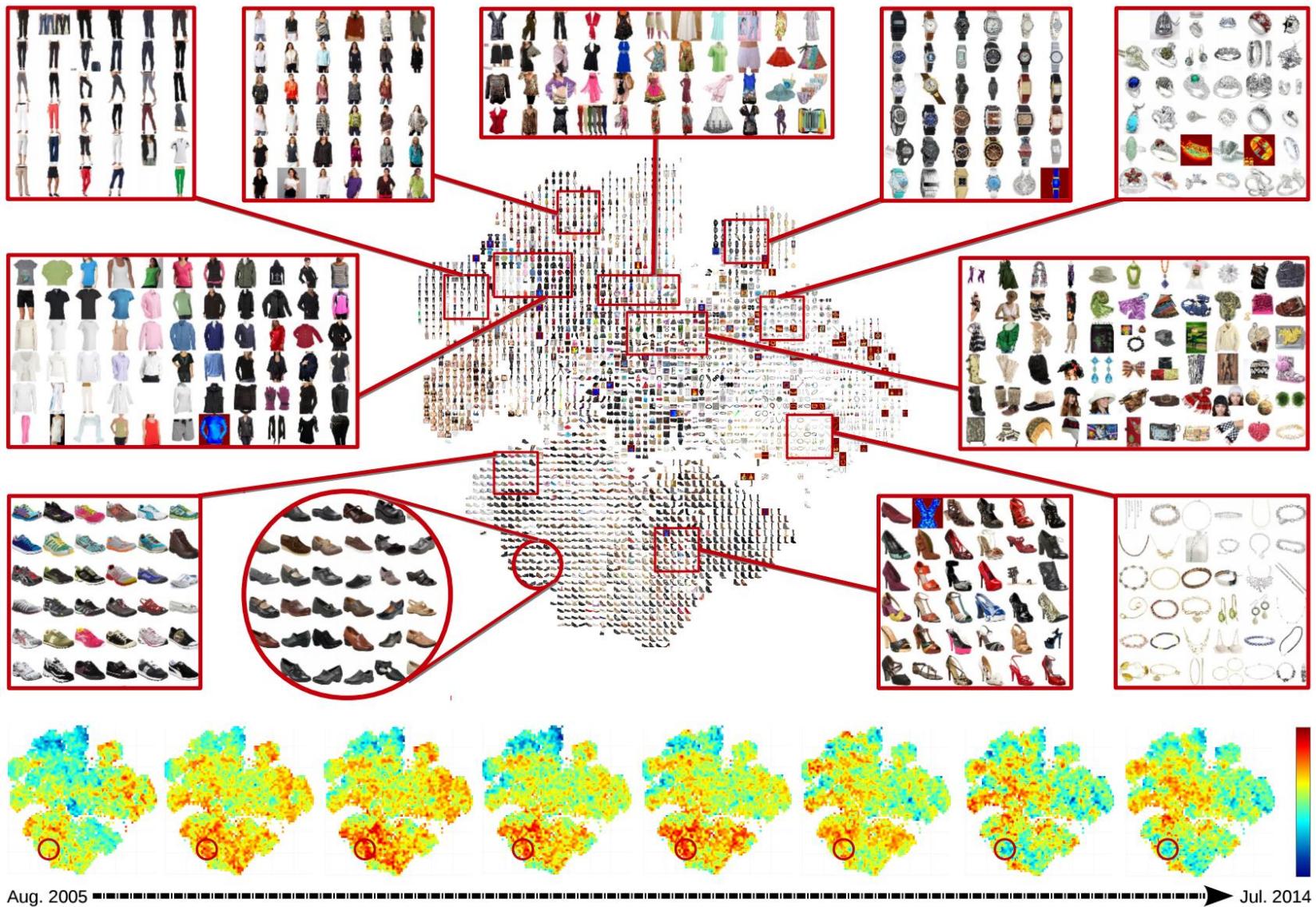
<http://www.cs.cornell.edu/~andreas/iccv15.pdf>

# Compute Coherence of Outfit

Least coordinated



Most coordinated



<http://cseweb.ucsd.edu/~jmcauley/pdfs/www16a.pdf>

# Recap

- Sparsity is often useful
  - Interpretability, data compression
  - Use Lasso/L1 objective
- Representation learning is often useful
  - Lower-dimensional embedding
  - Better suited to semantics of data domain