User-generated content mining: From collective disease rates to individual demographics

Vasileios Lampos

Computer Science @ UCL

@lampos | lampos.net

Language Technology Lab

University of Cambridge

Oct. 27, 2016

Structure of the presentation

1. **Introductory remarks**

2. **Collective disease surveillance from search query data**
   — Google Flu Trends and inference inaccuracies
   — Steps towards improvement

3. **Mining socio-economic demographics from social media users**
   — Occupational class
   — Income
   — Socioeconomic status

4. **Concluding remarks**

Context and Motivation

Context and Motivation

How can we use online user-generated content (UGC) to our benefit?

User-generated content for health. WHY?

+ Online content can potentially access a larger and **more representative** part of the population
  Note: *Health surveillance systems are based on the subset of people who actively seek medical attention*

+ More **timely** information (*almost instant*)

+ Geographical regions with **less established health monitoring systems** could benefit

+ Small **cost** when data access and modelling expertise are in place

Google Flu Trends — The idea

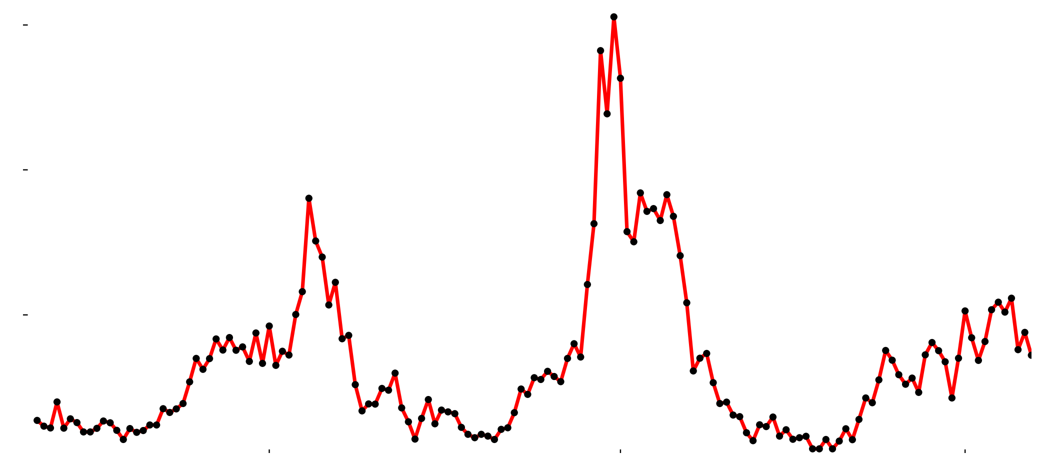
Can we turn **online search query statistics** to estimates about the rate of **influenza-like illness (ILI)** in the real-world population?

Google Flu Trends — Supervised learning

search query frequency time series

Flu rates from a health agency representing doctor consultations

Google



$\mathbf{X} \in \mathbb{R}^{M \text{ x } N}$

$\mathbf{y} \in \mathbb{R}^{M}$

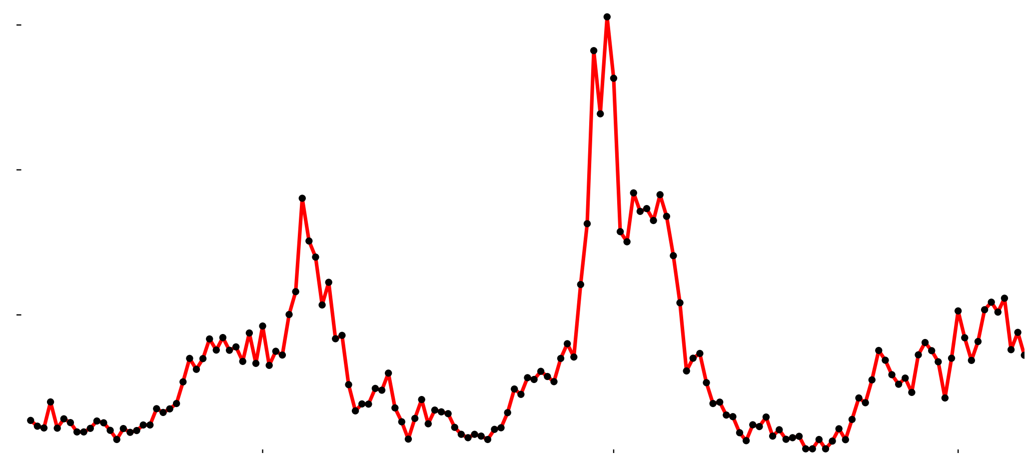
$$\boxed{\texttt{logit}(y) = \beta_0 + \beta_1 \times \texttt{logit}(q) + \varepsilon}$$

(*Ginsberg et al., 2009*)

Google Flu Trends — Supervised learning

search query frequency time series

Flu rates from a health agency representing doctor consultations



q is the aggregate frequency of a selected subset of the N candidate search queries

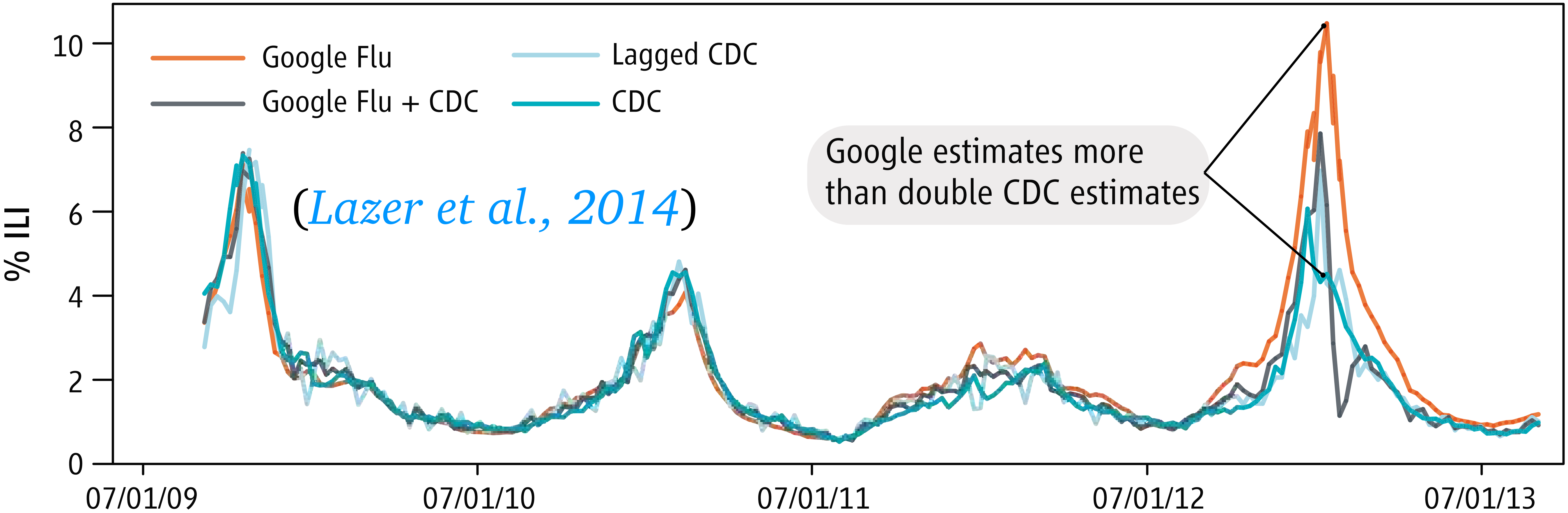
$\mathbf{X} \in \mathbb{R}^{M \times N}$

$\mathbf{y} \in \mathbb{R}^{M}$

$$\boxed{\mathtt{logit}(y) = \beta_0 + \beta_1 \times \mathtt{logit}(q) + \varepsilon}$$

(*Ginsberg et al., 2009*)

Google Flu Trends — Failure



(*Lazer et al., 2014*)

The estimates of the online Google Flu Trends tool were approx. **two times larger** than the ones from the CDC in 2012/13

Google Flu Trends — Hypotheses for failure

- "**Big Data**" criticism
- The **statistical learning model** was not good enough
- **Feature selection** was not good enough bringing in spurious search queries
- **Media hype** about flu significantly affects inference accuracy
- The **ground truth** is not perfect; it is rather a "silver" standard

Google Flu Trends — Hypotheses for failure

- X “**Big Data**” criticism
- ✓ The **statistical learning model** was not good enough
- ✓ **Feature selection** was not good enough bringing in spurious search queries
- ? **Media hype** about flu significantly affects inference accuracy
- ✓? The **ground truth** is not perfect; it is rather a “silver” standard

Advances in nowcasting influenza-like illness rates using online search logs

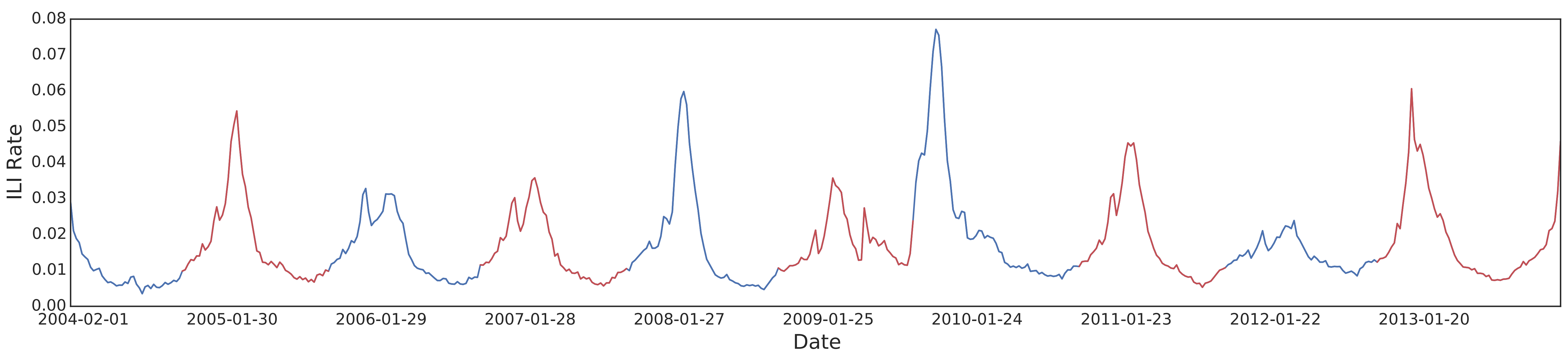
Lampos, Miller, Crossan & Stefansen
(Nature Scientific Reports, 2015)

Data

Google search logs

– weekly search counts of **49,708** search queries
– corresponding total volume of weekly searches
– user search sessions geolocated in the **US**
– anonymised & aggregate data
– Jan. 2004 to Dec. 2013 (521 weeks, ~**decade**)

ILI rates from CDC



Elastic Net for linear regularised regression

query frequency	$\mathbf{x}_i \in \mathbb{R}^m,\ i \in \{1, \ldots, n\}$	—	$\mathbf{X}$
ILI rates	$y_i \in \mathbb{R},\ i \in \{1, \ldots, n\}$	—	$\mathbf{y}$
weights, bias	$w_j, \beta \in \mathbb{R},\ j \in \{1, \ldots, m\}$	—	$\mathbf{w}_* = [\mathbf{w}; \beta]$

$$\underset{\mathbf{w},\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta - \sum_{j=1}^{m} x_{ij} w_j \right)^2 + \underbrace{\lambda_1 \sum_{j=1}^{m} |w_j|}_{\textbf{L1-norm}} + \underbrace{\lambda_2 \sum_{j=1}^{m} w_j^2}_{\textbf{L2-norm}} \right\}$$

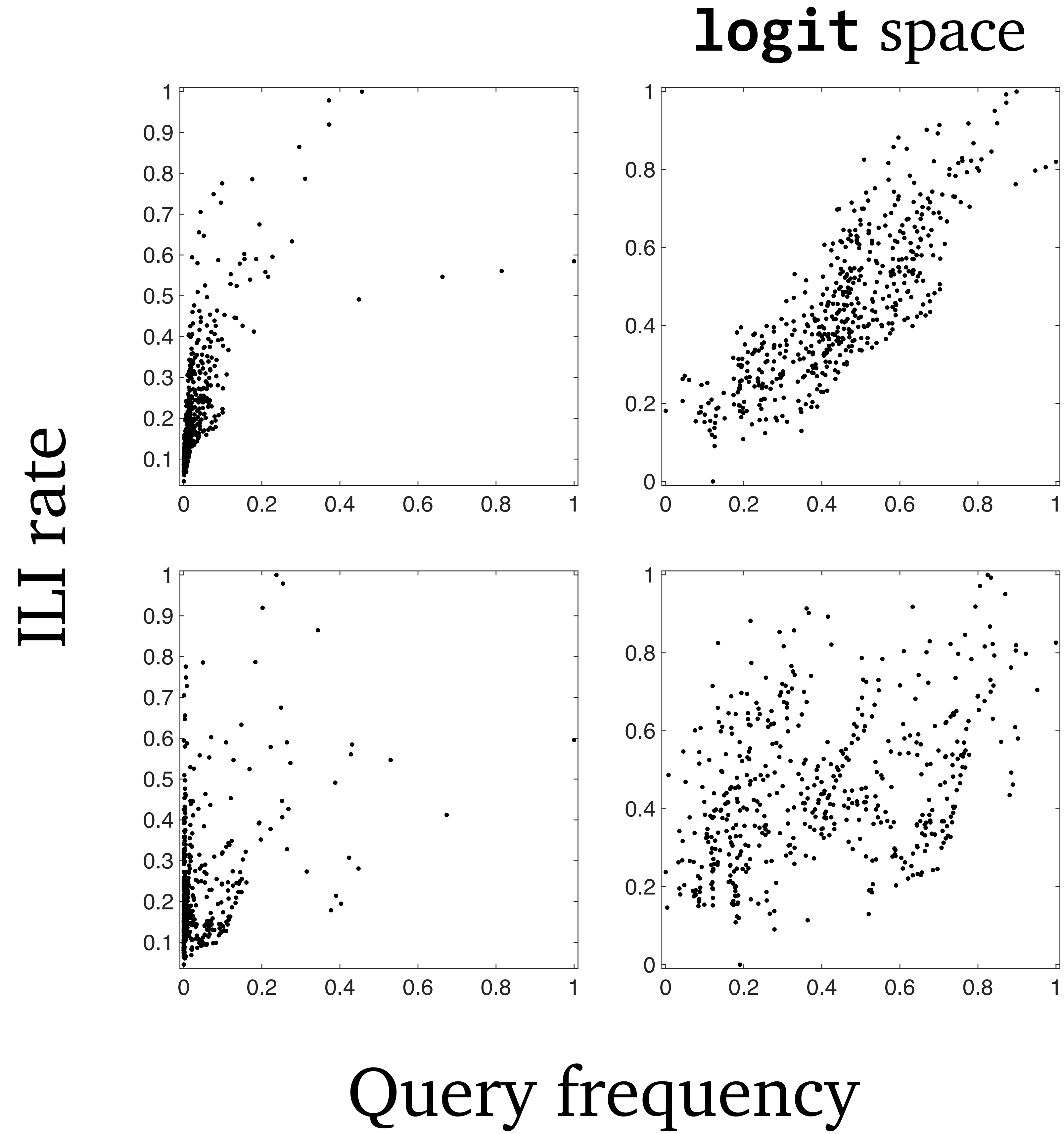
a sparse set of weights (**w**) is encouraged

(*Zou & Hastie, 2005*)

Nonlinearities in the data (1)

"flu symptoms in children"

"flu symptoms in adults"



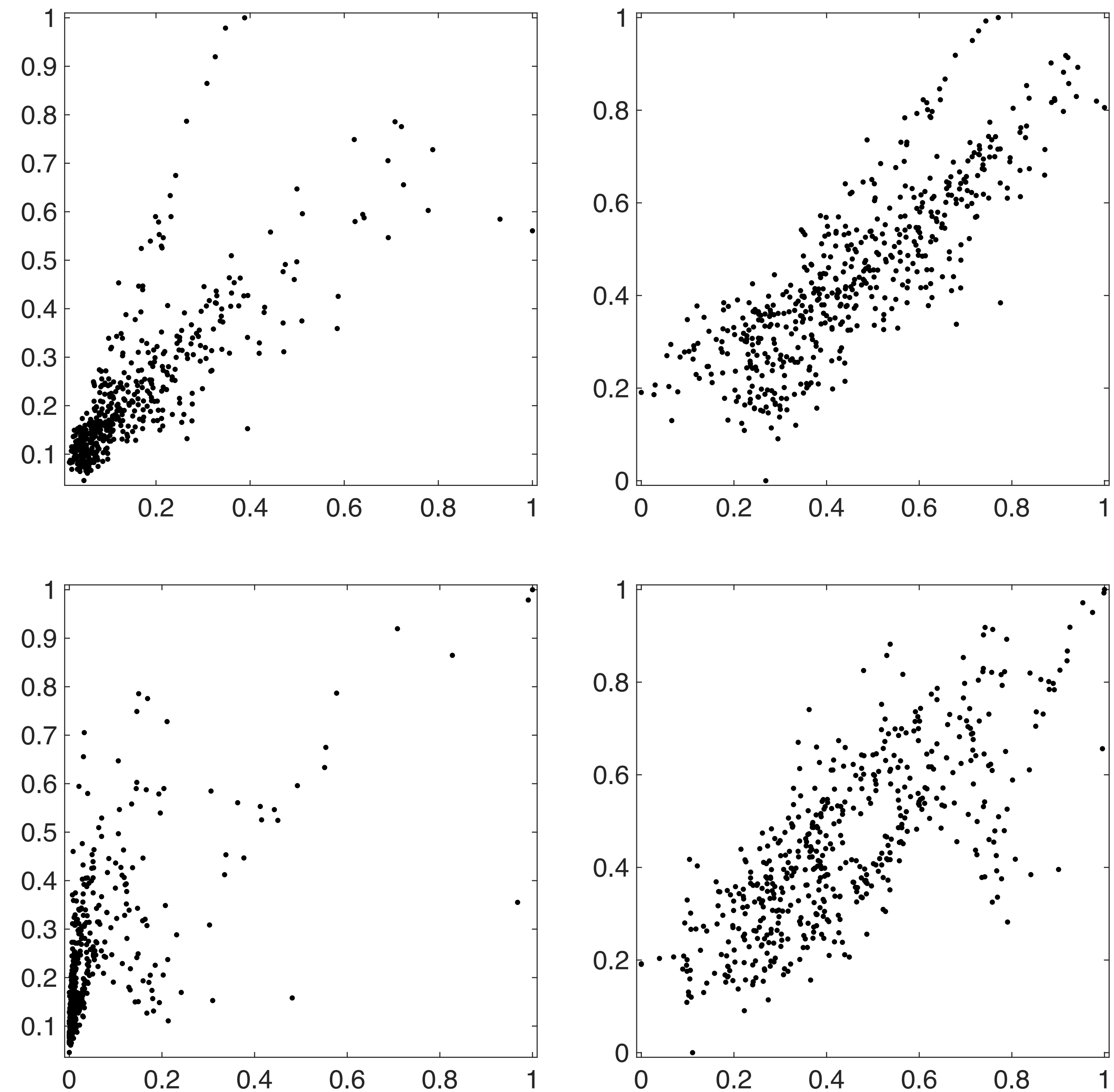
Nonlinearities in the data (2)

logit space

“flu remedies”

“tamiflu dosage”

ILI rate



Query frequency

Gaussian Processes for nonlinear modelling

Say $\boldsymbol{x} \in \mathbb{R}^d$ and we want to learn $f : \mathbb{R}^d \to \mathbb{R}$

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$$

mean function
drawn on inputs

covariance function (kernel)
drawn on pairs of inputs

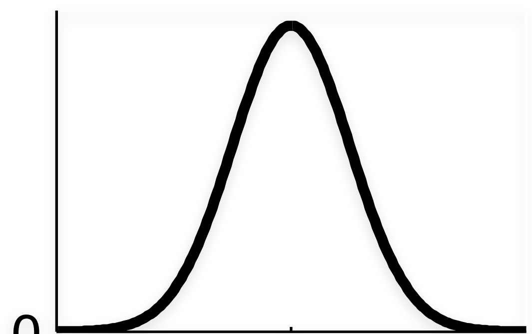
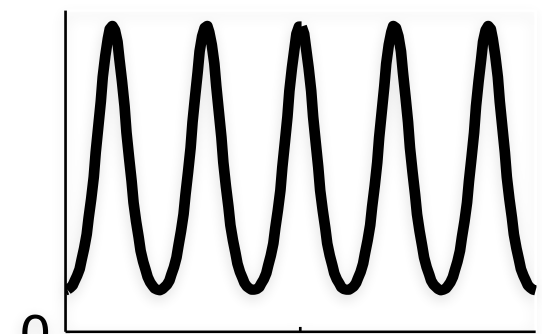
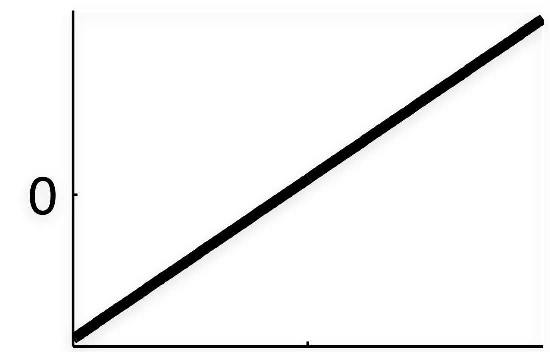
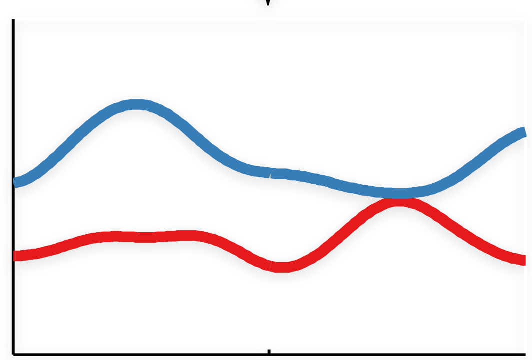
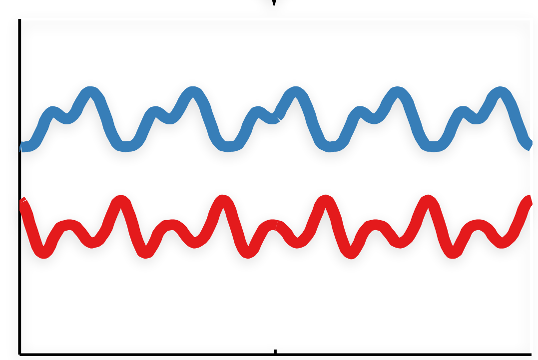
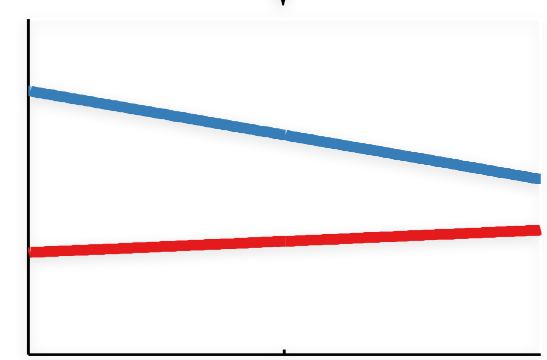
<u>Formally</u>: *Sets of random variables any finite number of which have a* ***multivariate Gaussian distribution***

Why do we use Gaussian Processes?

+ Kernelised, models nonlinearities
+ Interpretability (**A**uto**R**elevance **D**etermination)
+ Performance

(*Rasmussen & Williams, 2006*)

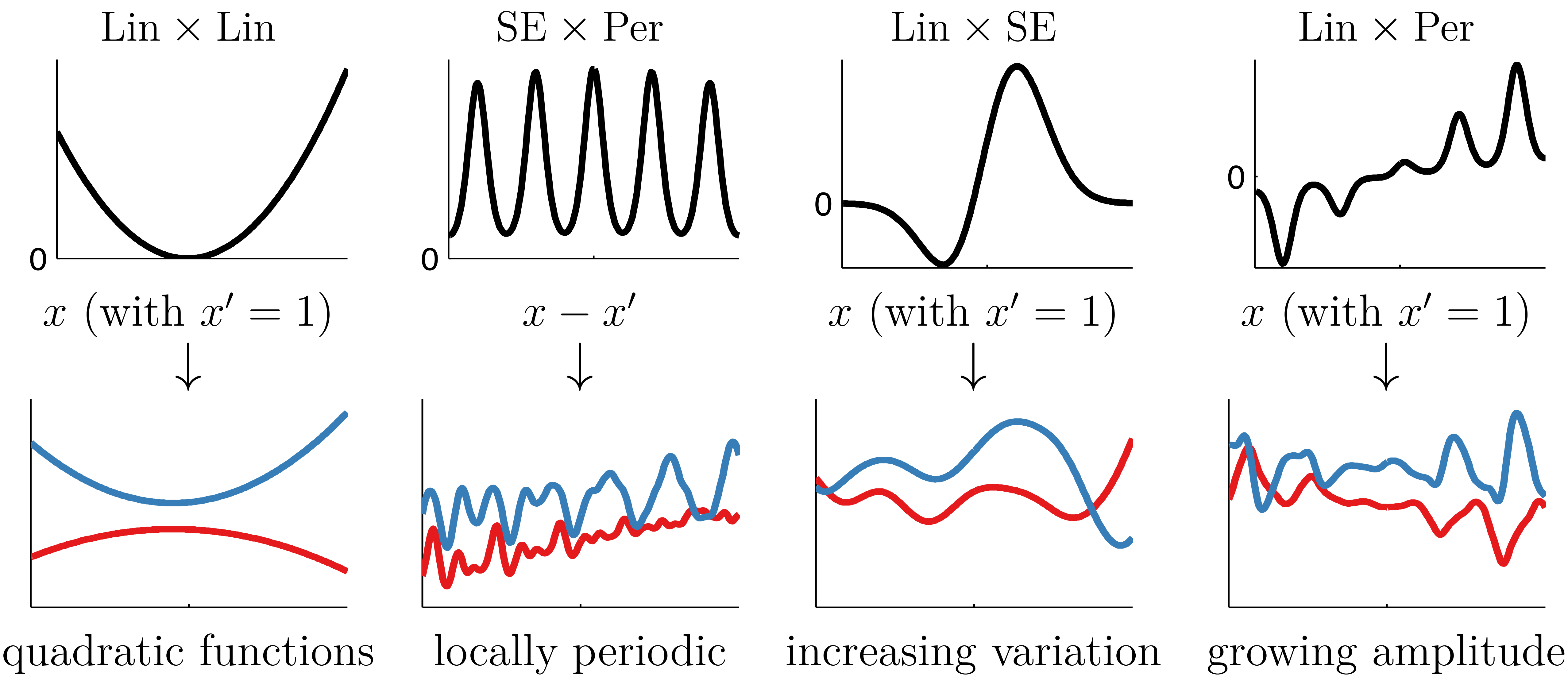
Common covariance functions (kernels)

Kernel name:	Squared-exp (SE)	Periodic (Per)	Linear (Lin)
$k(x, x') =$	$\sigma_f^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right)$	$\sigma_f^2 \exp\left(-\frac{2}{\ell^2}\sin^2\left(\pi\frac{x-x'}{p}\right)\right)$	$\sigma_f^2(x-c)(x'-c)$
Plot of $k(x, x')$:	 $x - x'$	 $x - x'$	 x (with $x' = 1$)
	$\downarrow$	$\downarrow$	$\downarrow$
Functions $f(x)$ sampled from GP prior:	 x	 x	 x
Type of structure:	local variation	repeating structure	linear functions

(*Duvenaud, 2014*)

Combining kernels in a GP

it is possible to **add** or **multiply** kernels
(among other operations)



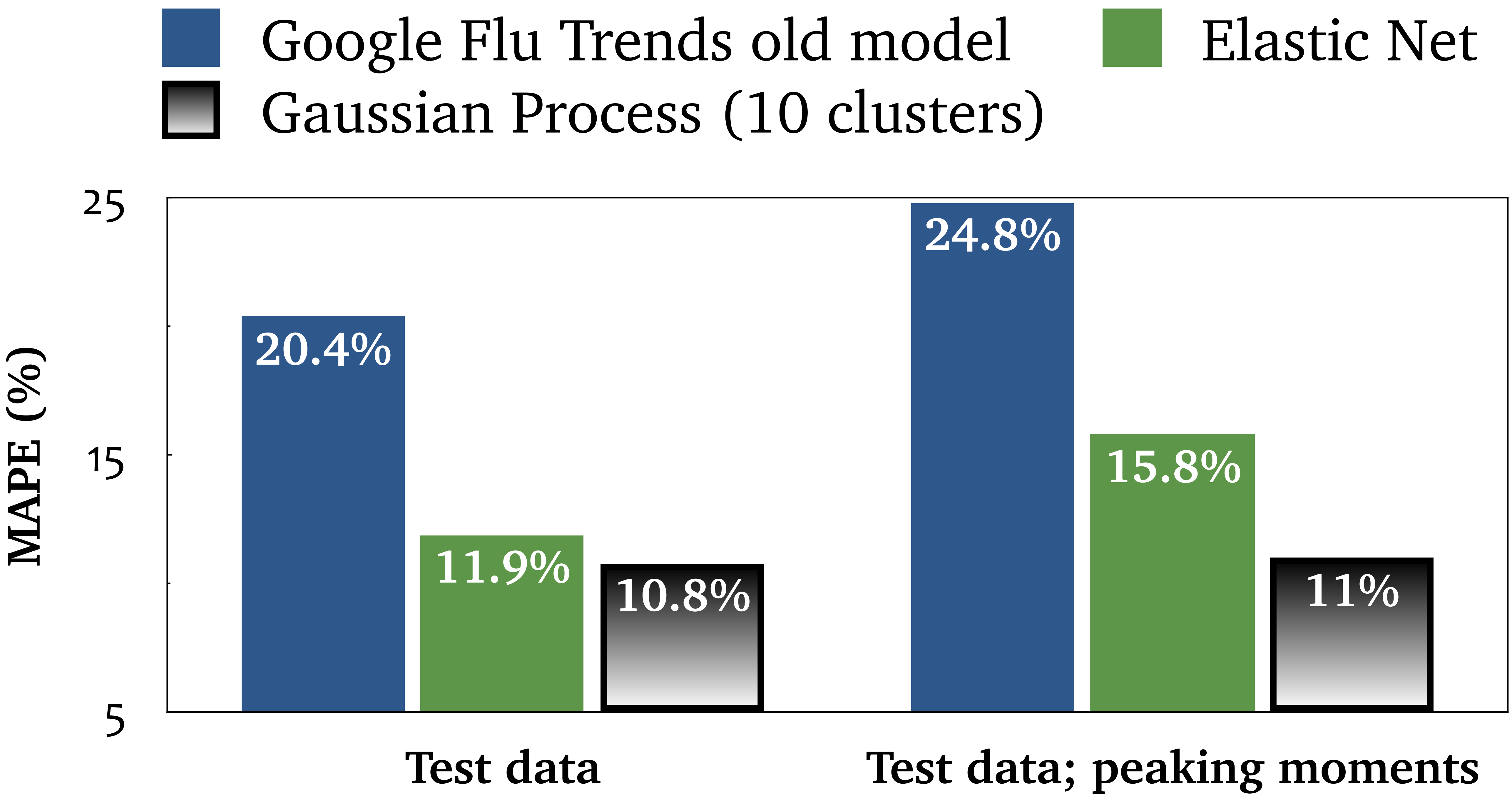
(*Duvenaud, 2014*)

GP kernel on query clusters

$$\boxed{k(\mathbf{x}, \mathbf{x}') = \left(\sum_{i=1}^{C} k_{\mathrm{SE}}(\mathbf{c}_i, \mathbf{c}_i')\right) + \sigma_{\mathrm{n}}^2 \cdot \delta(\mathbf{x}, \mathbf{x}')}$$

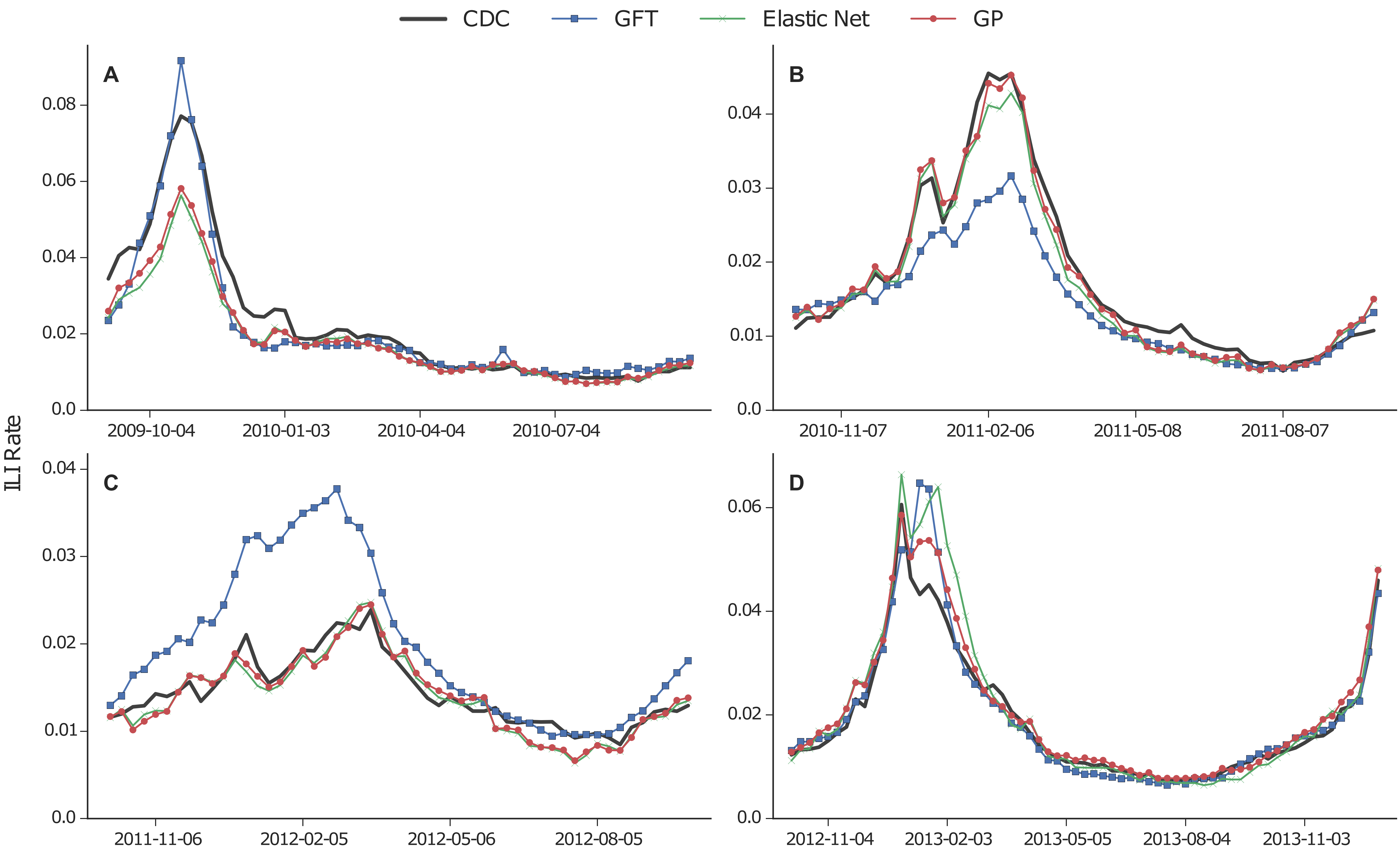
+ **protects inferences from radical changes** in the frequency of isolated queries
+ models the **contribution of various themes** (clusters) to the final prediction (*bi-product: interpretability*)
+ learns a sum of lower-dimensional functions: smaller input space, **easier learning task**, fewer samples required, more statistical traction obtained
– [***trade-off***] assumption that relationships between queries in separate clusters provide no information about ILI

Inference performance

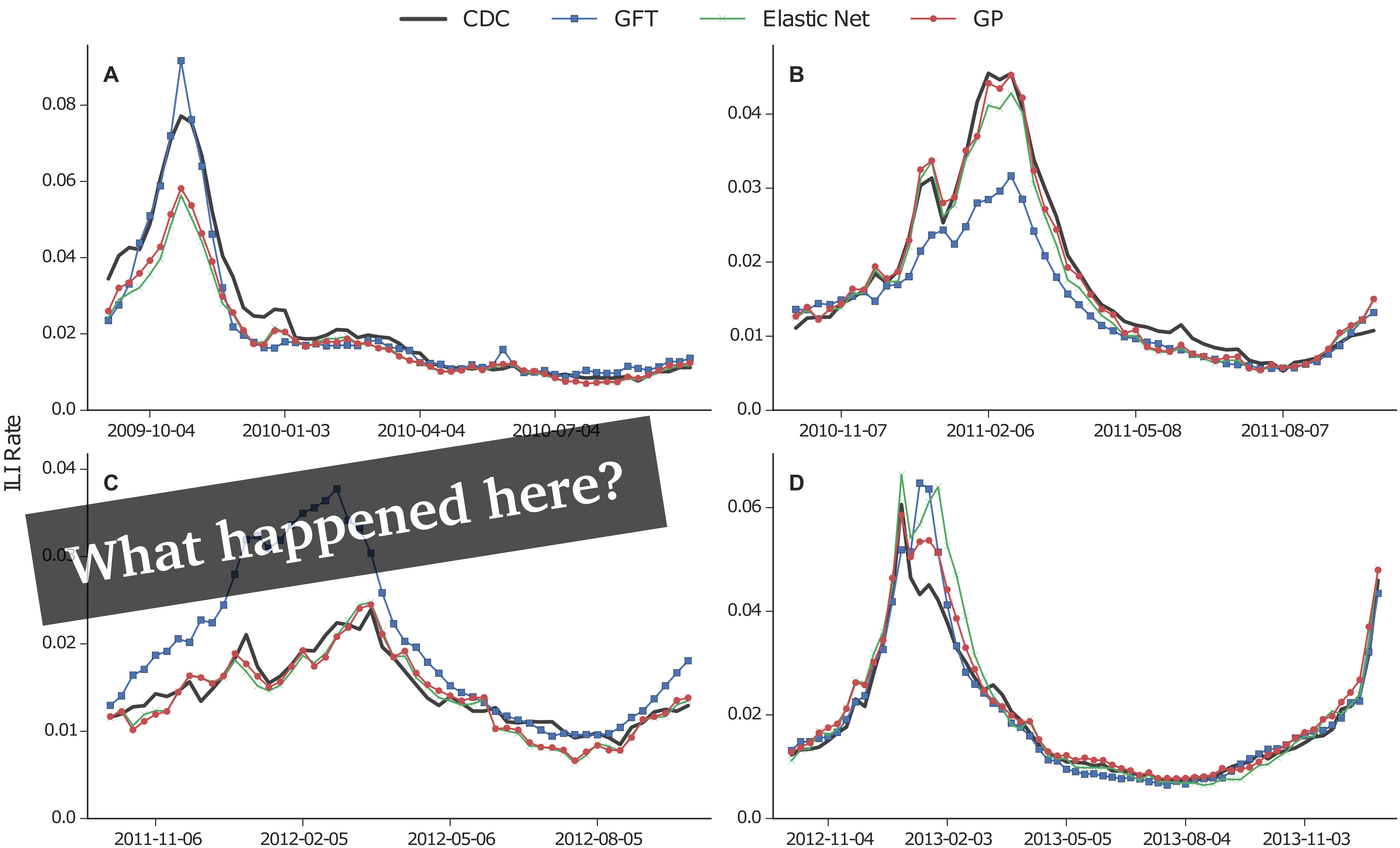


Mean absolute percentage (%) of error (MAPE) in flu rate estimates (2008-2013)

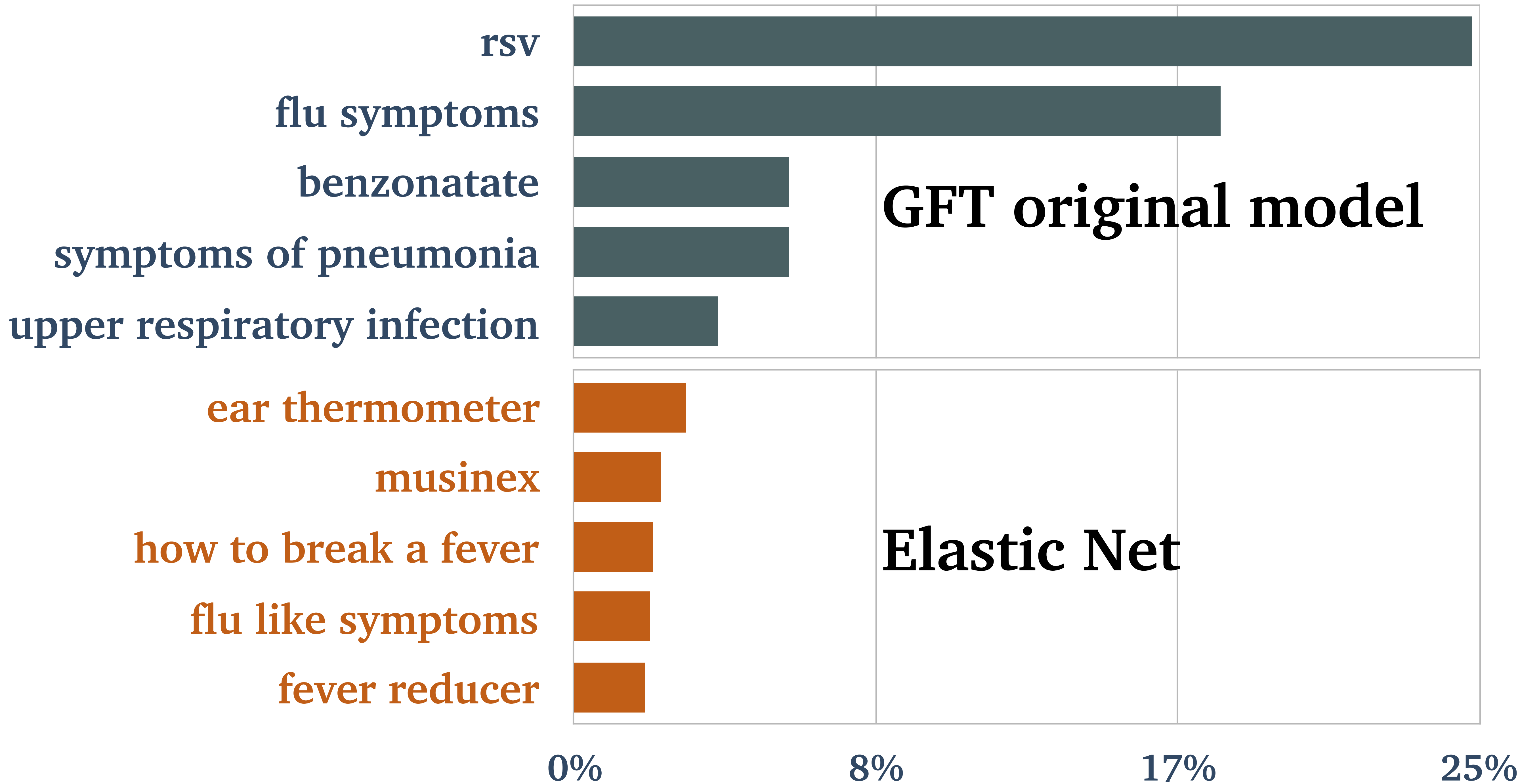
Comparative inference plots



Comparative inference plots



From 4 Dec. 2011 to 28 Apr. 2012…



Top-5 most influential search queries for flu rate inferences

I am skipping…

(1) How, and, hence, why the GP-clustering works

(2) The obvious auto-regressive extensions

(3) How we incorporated statistical NLP to further improve models (*submitted paper*)

Inferring user-level information from user-generated content

- occupational class
- income
- socio-economic status (SES)

Preotiuc-Pietro, Lampos & Aletras (ACL 2015)
Preotiuc-Pietro, Volkova, Lampos, Bachrach & Aletras (PLOS ONE, 2015)
Lampos, Aletras, Geyti, Zou & Cox (ECIR 2016)

About Twitter

And what about the statistical significance of the computed statistical significance? #inception_in_statistics

Reply Delete Favorite

RT if you love Justin Bieber. Delete ur account if you don't.

Reply Retweet Favorite

50 RETWEETS	1 FAVORITE

Why do I feel so happy today hihi. Bedtimeeee, good night. Yey thank You Lord for everything. Answered prayer ♥

Reply Retweet Favorite

i think i have the flu but i still look fabulous

Reply Retweet Favorite

About Twitter

- **140 characters** per published status (***tweet***)
- users can **follow** and **be followed**
- embedded usage of **topics** (using #hashtags)
- user **interaction** (re-tweets, @mentions, likes)
- **real-time** nature
- **biased demographics** (13-15% of UK's population, age bias etc.)
- information is **noisy** and **not always accurate**

i think i have the flu but i still look fabulous

Reply Retweet Favorite

Linguistic expression and demographics

"Socioeconomic variables are influencing language use."

(*Bernstein, 1960*; *Labov, 1972/2006*)

+ **Validate this hypothesis** on a broader, larger data set using social media
+ **Applications**
  > research, as in computational social science, health, and psychology
  > commercial

Standard Occupational Classification (SOC)

Major Group 1 (**C1**): Managers, Directors and Senior Officials
Sub-major Group 11: Corporate Managers and Directors
Minor Group 111: Chief Executives and Senior Officials
Unit Group 1115: Chief Executives and Senior Officials
•Job: chief executive, bank manager
Unit Group 1116: Elected Officers and Representatives
Minor Group 112: Production Managers and Directors
Minor Group 113: Functional Managers and Directors
Minor Group 115: Financial Institution Managers and Directors
Minor Group 116: Managers and Directors in Transport and Logistics
Minor Group 117: Senior Officers in Protective Services
Minor Group 118: Health and Social Services Managers and Directors
Minor Group 119: Managers and Directors in Retail and Wholesale
Sub-major Group 12: Other Managers and Proprietors
Major Group (**C2**): Professional Occupations
•Job: mechanical engineer, pediatrist
Major Group (**C3**): Associate Professional and Technical Occupations
•Job: system administrator, dispensing optician
Major Group (**C4**): Administrative and Secretarial Occupations
•Job: legal clerk, company secretary
Major Group (**C5**): Skilled Trades Occupations
•Job: electrical fitter, tailor
Major Group (**C6**): Caring, Leisure and Other Service Occupations
•Job: nursery assistant, hairdresser
Major Group (**C7**): Sales and Customer Service Occupations
•Job: sales assistant, telephonist
Major Group (**C8**): Process, Plant and Machine Operatives
•Job: factory worker, van driver
Major Group (**C9**): Elementary Occupations
•Job: shelf stacker, bartender

provided by the Office for National Statistics (UK)

9 major groups

25 sub-major groups

90 minor groups

369 unit groups

Standard Occupational Classification (SOC)

The 9 major occupational classes (C1-9)

C1 — Managers, Directors & Senior Officials
(*chief executive, bank manager)*

C2 — Professional Occupations (*postdoc, pediatrist*)

C3 — Associate Professional & Technical
(*system administrator, dispensing optician)*

C4 — Administrative & Secretarial (*legal clerk, secretary*)

C5 — Skilled Trades (*electrical fitter, tailor*)

C6 — Caring, Leisure, Other Service
(*nursery assistant, hairdresser)*

C7 — Sales & Customer Service (*sales assistant, telephonist*)

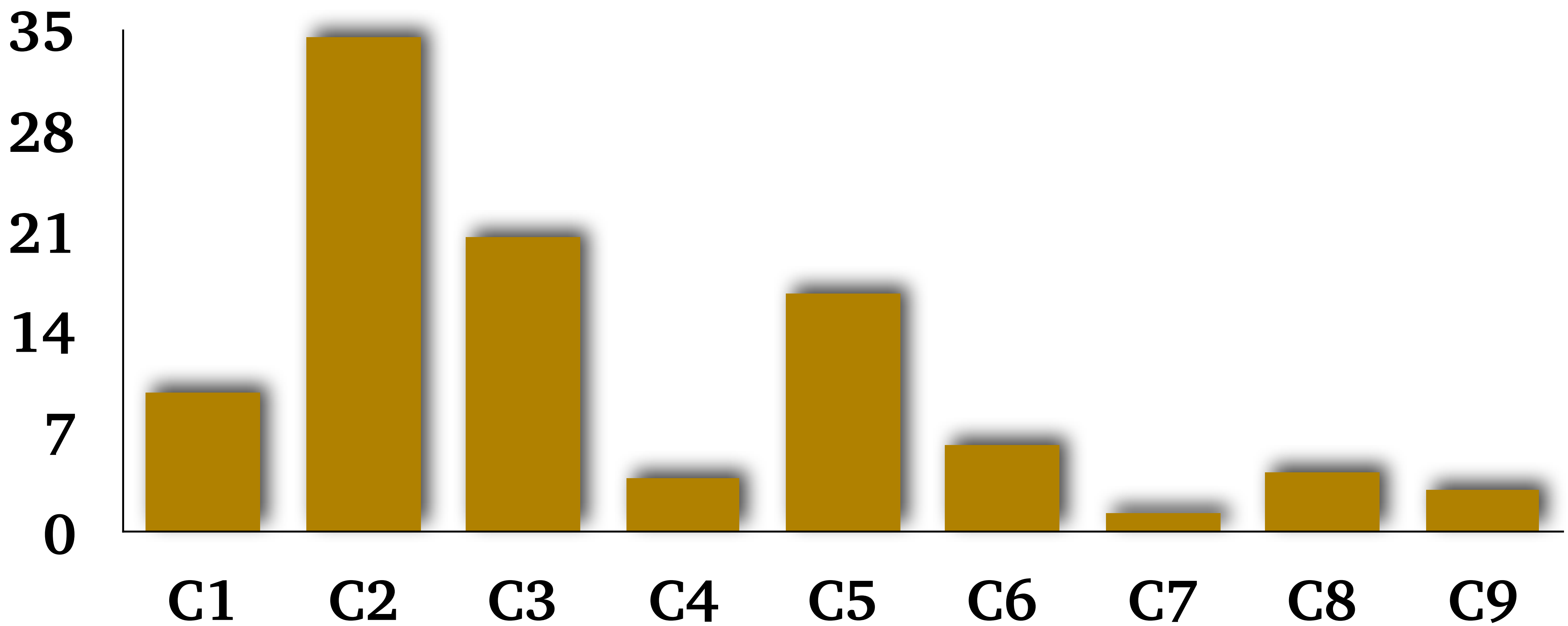
C8 — Process, Plant and Machine Operatives
(*factory worker, van driver)*

C9 — Elementary (*shelf stacker, bartender*)

Forming a Twitter user data set

+ **5,191** Twitter users mapped to their occupations, then mapped to one of the 9 SOC categories
+ 10 million tweets
+ **Download the data set**

% of users per SOC category



Twitter user attributes (*18 in total*)

number of

— followers
— friends
— followers/friends (ratio)
— times listed
— tweets
— favourites (likes)
— unique @-mentions
— tweets/day (avg.)
— retweets/tweet (avg.)

proportion of

— retweets done
— non duplicate tweets
— retweeted tweets
— hashtags
— tweets with hashtags
— tweets with @-mentions
— @-replies
— tweets with links
— tweets in English

Similarly to our paper for user impact estimation (*Lampos et al., 2014*)

Twitter user discussion topics (I)

Topics — Word clusters (#: 30, 50, 100, **200**)

+ ***SVD*** on the graph laplacian of the word by word similarity matrix using ***normalised PMI***, *i.e.* a form of spectral clustering
(*Bouma, 2009*; *von Luxburg, 2007*)

+ ***Word2vec*** (skip-gram with negative sampling) to learn word embeddings; pairwise ***cosine similarity*** on the embeddings to derive a word by word similarity matrix; then spectral clustering on the similarity matrix
(*Mikolov et al., 2013*)

Twitter user discussion topics (II)

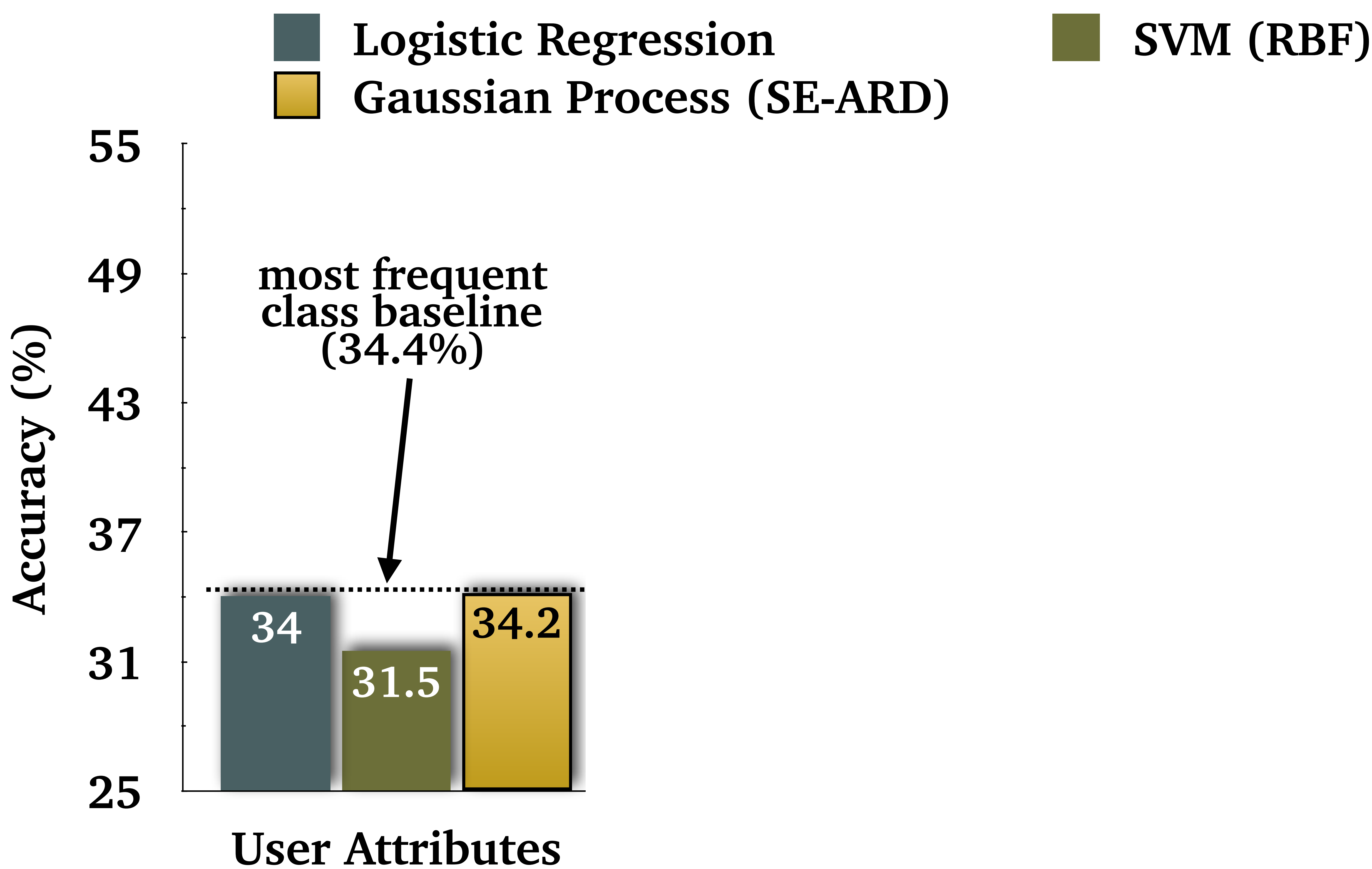
Topic	Most central words; *Most frequent words*
Arts	archival, stencil, canvas, minimalist; *art, design, print*
Health	chemotherapy, diagnosis, disease; *risk, cancer, mental, stress*
Beauty Care	exfoliating, cleanser, hydrating; *beauty, natural, dry, skin*
Higher Education	undergraduate, doctoral, academic, students, curriculum; *students, research, board, student, college, education, library*
Football	bardsley, etherington, gallas; *van, foster, cole, winger*
Corporate	consortium, institutional, firm's; *patent, industry, reports*
Elongated Words	yaaayy, wooooo, woooo, yayyyyy, yaaaaay, yayayaya, yayy; *wait, till, til, yay, ahhh, hoo, woo, woot, whoop, woohoo*
Politics	religious, colonialism, christianity, judaism, persecution, fascism, marxism; *human, culture, justice, religion, democracy*

Gaussian Process classifier

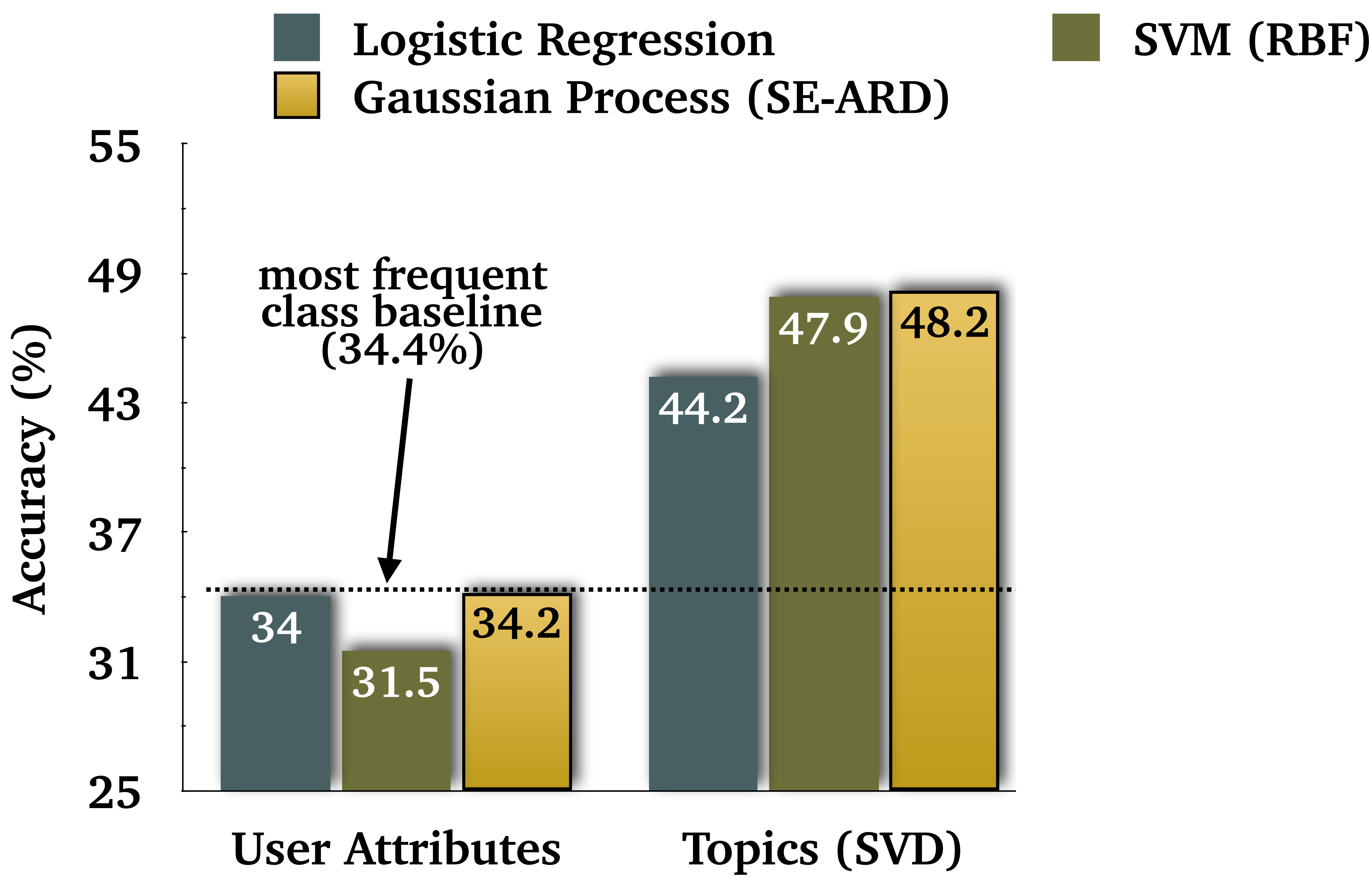
$$k_{\mathrm{ard}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left[\sum_{i}^{d} -\frac{(x_i - x_i')^2}{2l_i^2}\right]$$

+ Squared-exponential ARD covariance function: determines (quantify) the relevancy of each user feature, *i.e.* **the relevance of feature *i* is inversely proportional to the length-scale hyper-parameter l_i**
+ **9-class classification** using one vs. all
+ GP hyper-parameter learning with **Expectation Propagation**
+ Inference using **FITC** (500 inducing points)

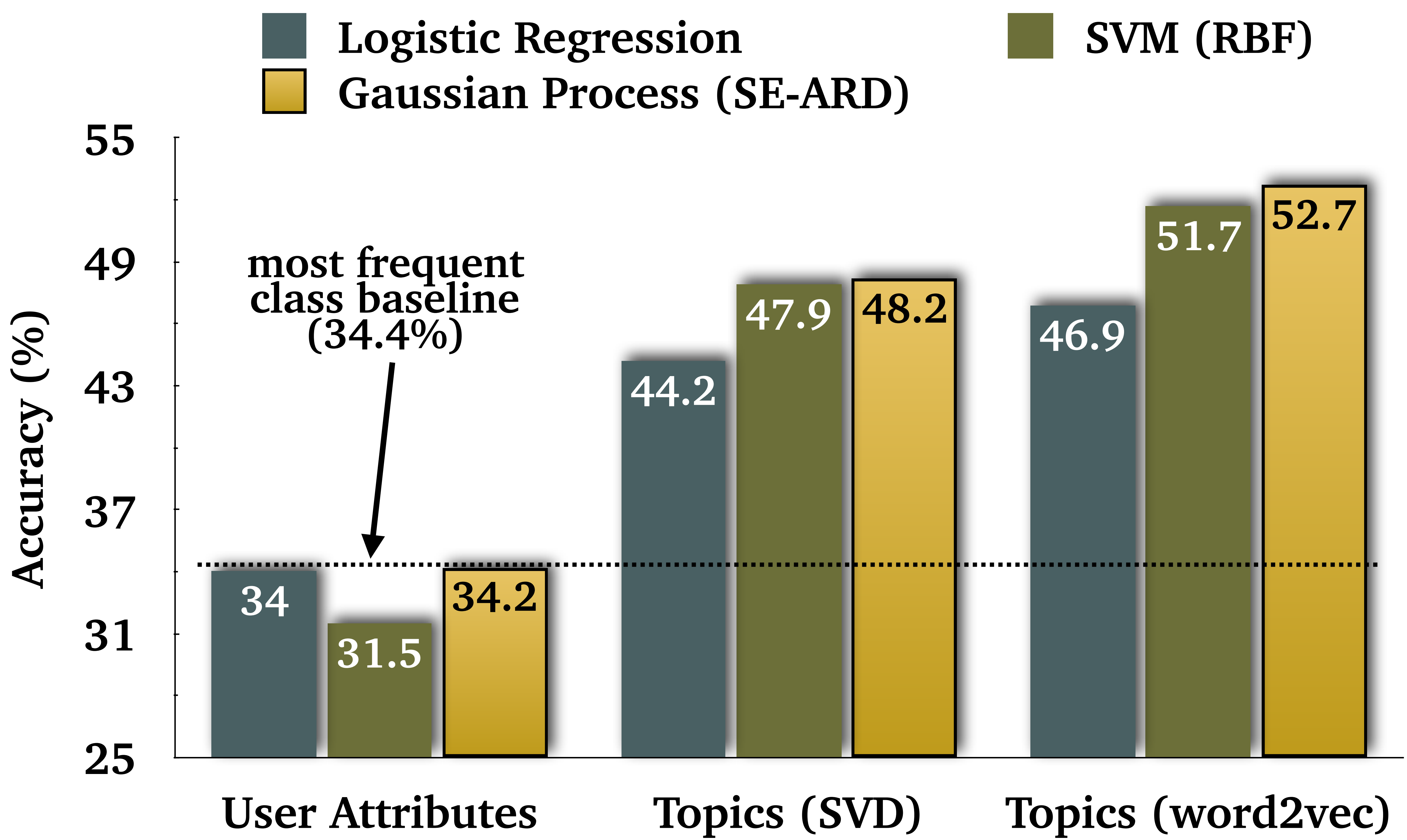
Occupation classification performance



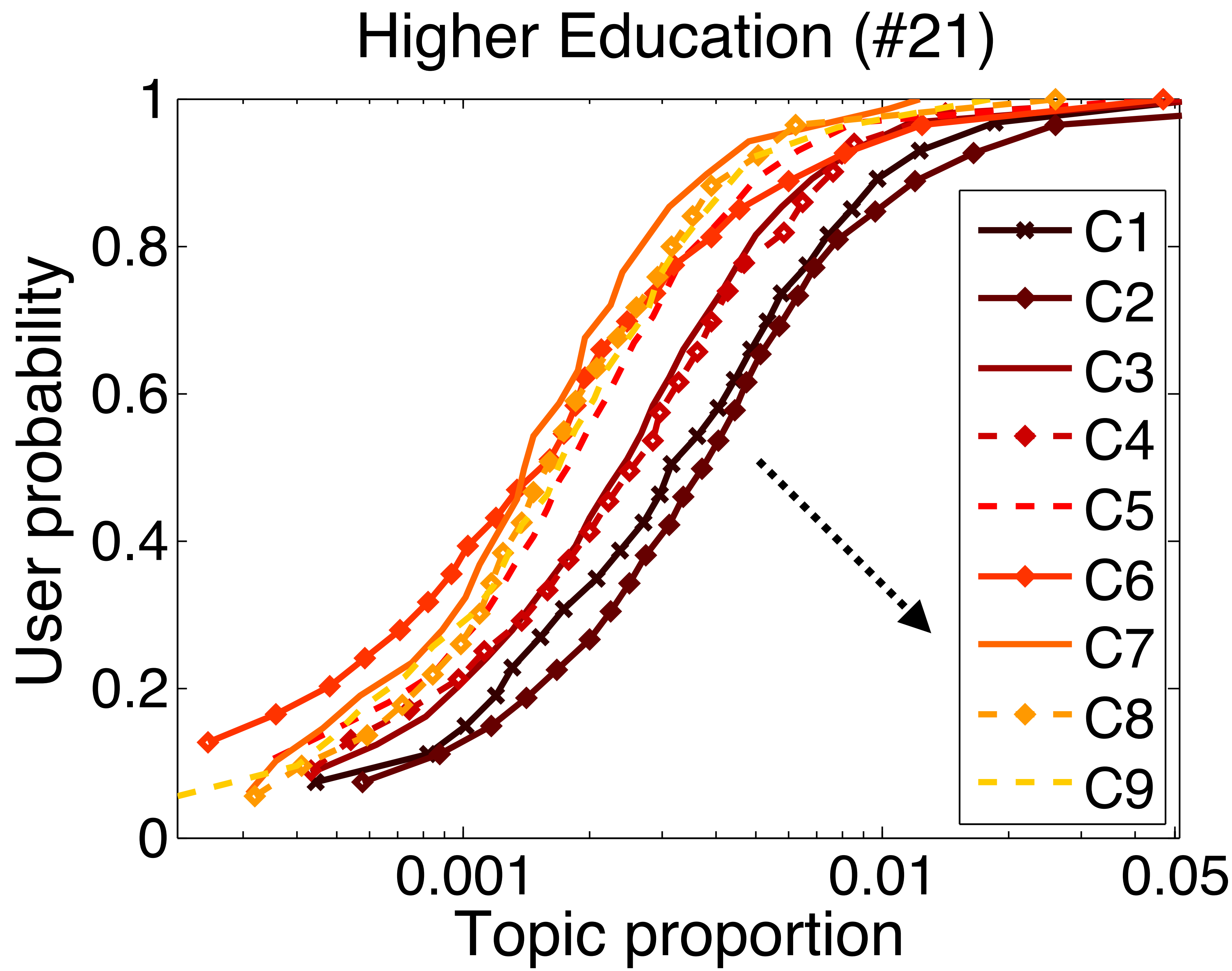
Occupation classification performance



Occupation classification performance

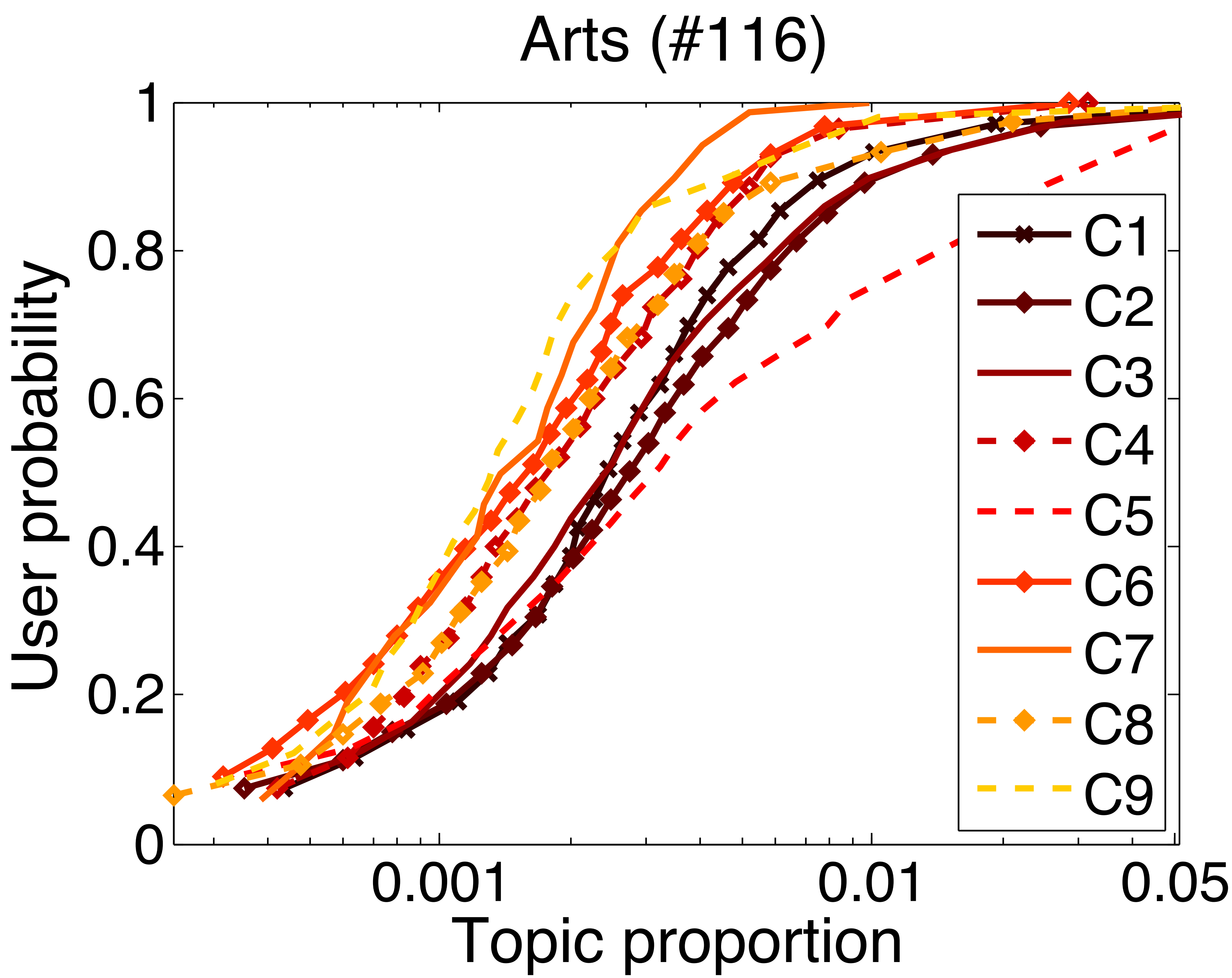


Occupation classification insights (I)



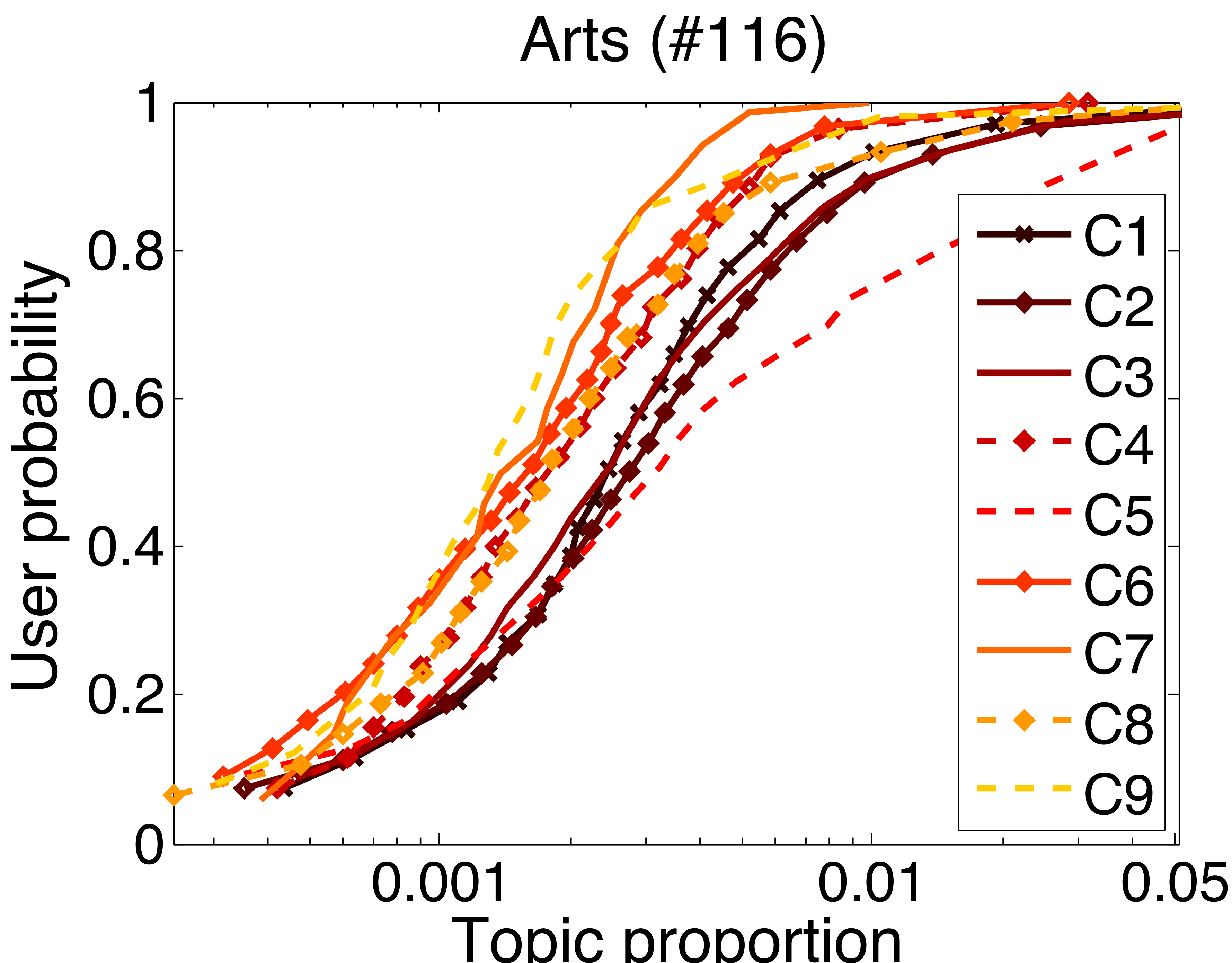
CDF of the topic "**Higher Education**": Topic **more prevalent in the upper classes** (C2, which includes education professionals, and C1), and less so in the lower classes

Occupation classification insights (II)



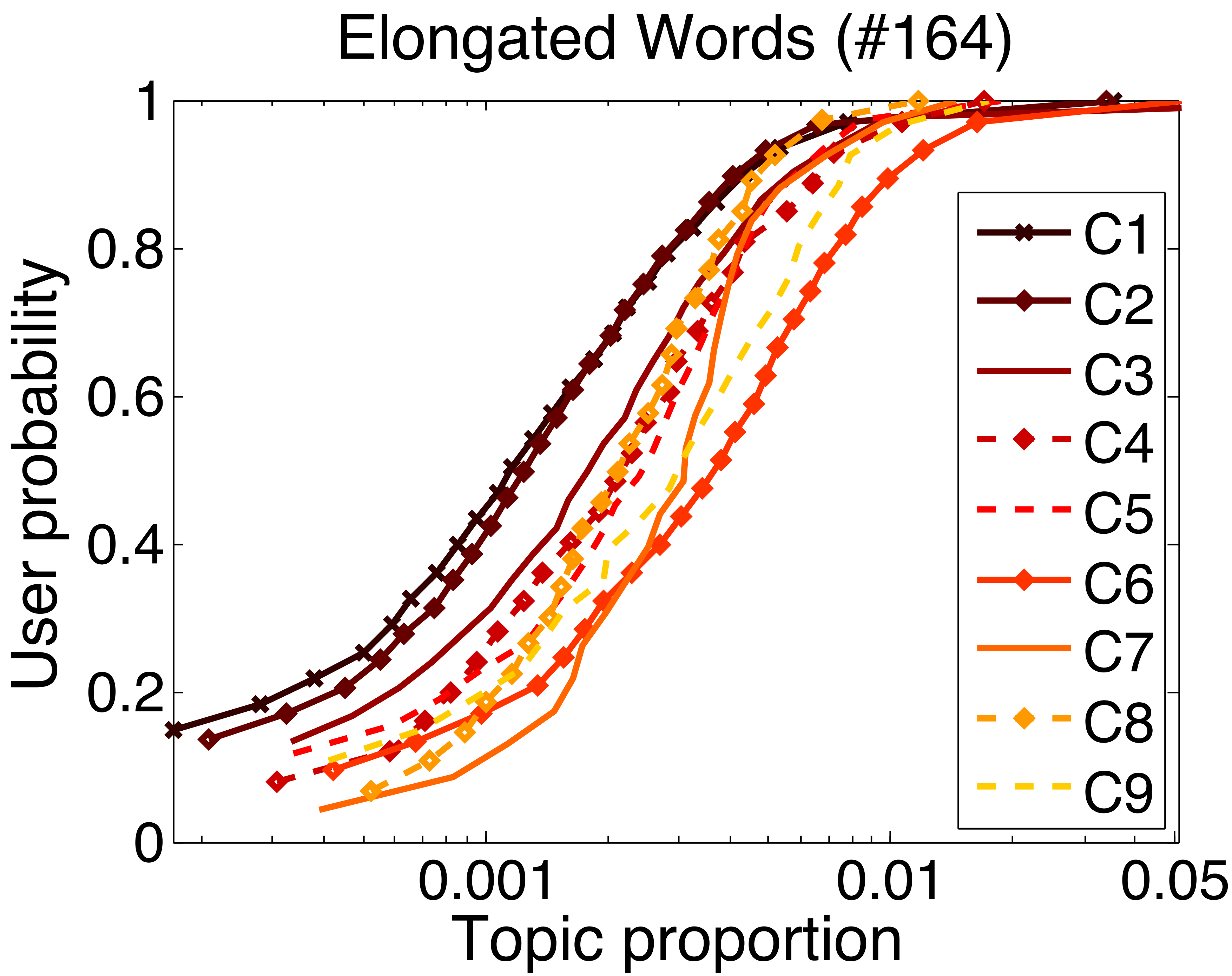
CDF of the topic "**Arts**": Topic **more prevalent in C5** (which includes artists) and **the upper classes**

Occupation classification insights (II)



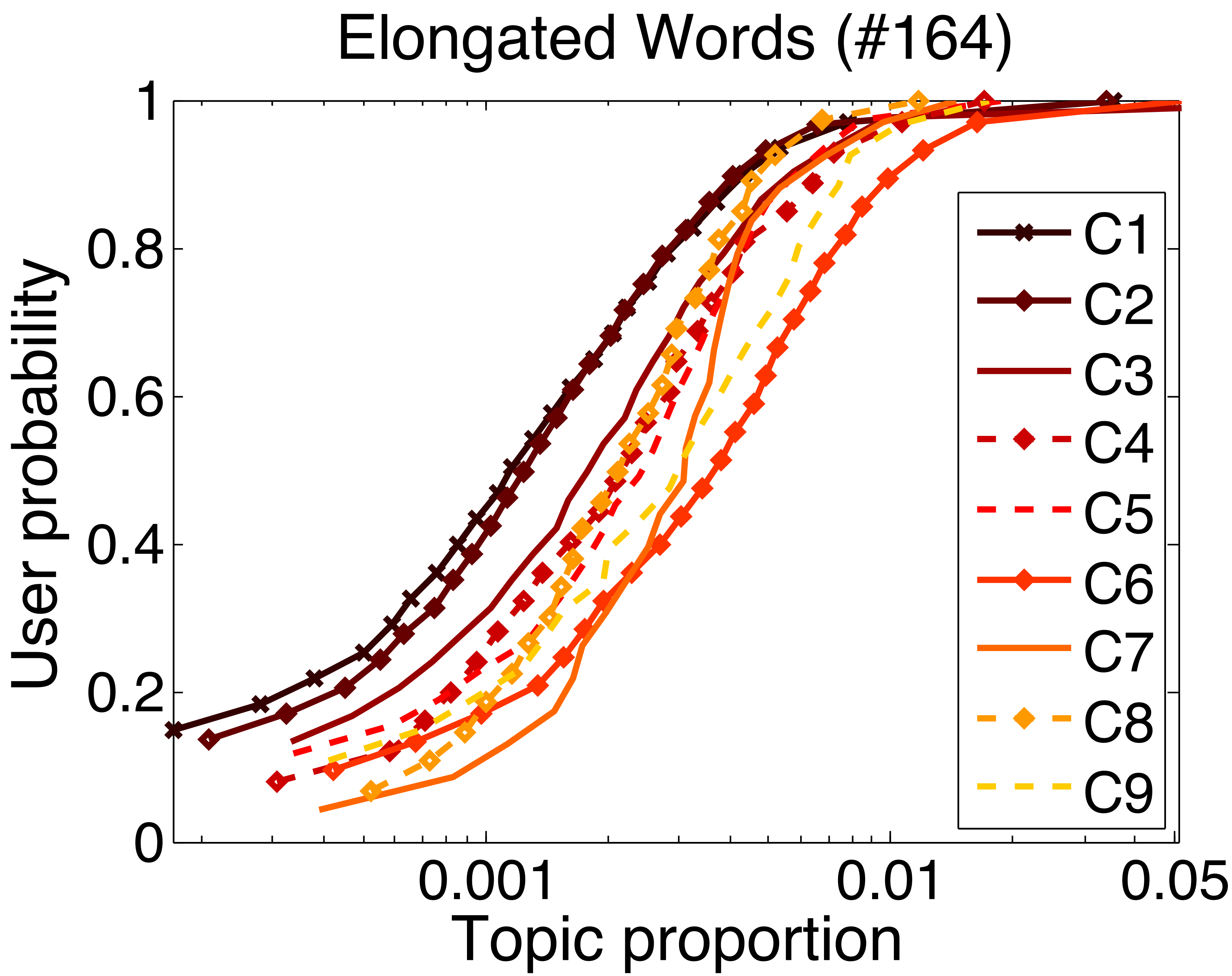
CDF of the topic "**Arts**": Topic **more prevalent in C5** (which includes artists) and **the upper classes**

Occupation classification insights (III)



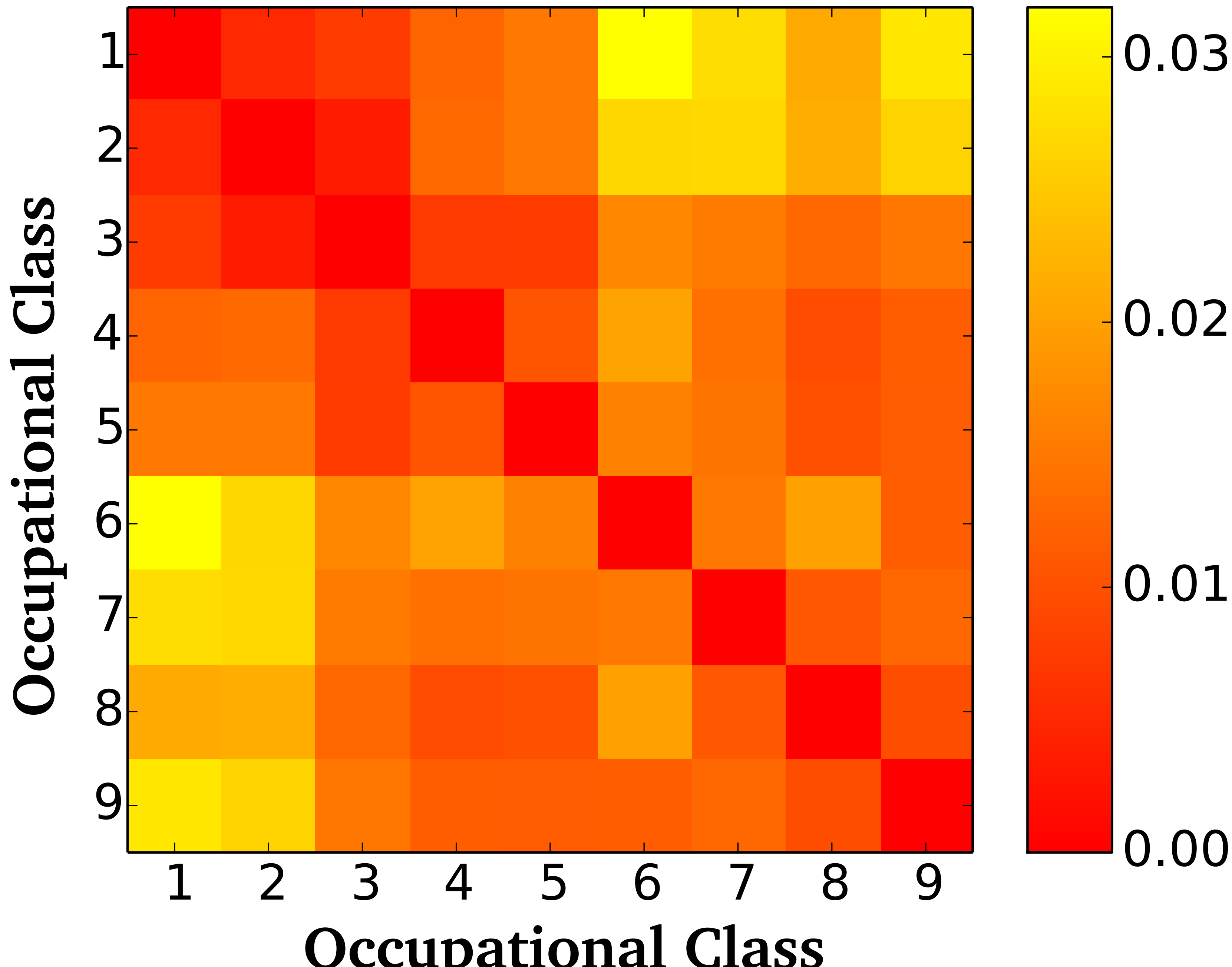
CDF of the topic “Elongated Words”: Topic more prevalent in the lower classes, and less so in the upper classes

Occupation classification insights (III)



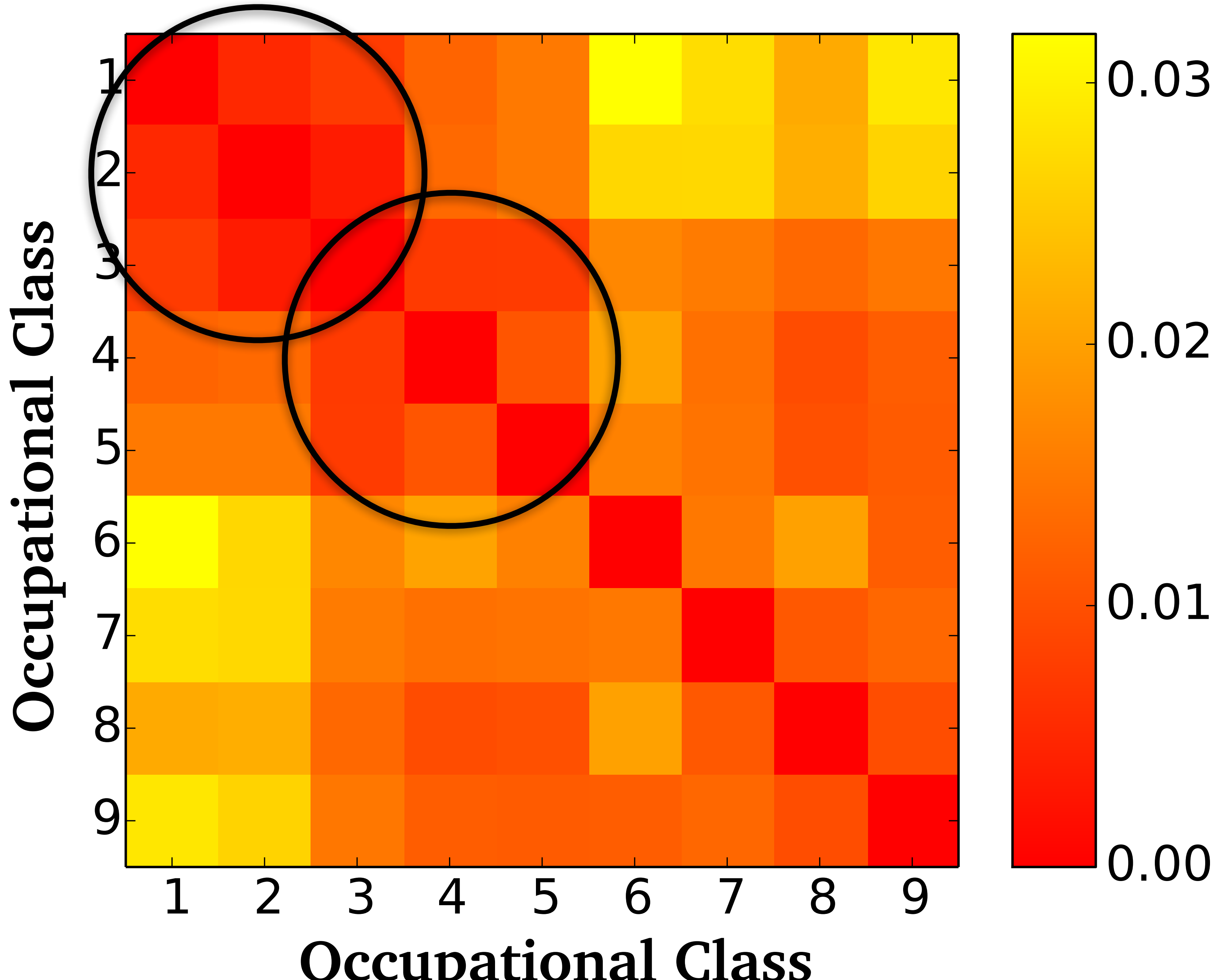
CDF of the topic “**Elongated Words**”: Topic **more prevalent in the lower classes**, and less so in the upper classes

Occupation classification insights (IV)



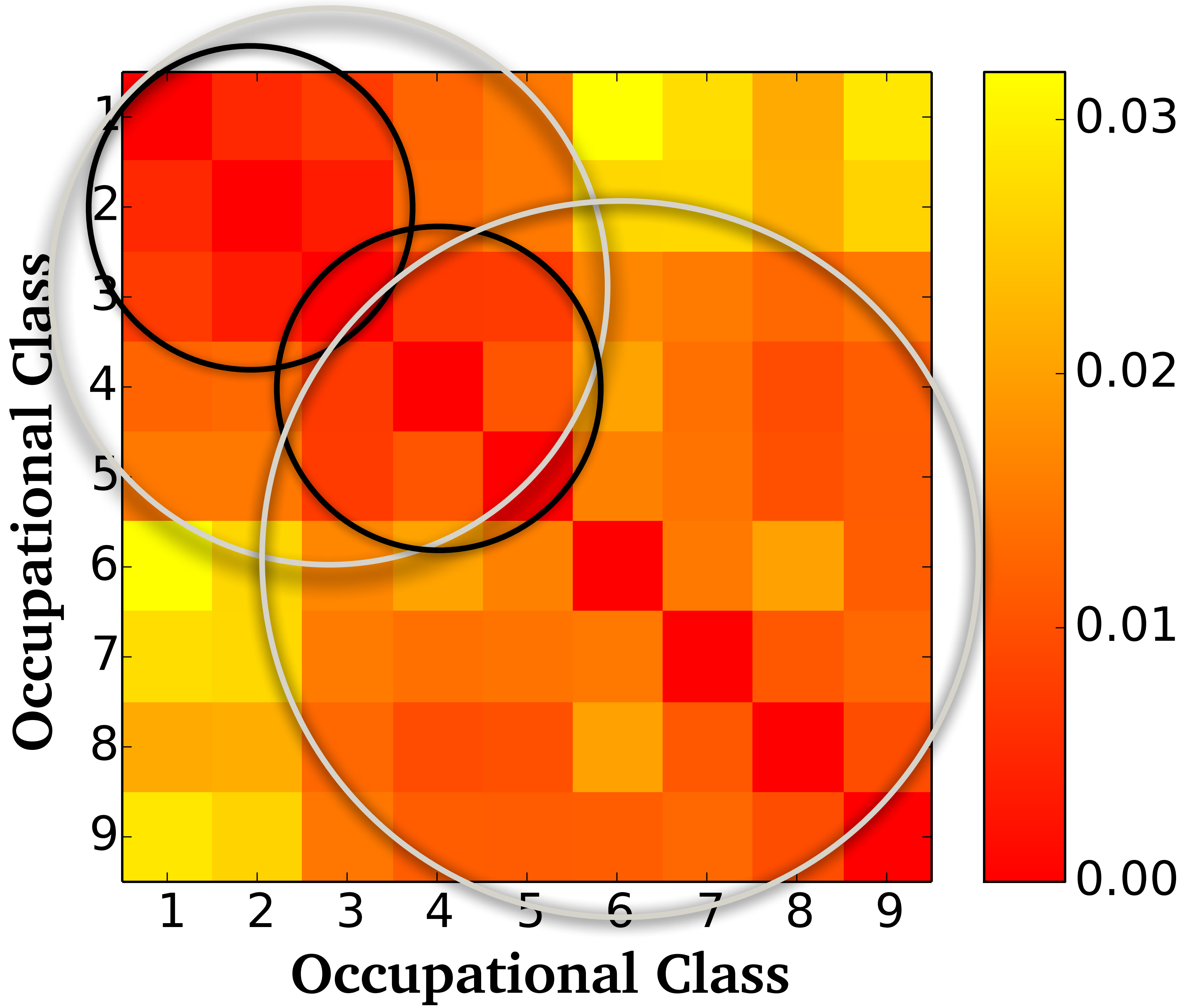
Topic distribution distance (*Jensen-Shannon divergence*) for the different occupational classes (1-9)

Occupation classification insights (IV)



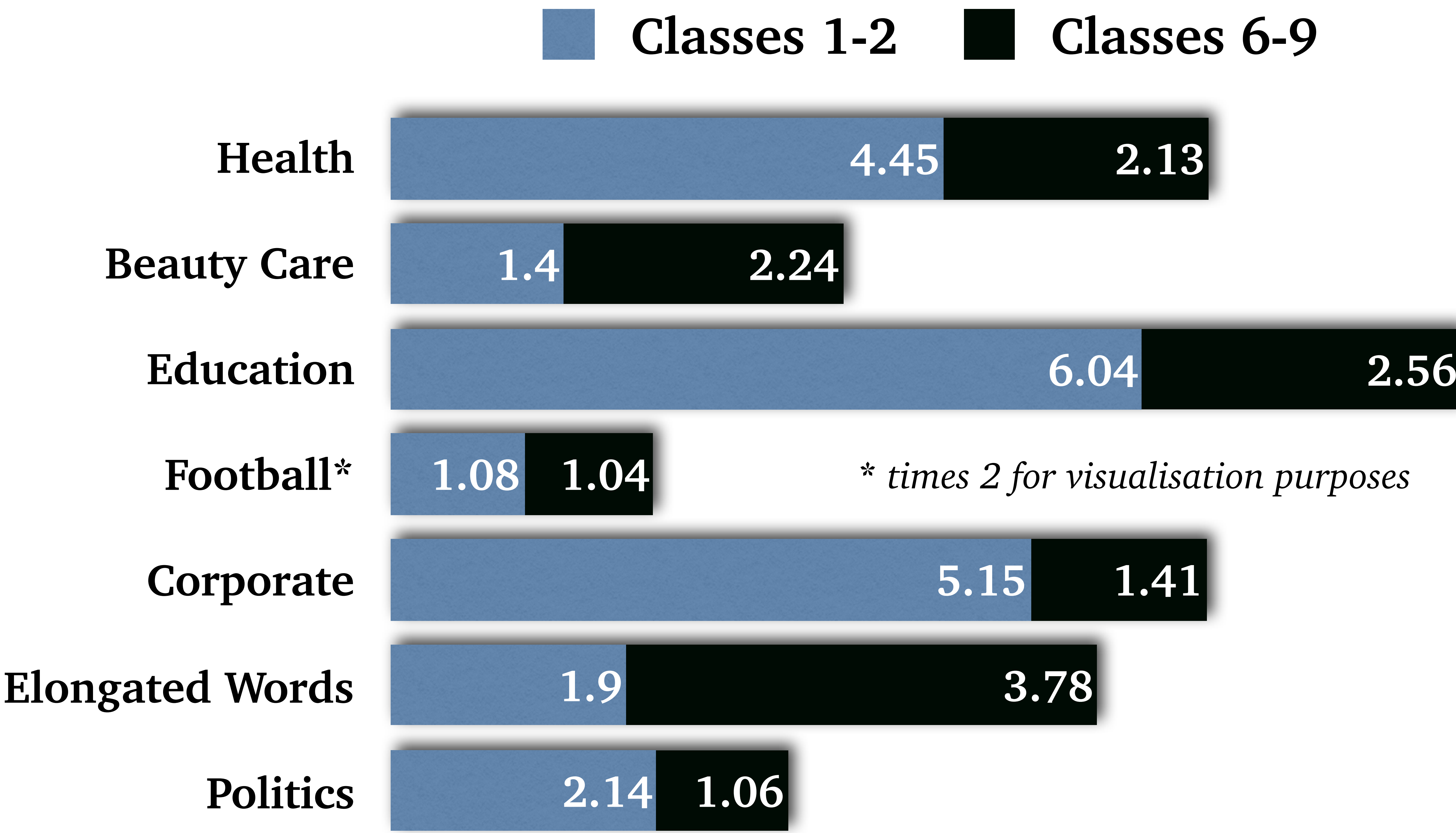
Topic distribution distance (*Jensen-Shannon divergence*) for the different occupational classes (1-9)

Occupation classification insights (IV)



Topic distribution distance (*Jensen-Shannon divergence*)
for the different occupational classes (1-9)

Occupation classification insights (V)



Topic scores for occupational class supersets

Additional ‘perceived’ user features

- Previously used features: **Profile** features, **Shallow** profile features, and **Topics**
- Based on the work of *Volkova et al. (2015)*, we also incorporated:
  - Inferred Psycho-**Demo**graphic features (**15**)
    e.g. gender, age, education level, religion, life satisfaction, excitement, anxiety etc.
  - **Emotions** (9)
    e.g. positive / negative sentiment, joy, anger, fear, disgust, sadness, surprise etc.

Defining the user income regression task

Group 112: Production Managers and Directors (50,952 GBP/year)	
•Job titles: engineering manager, managing director, production manager, construction manager, quarry manager, operations manager	
Group 241: Conservation and Environment Professionals (53,679 GBP/year)	
•Job titles: conservation officer, ecologist, energy conservation officer, heritage manager, marine conservationist, energy manager, environmental consultant, environmental engineer, environmental protection officer, environmental scientist, landfill engineer	
Group 312: Draughtspersons and Related Architectural Technicians (29,167 GBP/year)	
•Job titles: architectural assistant, architectural, technician, construction planner, planning enforcement officer, cartographer, draughtsman, CAD operator	
Group 411: Administrative Occupations: Government and Related Organisations (20,373 GBP/year)	
•Job titles: administrative assistant, civil servant, government clerk, revenue officer, benefits assistant, trade union official, research association secretary	
Group 541: Textiles and Garments Trades (18,986 GBP/year)	
•Job titles: knitter, weaver, carpet weaver, curtain maker, upholsterer, curtain fitter, cobbler, leather worker, shoe machinist, shoe repairer, hosiery cutter, dressmaker, fabric cutter, tailor, tailoress, clothing manufacturer, embroiderer, hand sewer, sail maker, upholstery cutter	
Group 622: Hairdressers and Related Services (10,793 GBP/year)	
•Job titles: barber, colourist, hair stylist, hairdresser, beautician, beauty therapist, nail technician, tattooist	
Group 713: Sales Supervisors (18,383 GBP/year)	
•Job titles: sales supervisor, section manager, shop supervisor, retail supervisor, retail team leader	
Group 813: Assemblers and Routine Operatives (22,491 GBP/year)	
•Job titles: assembler, line operator, solderer, quality assurance inspector, quality auditor, quality controller, quality inspector, test engineer, weightbridge operator, type technician	
Group 913: Elementary Process Plant Occupations (17,902 GBP/year)	
•Job titles: factory cleaner, hygene operator, industrial cleaner, paint filler, packaging operator, material handler, packer	

Same Twitter data set as in the job classification task

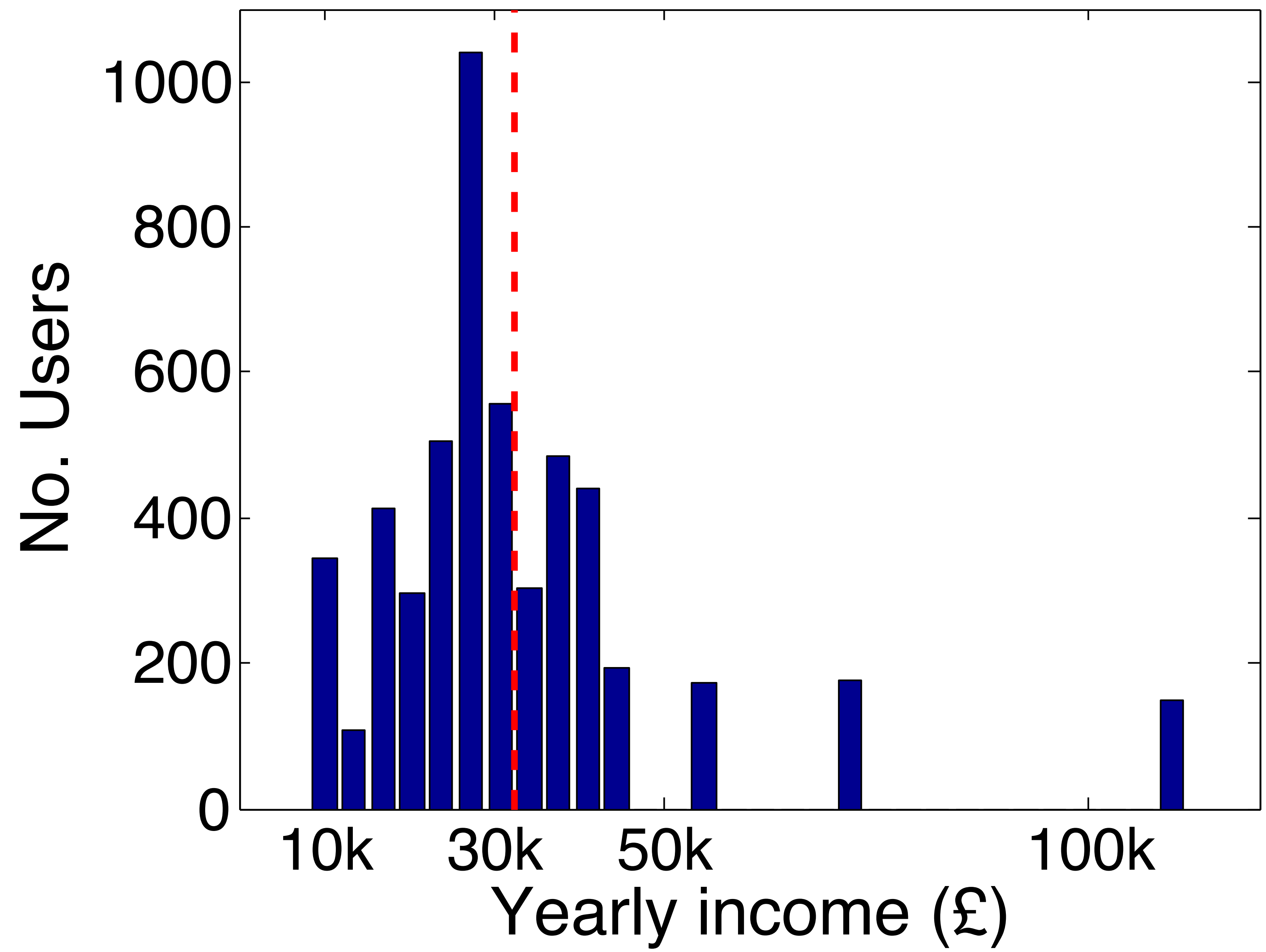
Use an income mapping from SOC to create real-valued target data for the regression task

User income regression: data

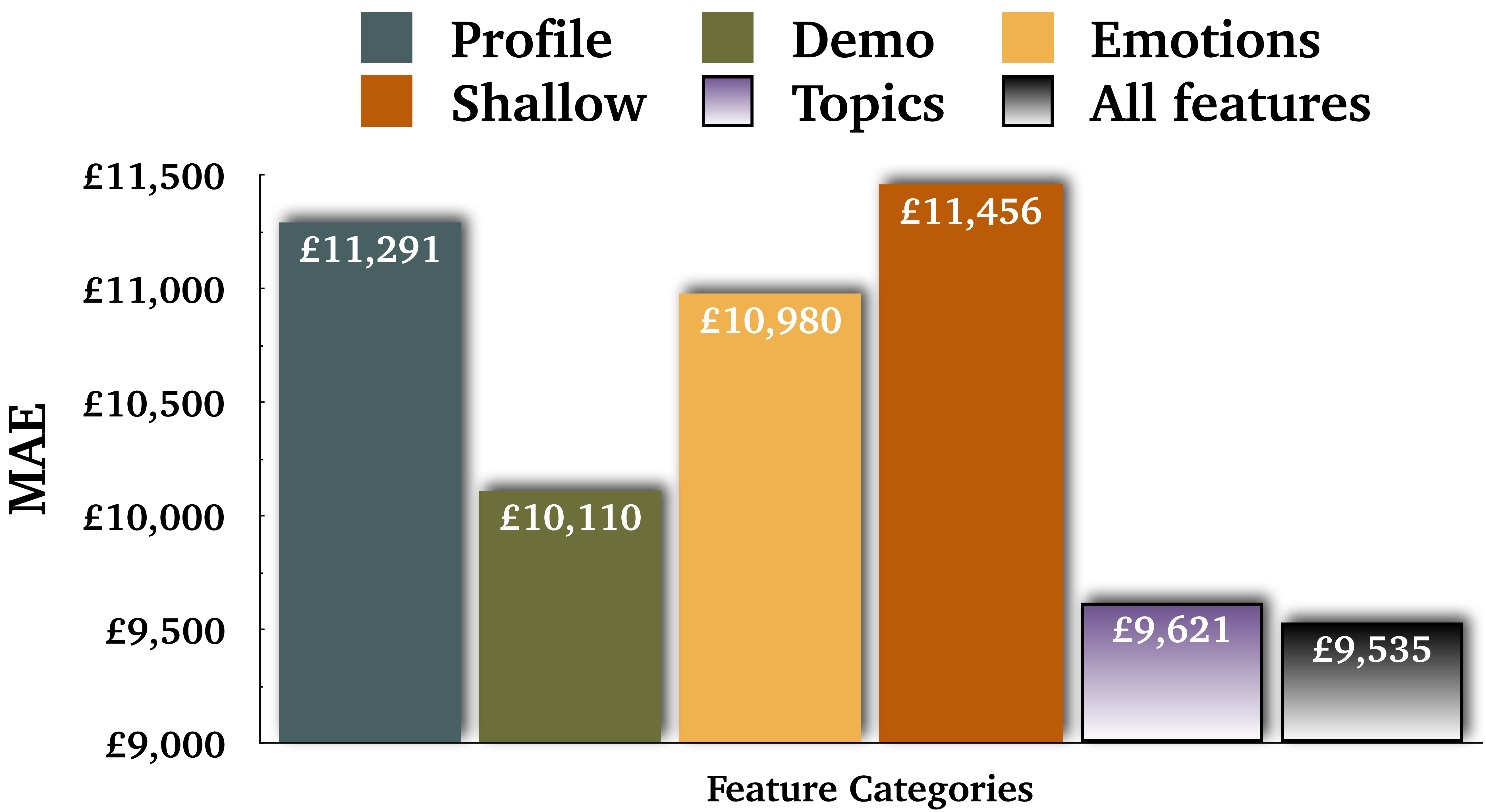
+ **5,191** Twitter users mapped to their occupations, then mapped to an average income in GBP (£) using the *SOC* taxonomy

+ ~11 million tweets

+ Download the data

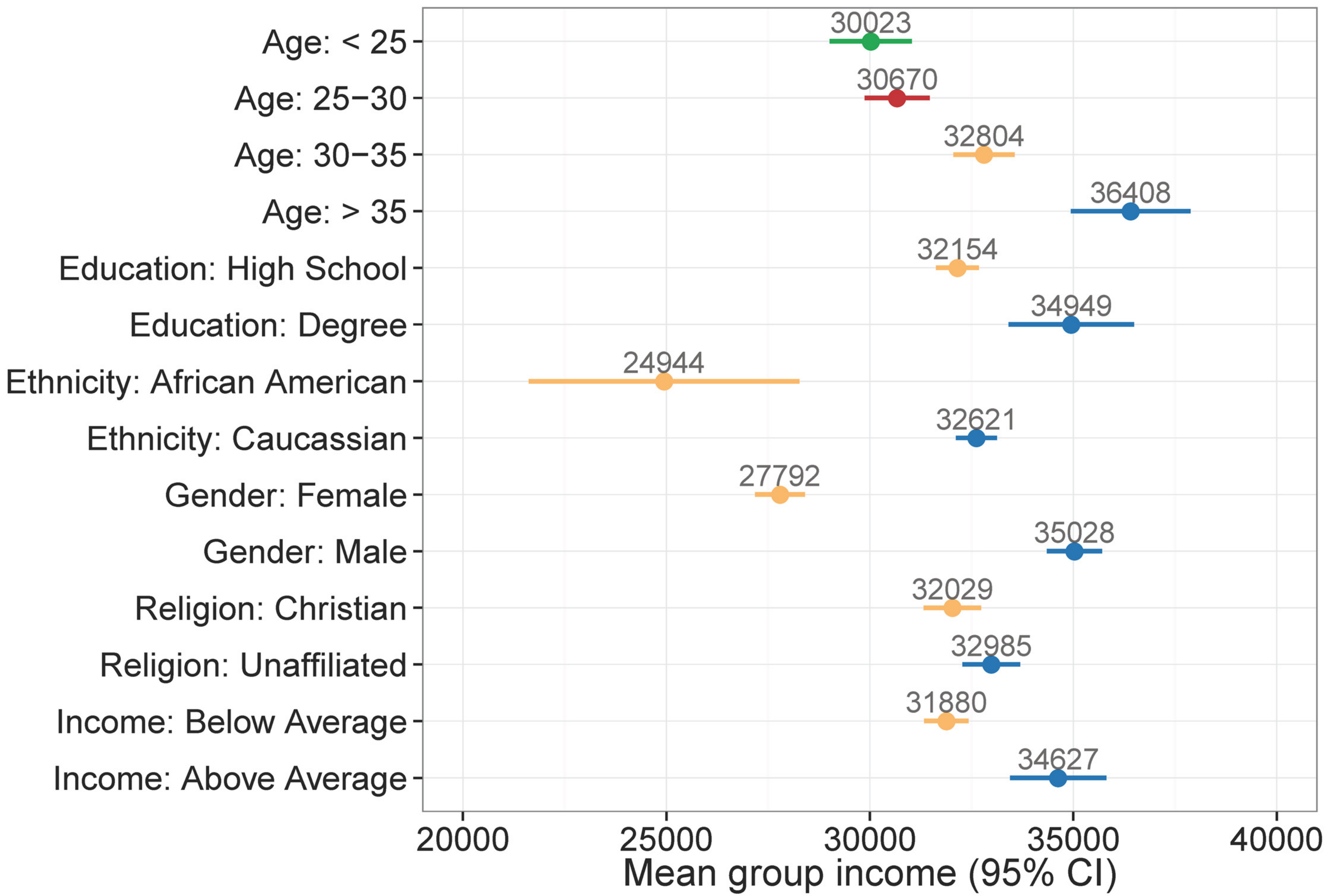


User income regression performance



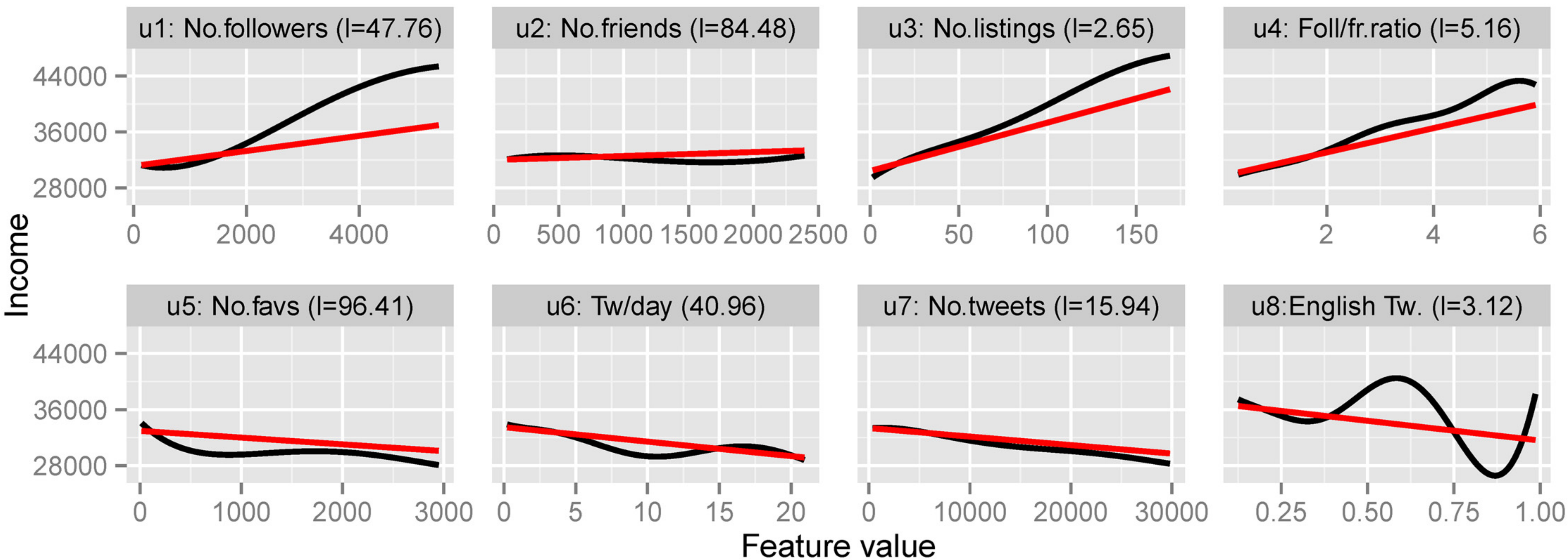
Income inference error (Mean Absolute Error) using GP regression or a linear ensemble for all features

User income regression insights (I)



User income regression insights (II)

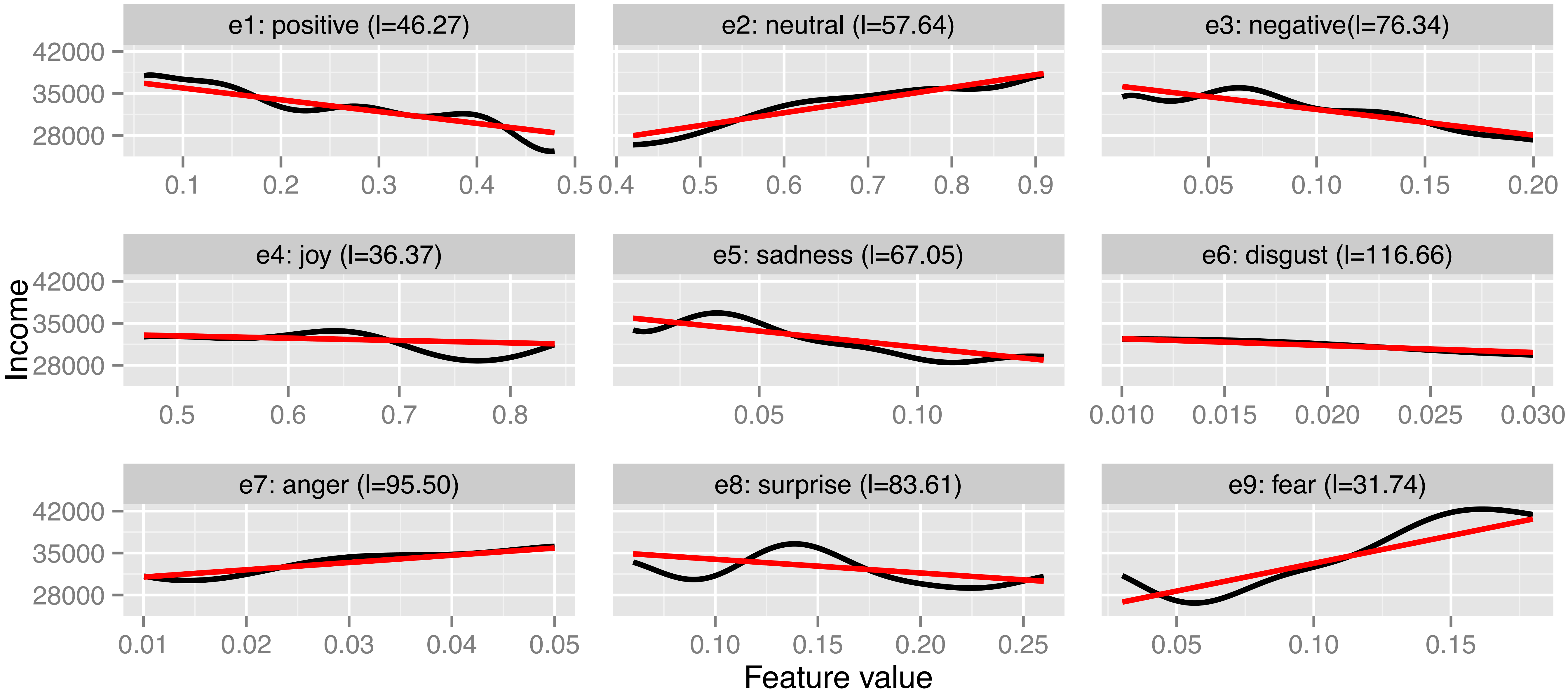
Relating income and user attributes



Linear vs **GP** fit

User income regression insights (III)

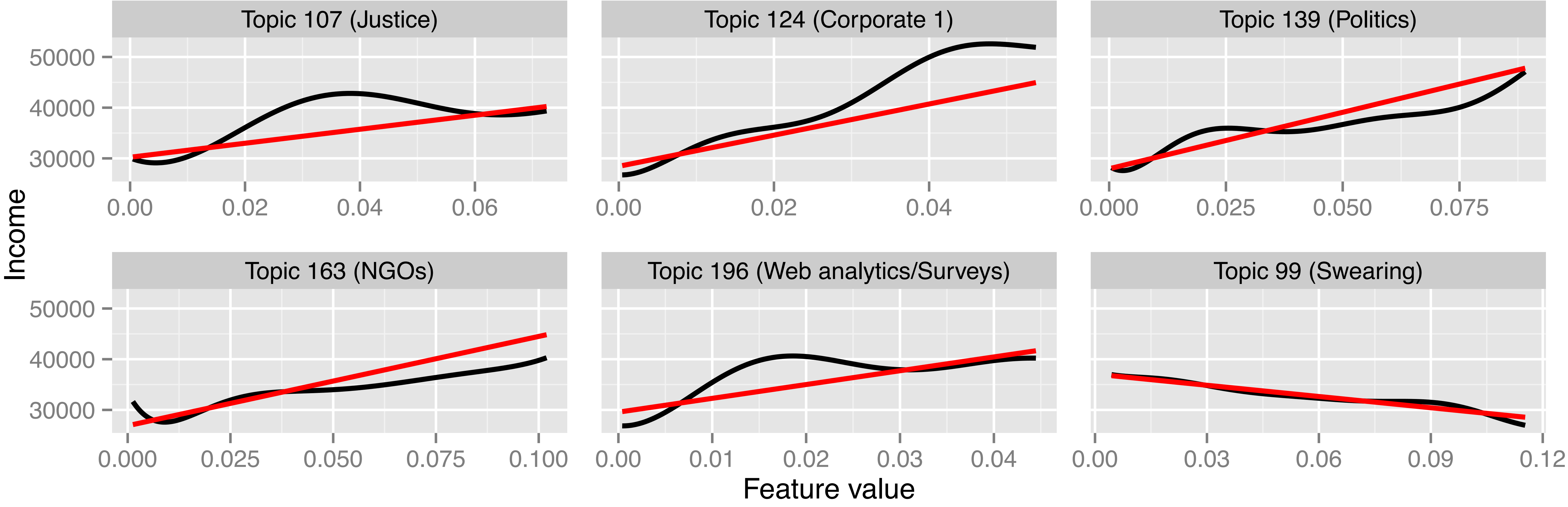
Relating income and emotion



Linear vs **GP** fit

User income regression insights (IV)

Relating income and topics of discussion



Linear vs **GP** fit

Defining a user SES classification task

Profile description on Twitter → Occupation → **SOC** category[1] → **NS-SEC**[2]

Office for National Statistics

1. **S**tandard **O**ccupational **C**lassification job groups
2. **N**ational **S**tatistics **S**ocio-**E**conomic **C**lassification: Map from the job groups in the SOC to a socioeconomic status (SES): *upper*, *middle* or *lower*

UK Twitter user data set for SES classification

+ **1,342** UK Twitter user profiles
+ 2 million tweets
+ Date interval: Feb. 1, 2014 to March 21, 2015
+ Labelled with a **socioeconomic status** (SES), using the occupational class proxy from SOC and NS-SEC: ***upper***, ***middle***, or ***lower***
+ 1,291 **user features** following the previous paradigms, *i.e.* quantifying behaviour, impact, profile info, text in tweets and topics from tweets
+ **Download the data set**

SES classification performance

3-class classification

	T1	T2	T3	P
O1	606	84	53	81.6%
O2	49	186	45	66.4%
O3	55	48	216	67.7%
R	854%	58.5%	68.8%	75.1%

middle & lower merged

	T1	T2	P
O1	584	115	83.5%
O2	126	517	80.4%
R	82.3%	81.8%	82.0%

… using a Gaussian Process classifier

Classification	Accuracy (%)	Precision (%)	Recall (%)	F1
2 classes	82.05 (2.4)	82.2 (2.4)	81.97 (2.6)	.821 (.03)
3 classes	75.09 (3.3)	72.04 (4.4)	70.76 (5.7)	.714 (.05)

Conclusions — UGC mining: From collective disease rates to individual demographics

- influenza-like illness rates
- occupational class
- income
- socio-economic status

Further thoughts

- **User-generated content** is a **valuable asset**
- **Nonlinear models** tend to perform better given the multimodality of the feature space
- *Deeper* **representations** of text tend to improve performance
- **Qualitative analysis** is important
  - Evaluation
  - Interesting insights

Some of the future research challenges

+ Work closer with **domain experts**
  http://fludetector.cs.ucl.ac.uk
+ Better understanding of online media **biases**, *e.g.* demographics, external influence etc.
+ **Generalisation**, defining **limitations**, more rigorous **evaluation** frameworks
+ Methodological improvements
+ **Ethical concerns**

Acknowledgements

All **collaborators** (*in alphabetical order*)
in research mentioned today

Nikolaos Aletras (*Amazon*)
Yoram Bachrach (*Microsoft Research*)
Ingemar J. Cox (*UCL*)
Steve Crossan (*Google*)
Jens K. Geyti (*UCL*)
Andrew C. Miller (*Harvard University*)
Daniel Preotiuc-Pietro (*Penn*)
Christian Stefansen (*Google*)
Svitlana Volkova (*PNNL*)
Bin Zou (*UCL*)

Currently **funded by**

i-sense

Thank you!

Any questions?

Slides can be downloaded from
lampos.net/talks

@lampos | lampos.net

References

Bernstein. ***Language and social class*** (Br J Sociol, 1960)

Bouma. ***Normalized (pointwise) mutual information in collocation extraction*** (GSCL, 2009)

Duvenaud. ***Automatic Model Construction with Gaussian Processes*** (Ph.D. Thesis, Univ of Cambridge, 2014)

Labov. ***The Social Stratification of English in New York City*** (Cambridge Univ Press, 1972; 2006, 2nd ed.)

Lampos, Aletras, Geyti, Zou & Cox. ***Inferring the Socioeconomic Status of Social Media Users based on Behaviour and Language*** (ECIR, 2016)

Lampos, Miller, Crossan & Stefansen. ***Advances in nowcasting influenza-like illness rates using search query logs*** (Nature Sci Rep, 2015)

Lampos, Preotiuc-Pietro, Aletras & Cohn. ***Predicting and Characterising User Impact on Twitter*** (EACL, 2014)

Mikolov, Chen, Corrado & Dean. ***Efficient estimation of word representations in vector space*** (ICLR, 2013)

Preotiuc-Pietro, Lampos & Aletras. ***An analysis of the user occupational class through Twitter content*** (ACL, 2015)

Preotiuc-Pietro, Volkova, Lampos, Bachrach & Aletras. ***Studying User Income through Language, Behaviour and Affect in Social Media*** (PLoS ONE, 2015)

Rasmussen & Williams. ***Gaussian Processes for Machine Learning*** (MIT Press, 2006)

Volkova, Bachrach, Armstrong & Sharma. ***Inferring Latent User Properties from Texts Published in Social Media*** (AAAI, 2015)

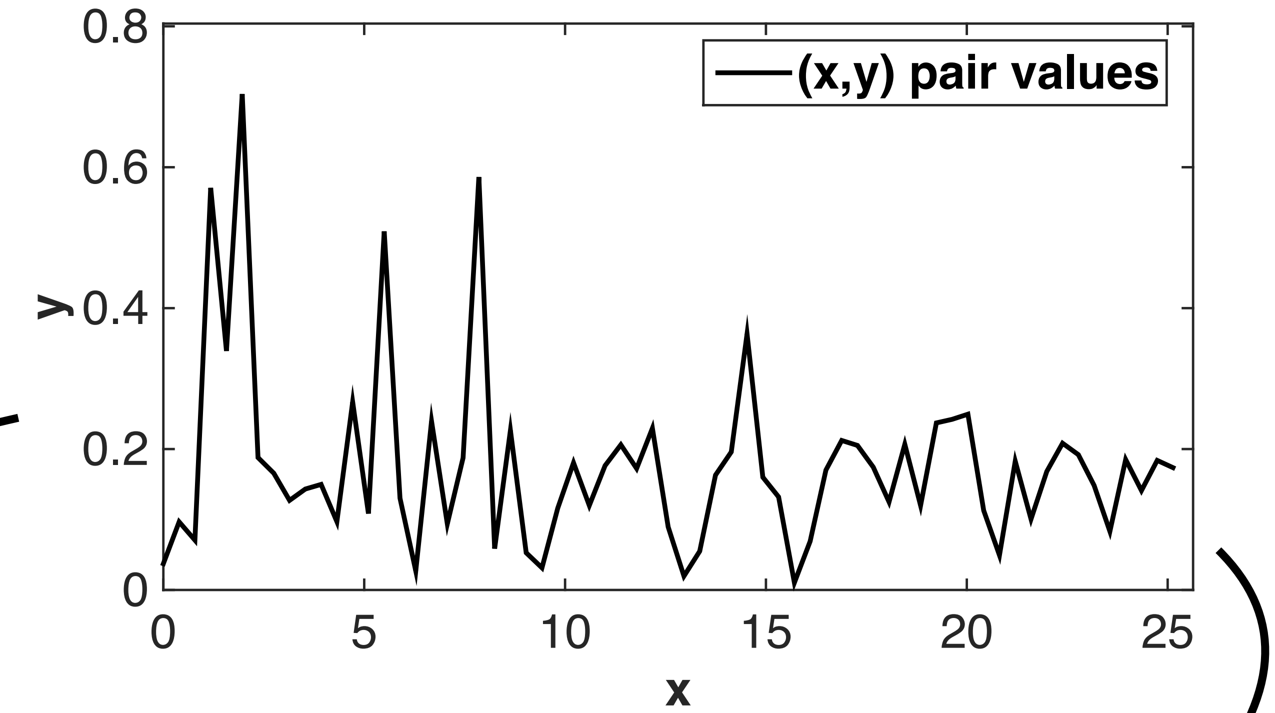
von Luxburg. ***A tutorial on spectral clustering*** (Stat Comput, 2007)

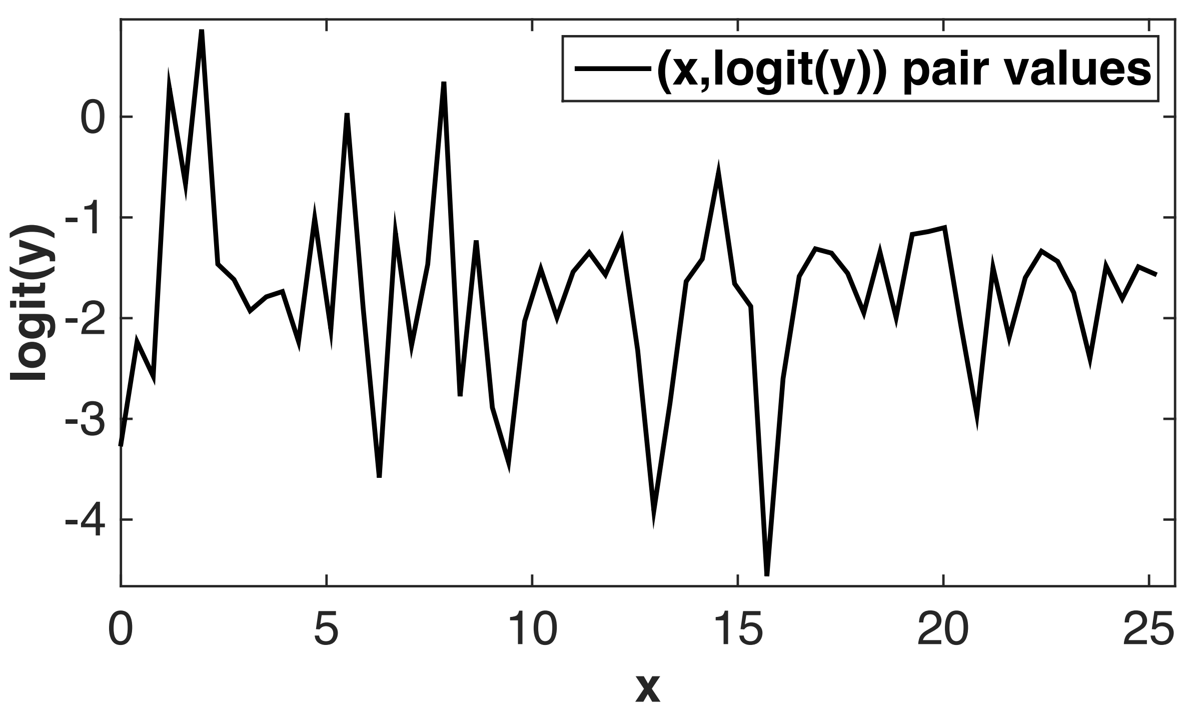
Zou & Hastie. ***Regularization and variable selection via the elastic net*** (J R Stat Doc Series B Stat Methodol, 2005)

Logit function

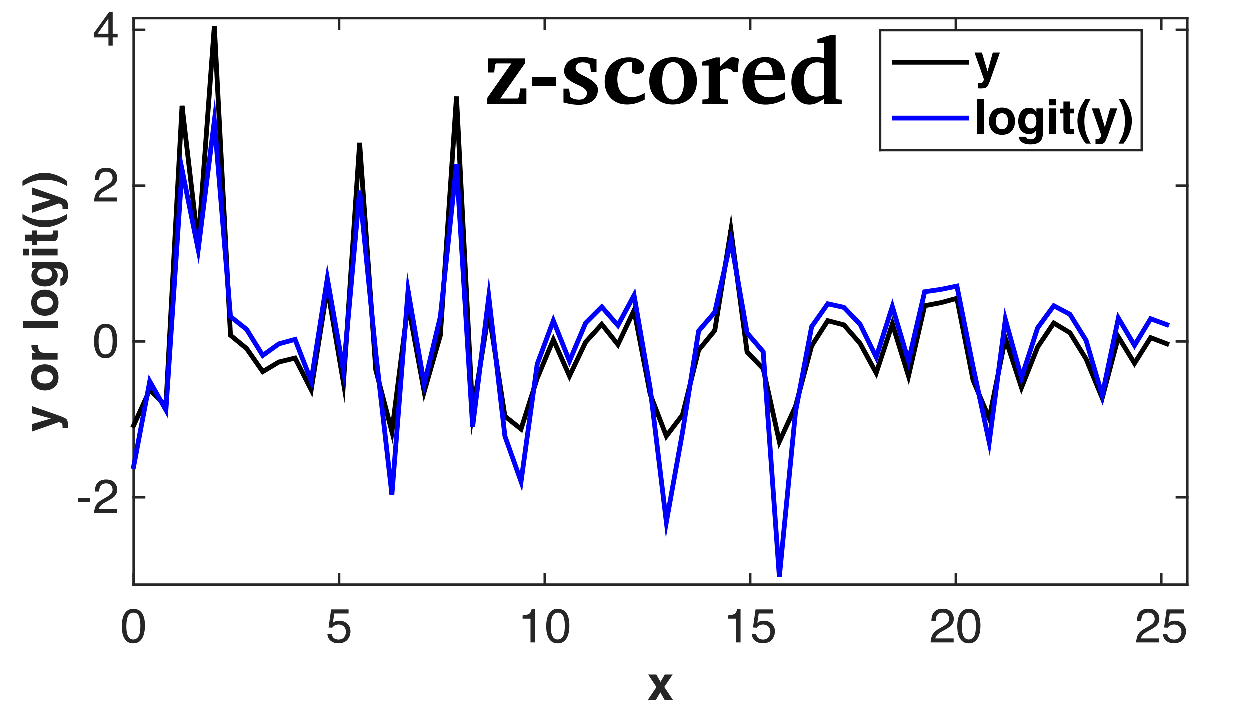
$$\boxed{\mathtt{logit}(a) = \mathtt{log}(a/(1-a))}$$

logit





z-scored



— intermediate values are ‘squashed’

— border values are ‘emphasised’

More information about Gaussian Processes

+ Book: "*Gaussian Processes for Machine Learning*"
  http://www.gaussianprocess.org/gpml/
+ Video-lecture: "*Gaussian Process Basics*"
  http://videolectures.net/gpip06_mackay_gpb/
+ Tutorial tailored to statistical NLP tasks: "*Gaussian Processes for Natural Language Processing*"
  http://people.eng.unimelb.edu.au/tcohn/tutorial.html
+ Software I — *GPML* for Octave or MATLAB
  http://www.gaussianprocess.org/gpml/code
+ Software II — *GPy* for Python
  http://sheffieldml.github.io/GPy/