Machine Learning and Finance: Prediction and Measurement

Sanmay Das

Washington University in St. Louis

MFM Summer School 2018

Statistics/Econometrics “vs.” Machine Learning

- Differences in communities, culture, practices
  - BUT: communities are learning a lot from each other and coming closer
  - Ideas come from both sides (e.g. bootstrap came from statistics to ML)
- (Supervised) ML has more of a focus on:
  - (Out-of-sample) prediction (vs. hypothesis testing or causal inference)
  - Algorithmic questions (esp. in high-dimensional problems)
  - Empirical (finite-sample) evaluation vs. asymptotics

Predictive Inference

- Clear use case: Solve a prediction problem and use the predictions for a meaningful task.
- E.g. "Prediction Policy Problems" (Kleinberg et al., 2015) example:
  - $\approx 500,000$ Medicare beneficiaries receive hip or knee replacements every year
  - Costs are both monetary and quality-of-life (first 6 months are particularly tough for recipients, but outcomes improve by 12 months)
  - However, 1.4% of recipients die in the month after surgery and 4.2% in months 1-12
  - These 4.2% are highly predictable using ML methods. For this population, having the surgery was probably a bad decision in terms of QOL.
  - Don't need to establish causality in order to improve outcomes in expectation (ethical issues can be a concern).

Goals

- Explain how machine learners view the world.
- Discuss what we're good at, and what the state of the art is, with a focus on things that might be of value to MFM scholars.
- Talk about a couple of examples in a bit of detail.

Goals

- Explain how machine learners view the world.
- Discuss what we're good at, and what the state of the art is, with a focus on things that might be of value to MFM scholars.
- Talk about a couple of examples in a bit of detail.
- ~~Discuss causality~~: Machine learning has a lot to learn
- ~~Talk about whether superintelligent AIs will take over the world, or about using ML to sell ad space on the web~~

Outline

- Basic framework: How (supervised) machine learners think
- State of practice in prediction problems
  - Algorithms: SVMs, ensemble methods, etc.
  - Optimization and design choices
- Prediction problems and applications in finance
- A measurement example: Text analysis

Outline

- **Basic framework: How (supervised) machine learners think**
- State of practice in prediction problems
  - Algorithms: SVMs, ensemble methods, etc.
  - Optimization and design choices
- Prediction problems and applications in finance
- A measurement example: Text analysis

The Supervised Learning Problem

- Unknown target function $f : \mathcal{X} \to \mathcal{Y}$
- Training data $\mathcal{D} : (\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots (\mathbf{x}_n, y_n)$ where $y_i = f(\mathbf{x}_i)$ (possibly noisy).
- Want to learn h "close to" f.
- Two central questions:
  - How do we learn h?
  - What can we say about how close h is to f?

(Note: My development and notation here follows that of Abu-Mostafa, Magdon-Ismail, and Lin (2012)).

Generalization Error

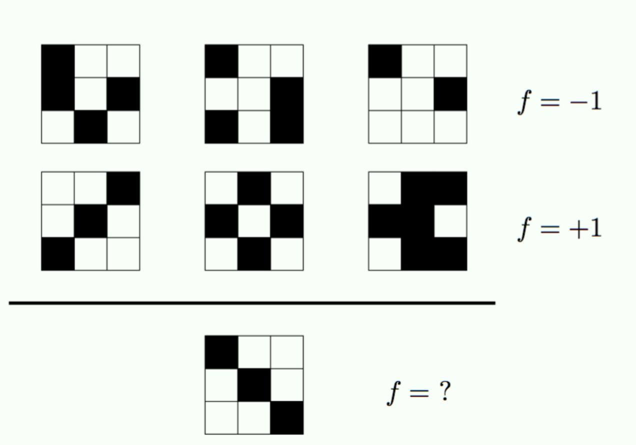
- Standard for closeness that we care about
  $E_{\text{out}}(h) = \Pr[h(\mathbf{x}) \neq f(\mathbf{x})]$, where the probability is based on the sampling distribution on $\mathcal{X}$.
- In practice, we estimate E_{out} by evaluating on a (held-out) *test set*.
- There are a ton of interesting problems when the sampling distribution for test data is not the same as that for $\mathcal{D}$.

How Do We Learn f?

- Pick a *hypothesis set* $\mathcal{H} = \{h_1, h_2, \ldots, \}$
- Use a *learning algorithm* to select a hypothesis from $\mathcal{H}$ on the basis of $\mathcal{D}$.
- The choice of $\mathcal{H}$ and the learning algorithm are intertwined

How Do We Learn *f*?

- Pick a *hypothesis set* $\mathcal{H} = \{h_1, h_2, \ldots,\}$
- Use a *learning algorithm* to select a hypothesis from $\mathcal{H}$ on the basis of $\mathcal{D}$.
- The choice of $\mathcal{H}$ and the learning algorithm are intertwined
- No free lunch in machine learning

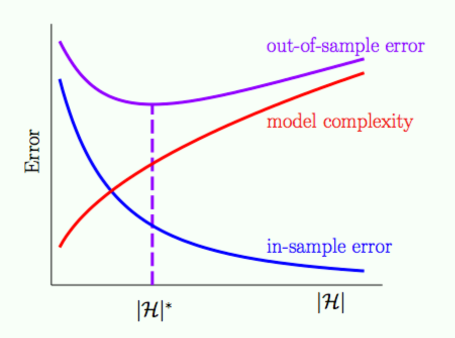


Choosing h from $\mathcal{H}$

- First thought: Minimize $E_{\text{in}}(g) = \frac{1}{n}\sum_{i=1}^{n}[h(\mathbf{x}_i) \neq f(\mathbf{x}_i)]$
- Many algorithms can be thought of within this broad framework.
  - Linear regression: Find a weight vector $\mathbf{w}$ that minimizes $E_{\text{in}}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^T\mathbf{x}_i - f(\mathbf{x}_i))^2$
  - Logistic regression: Find a linear function that minimizes $E_{\text{in}}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\ln(1 + e^{-y_i\mathbf{w}^T\mathbf{x}_i})$
  - Decision trees: Find the tree that directly minimizes the above. Problem: Computationally intractable, so we use greedy heuristics

Minimizing E_{out}

- But E_{in} is not really our objective. Vapnik-Chervonenkis and PAC theory tell us (roughly) that $E_{\text{out}}(g) \leq E_{\text{in}}(g) + \Omega(\mathcal{H})$ where $\Omega(\mathcal{H})$ penalizes complexity of the hypothesis class. Gives us two objectives:
  - Control hypothesis-class complexity
  - Minimize E_{in}



(from (Abu-Mostafa, Magdon-Ismail, and Lin, 2012))

The Central Problems

- There are deep relationships between the stability and variance of a learning algorithm, hypothesis complexity, and generalization ability.
- Bigger data $\rightarrow$ more complex hypothesis spaces can generalize better.
- Different ML algorithms arise from different choices related to two questions:
  - What $\mathcal{H}$ to search
  - What and how to optimize in the search process

Outline

- Basic framework: How (supervised) machine learners think
- **State of practice in prediction problems**
  - **Algorithms: SVMs, ensemble methods, etc.**
  - Optimization and design choices
- Prediction problems and applications in finance
- A measurement example: Text analysis

Linear Models and SVMs

- Regularized logistic regression: Minimize
  $E_{\text{in}}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} \ln(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i}) + \lambda \mathbf{w}^T \mathbf{w}$.
  - Often used with text data.
  - In general performs better than Naive Bayes.

Linear Models and SVMs

- Regularized logistic regression: Minimize $E_{\text{in}}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} \ln(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i}) + \lambda \mathbf{w}^T \mathbf{w}$.
  - Often used with text data.
  - In general performs better than Naive Bayes.
- Support vector machines: Minimize $E_{\text{in}}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(1 - y(\mathbf{w}^T \mathbf{x}_i)_+) + \lambda \mathbf{w}^T \mathbf{w}$ where $(z)_+ \equiv \max(z, 0)$ is the *hinge loss*.
  - Same thing as maximizing the margin. The dual QP has a nice interpretation in terms of "support vectors" (points at or within the margin).
  - When combined with the kernel trick, SVMs give a natural and powerful regularized non-linear classifier.
  - Try quadratic and RBF kernels if needed.

Decision Trees

- Flexible and rich hypothesis space. Easy to express trees in terms of a combination of rules.
- Impossible to find the optimal tree, so we use greedy heuristics to construct the tree.
- High variance classifier
  - Because of the greedy nature, a small change can have huge cascading effect, leading to totally different trees.
- Frequently used as a building block for ensemble methods.

Random Forests

- Construct *k* decision trees on bootstrapped replicates of the training set.
- Use random feature selection on the nodes (typically $\sqrt{\text{\# features}}$) to decorrelate trees further.
- The ensemble votes on any new example.

Random Forests

- Construct k decision trees on bootstrapped replicates of the training set.
- Use random feature selection on the nodes (typically $\sqrt{\text{\# features}}$) to decorrelate trees further.
- The ensemble votes on any new example.
- Off-the-shelf state of the art performance.
- Out-of-bag error provides a great estimate of out-of-sample error for free.
- Implementations come along with a feature scoring method that gets used a lot (measuring decrease in accuracy when permuting a feature in all OOB samples)

Boosting

- Arose as the answer to a theoretical question: Given a black-box weak learner (slightly better than chance), can we “boost” it into a strong learner.
- Key idea: Each new weak learner is trained on a distribution of examples modified to give more weight to those the existing ensemble has most trouble with.

Boosting

- Arose as the answer to a theoretical question: Given a black-box weak learner (slightly better than chance), can we “boost” it into a strong learner.
- Key idea: Each new weak learner is trained on a distribution of examples modified to give more weight to those the existing ensemble has most trouble with.
- A variant called *gradient boosting* is also state-of-the-art.
- Typically use decision stumps or short decision trees as the weak learners.
- If you need a powerful ML algorithm, try random forests and/or gradient boosting first.

Deep Learning

- (Massively) multilayer neural networks, typically trained using GPUs with huge training datasets.
- Optimization, structure, and regularization are all arts that are slowly becoming more “off the shelf” but aren’t there yet.
- It appears they learn interesting intermediate representations.
- Have been hugely influential in computer vision and NLP.
- Why they work is still a major theoretical puzzle.

Outline

- Basic framework: How (supervised) machine learners think
- **State of practice in prediction problems**
  - Algorithms: SVMs, ensemble methods, etc.
  - **Optimization and design choices**
- Prediction problems and applications in finance
- A measurement example: Text analysis

Model Selection

- How strong a regularizer? How many trees?
- Classic answer: Cross-validation and grid search over the space of parameters

Model Selection

- How strong a regularizer? How many trees?
- Classic answer: Cross-validation and grid search over the space of parameters
- Taking over: Bayesian Optimization
  - Use a Gaussian Process prior over the value of the function being optimized (say error) and iteratively evaluate at different hyperparameter values (Shahriari et al., 2016; Snoek, Larochelle, and Adams, 2012; Osborne, Garnett, and Roberts, 2009).
  - Spearmint (`https://github.com/HIPS/Spearmint`) is a well-known package for BO.

Outline

- Basic framework: How (supervised) machine learners think
- State of practice in prediction problems
  - Algorithms: SVMs, ensemble methods, etc.
  - Optimization and design choices
- **Prediction problems and applications in finance**
- A measurement example: Text analysis

Some Trends in Current Research

- Can predictive analytics provide business or regulatory value and insight?
  - Khandani, Kim, and Lo (2010) show that banks can use ML on combined credit bureau and monthly transaction data to manage credit lines more profitably.

Some Trends in Current Research

- Can predictive analytics provide business or regulatory value and insight?
  - Khandani, Kim, and Lo (2010) show that banks can use ML on combined credit bureau and monthly transaction data to manage credit lines more profitably.
  - Kong and Saar-Tsechansky (2014) discuss active learning in order to identify tax returns that would be particularly informative to audit.

Some Trends in Current Research

- Can predictive analytics provide business or regulatory value and insight?
  - Khandani, Kim, and Lo (2010) show that banks can use ML on combined credit bureau and monthly transaction data to manage credit lines more profitably.
  - Kong and Saar-Tsechansky (2014) discuss active learning in order to identify tax returns that would be particularly informative to audit.
  - Giesecke, Sirignano, and Sadhwani (2016) use deep learning to estimate state transitions for mortgages.

Some Trends in Current Research

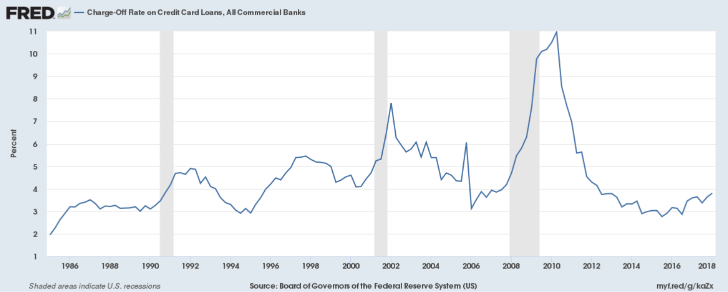
- Can predictive analytics provide business or regulatory value and insight?
  - Khandani, Kim, and Lo (2010) show that banks can use ML on combined credit bureau and monthly transaction data to manage credit lines more profitably.
  - Kong and Saar-Tsechansky (2014) discuss active learning in order to identify tax returns that would be particularly informative to audit.
  - Giesecke, Sirignano, and Sadhwani (2016) use deep learning to estimate state transitions for mortgages.
  - Butaru et al. (2016) learn models for predicting CC delinquency across several large banks (and compare risk management practices across banks).

Some Trends in Current Research

- Can predictive analytics provide business or regulatory value and insight?
  - Khandani, Kim, and Lo (2010) show that banks can use ML on combined credit bureau and monthly transaction data to manage credit lines more profitably.
  - Kong and Saar-Tsechansky (2014) discuss active learning in order to identify tax returns that would be particularly informative to audit.
  - Giesecke, Sirignano, and Sadhwani (2016) use deep learning to estimate state transitions for mortgages.
  - Butaru et al. (2016) learn models for predicting CC delinquency across several large banks (and compare risk management practices across banks).
- Many of these have a “horse race” element.
- One major issue is the Lucas Critique. Will get back to this...

Predicting Credit Card Delinquency

- Rapidly growing consumer credit market.
- High charge-off rate and severe delinquency rate during the financial crisis highlight the need for robust and improved out-of-sample prediction models for risk management



- With good predictions of delinquency, banks can actively manage credit to manage their exposure (Khandani, Kim, and Lo, 2010).

Project Overview

- Immediate goal: Support risk-based supervision
  - Compare performance of ML methods vs. “traditional” logistic regression.
  - Compare risk management across banks.
- Future goal: Measure systemic risk
  - Combine data across institutions.
  - Generate aggregate forecast models.
  - Use in stress testing?

Data Sources

- Credit card data
  - Detailed account-level characteristics (days past due, balance, utilization, FICO, etc.)
    - Cannot link the accounts across individuals
  - Collected by a major US regulator
  - Entire credit card portfolios of 6 large banks
  - Monthly data from January 2008
- Attribute data
  - Detailed borrower characteristics from a credit bureau (linked by account)
  - Quarterly starting in 2009
- Macroeconomic variables
  - Collected from various sources (linked by account ZIP)
  - Employment data, HPI, average wages, average hours, etc.
- In total, many TB of raw data

Sample Description

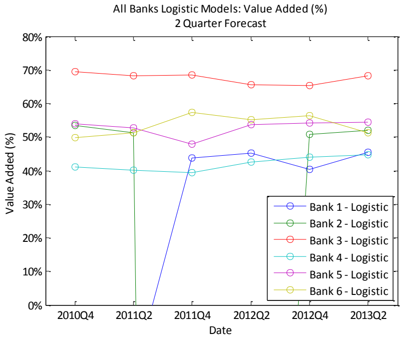
- Approx. 1% of each bank's CC portfolio.
- Yields between 90K and 1M observations per period per bank.
- Portfolio size varies over time (some grow, some decline) and the sample size is representative of the true portfolio.
- Substantial heterogeneity in the time series and cross-sectional distribution of delinquency rates.

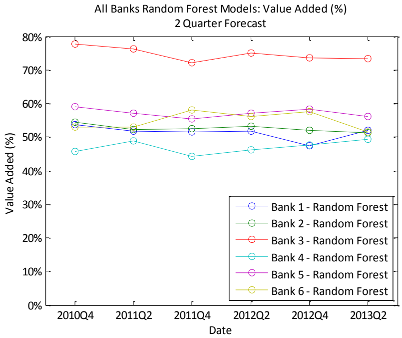
Modeling Dominates Credit Scores

Value-Added Analysis

- Compute a hypothetical cost-saving per customer, assuming the bank would cut lines of those predicted to default. Depends on:
  - Classification accuracy (the whole confusion matrix): **Quality of the model**
  - Run-up ratio (how much customers will increase their balance after time of prediction)
  - Profitability margin: Expected future cash-flows (opportunity cost of misclassification)
- Based on the framework of (Khandani, Kim, and Lo, 2010).

Value-Added Comparison





Most Important Attributes For Prediction

- Days past due
- Behavioral score
- Refreshed credit score
- Ratio of actual to minimum payents
- 1 month change in monthly utilization
- Payment equal to minimum in last 3 months

Further Insights

- Banks vary greatly in their risk management practices and effectiveness
  - 3 of the banks are very effective at cutting lines of accounts predicted to become delinquent
- Risk factors affecting delinquency vary across banks
- Macroeconomic variables affect different banks differently
  - Were more important factors in the model in 2012 and earlier than later.

Are Models For Predicting Failure Destined to Fail?

- So what if you can predict these things well in hindsight?
- The basic problem: Agent incentives and Goodhart's law (Wikipedia's formulation: "When a measure becomes a target, it ceases to be a good measure").
- If the ML model (or any risk model (Danielsson, 2002)) is part of the regulatory system, it will break down, because the data generating process will change.

Are Models For Predicting Failure Destined to Fail?

- So what if you can predict these things well in hindsight?
- The basic problem: Agent incentives and Goodhart's law (Wikipedia's formulation: "When a measure becomes a target, it ceases to be a good measure").
- If the ML model (or any risk model (Danielsson, 2002)) is part of the regulatory system, it will break down, because the data generating process will change.
- Rajan, Seru, and Vig (2015) study this empirically in the context of subprime mortgage loans originated from 1997–2006.
    - Securitization changes lenders' incentives and behavior. Over time, borrowers with similar "hard" information in the data become worse risks because they are worse along the "soft" information dimension (unobservable to investors).
    - "[When] incentive effects lead to a change in the underlying regime, the coefficients from a statistical model estimated on past data have no validity going forward, regardless of how sophisticated the model is or how well it fits the prior data."

Possible Approach: Don't Share the Model

New York Times, "Not communicating to banks exactly what they need to do to get their bankruptcy plans to pass muster is frustrating, confusing and part of the plan."

TheUpshot

FOLLOW US:
GET THE UPSHOT IN YOUR INBOX

RULES, RULES

How Regulators Mess With Bankers' Minds, and Why That's Good

Peter Eavis @petereavis APRIL 14, 2016

23

Bank regulators on Wednesday sent a message that big banks are still too big and too complex. They rejected special plans, called living wills, that the banks have to submit to show they can go through an orderly bankruptcy.

The thinking behind the regulators' call for living wills is that if a large bank crash is orderly, there will be no need to save it and no need for taxpayer bailouts.

Pretty straightforward, right? Not for the banks. The regulators deliberately did not communicate the exact things the banks needed to do for their plans to pass muster. In this way, they kept them on their toes — and treating powerful banks this way may end up playing a surprisingly important role in keeping the financial regulation effective over time.

Issues

- The regulatory model will still affect the system even if it doesn't explicitly change agent behavior
- Transparency can be important in many contexts
- Can the model be backed out from results?
- Not insurmountable in many applications

Adoption of ML

- These problems are really not specific to using ML techniques. Banks and regulators already use models to predict the future. ML typically does it better.
- Not meant to replace causal models where needed.
- In practice, have to re-train models regularly to deal with nonstationarities.

Adoption of ML

- These problems are really not specific to using ML techniques. Banks and regulators already use models to predict the future. ML typically does it better.
- Not meant to replace causal models where needed.
- In practice, have to re-train models regularly to deal with nonstationarities.
- The myth of interpretability
  - Are logistic regression coefficients really meaningful?
  - Single decision trees are equivalent to sets of rules, but decision tree learners are inherently unstable. How meaningful is the tree?
  - Often used synonymously with “manually sanity checkable”
  - There are ways to sanity check more complex models as well (random forest feature scores, checking which features drive changed in NN output, etc.)

Outline

- Basic framework: How (supervised) machine learners think
- State of practice in prediction problems
  - Algorithms: SVMs, ensemble methods, etc.
  - Optimization and design choices
- Prediction problems and applications in finance
- **A measurement example: Text analysis**

Text Analysis

- Big idea: Ton of information in text that hasn't been quantified and analyzed, but could be usable.
- Examples:
  - Gentzkow and Shapiro (2010) measure ideological slant by identifying "phrases" used much more frequently by Democrats or Republicans in the congressional record, and use this to estimate the drivers of media slant.

Text Analysis

- Big idea: Ton of information in text that hasn't been quantified and analyzed, but could be usable.
- Examples:
  - Gentzkow and Shapiro (2010) measure ideological slant by identifying "phrases" used much more frequently by Democrats or Republicans in the congressional record, and use this to estimate the drivers of media slant.
  - Tetlock, Saar-Tsechansky, and Macskassy (2008) use fraction of negative words in firm-specific news stories to predict accounting earnings and stock returns.

Text Analysis

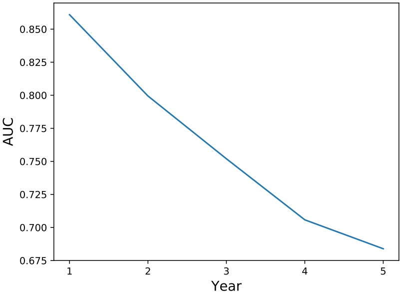
- Big idea: Ton of information in text that hasn't been quantified and analyzed, but could be usable.
- Examples:
  - Gentzkow and Shapiro (2010) measure ideological slant by identifying "phrases" used much more frequently by Democrats or Republicans in the congressional record, and use this to estimate the drivers of media slant.
  - Tetlock, Saar-Tsechansky, and Macskassy (2008) use fraction of negative words in firm-specific news stories to predict accounting earnings and stock returns.
  - Manela and Moreira (2016) construct a measure of aggregate uncertainty "News Implied Volatility" (NVIX). ML (SVR) to map from bag-of-bigrams representations of WSJ articles to VIX in the observed period, and can then generalize to before VIX existed.

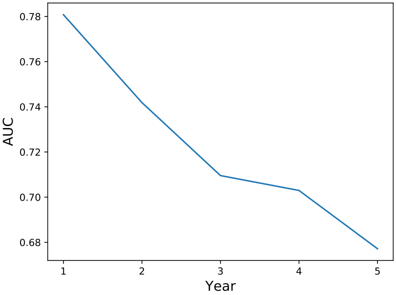
A Cautionary Tale: Political Ideology Measurement

- Idea: Measure ideology by building a classifier of partisanship (Yan et al., 2018)
- Data:
  - Congressional Record and press releases (politicians)
  - Salon and Townhall (media)
  - Conservapedia, RationalWiki, Wikipedia (the crowd)
- Algorithms:
  - Logistic regression on n-grams (Bag-of-bigrams, TFIDF, feature hashing)
  - Recursive autoencoder (RAE) (Socher et al., 2011).

Generalizing Across Time

Out-of-time predictions for Salon-Townhall (left) and the Congressional Record (right) when trained on two years of data and tested going forward





Generalizing Across Datasets

Testing and Averaging By Year

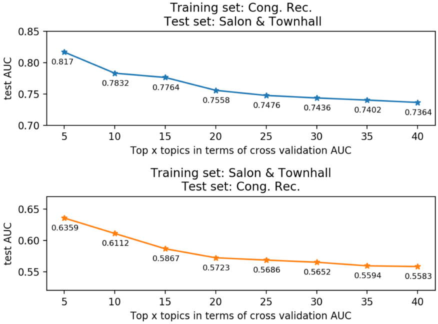
Training \ Test	Cong. Rec.	Salon & Townhall	Wikis
Cong. Rec.	0.82 (LR) 0.81 (RAE)	0.64 (LR) 0.59(RAE)	0.51 (LR) 0.47 (RAE)
Salon & Townhall	0.56 (LR) 0.54 (RAE)	0.91 (LR) 0.90 (RAE)	0.50 (LR) 0.55 (RAE)
Wikis	0.50 (LR) 0.47 (RAE)	0.54 (LR) 0.57 (RAE)	0.82 (LR) 0.82 (RAE)

A Silver Lining?

- Predictability may be a function of topic
- Learn a topic model jointly on CR and ST
- Learn individual classifiers for texts "hard classified" to each of 40 topics

A Silver Lining?

- Predictability may be a function of topic
- Learn a topic model jointly on CR and ST
- Learn individual classifiers for texts “hard classified” to each of 40 topics



High AUC Topics

5 topics with the highest cross validation AUC

Topic 28
(CV AUC: 0.960)
republican parti 0.007
social secur 0.006
tax cut 0.005
american peopl 0.003
wall street 0.003
great depress 0.002
liber democrat 0.002
econom polici 0.002
bill clinton 0.002
georg bush 0.002
…

Topic 36
(CV AUC: 0.950)
health care 0.021
health insur 0.006
small busi 0.006
incom tax 0.003
tax rate 0.002
tax cut 0.002
insur compani 0.002
balanc budget 0.002
million american 0.002
care system 0.002
…

Topic 11
(CV AUC: 0.945)
pro lif 0.005
onlin edit 0.004
richard dawkin 0.003
stem cell 0.002
plan parenthood 0.002
scientif medic 0.001
abort time 0.001
theori evolut 0.001
cell research 0.001
unit state 0.001
…

Topic 29
(CV AUC: 0.928)
global warm 0.006
climat chang 0.006
unit state 0.003
oil ga 0.002
natur ga 0.002
oil compani 0.002
carbon dioxid 0.002
renew energi 0.002
nuclear power 0.001
fossil fuel 0.001
…

Topic 6
(CV AUC: 0.919)
civil war 0.006
war iraq 0.003
saddam hussein 0.002
liber bia 0.002
de gaull 0.002
foreign polici 0.002
bin laden 0.002
war terror 0.002
al qaeda 0.001
middl east 0.001
…

- 28 Politics & the economy
- 36 Health care, insurance, and taxes
- 11 Evolutionary science, medicine
- 29 Climate change
- 6 Foreign policy in the middle east

Low AUC Topics

5 topics with the lowest cross validation AUC

Topic 8
(CV AUC: 0.737)
pro choic 0.004
first amend 0.003
time limit 0.003
plan parenthood 0.002
democrat nomin 0.002
anti abort 0.002
daili show 0.002
use describ 0.002
vote bill 0.002
amend offer 0.002
…

Topic 2
(CV AUC: 0.734)
balanc time 0.004
mr speaker 0.004
breast cancer 0.003
urg colleagu 0.003
back balanc 0.002
yield back 0.002
support homeopathi 0.002
nativ american 0.002
reserv balanc 0.002
sexual assault 0.002
…

Topic 37
(CV AUC: 0.728)
talk point 0.002
rush limbaugh 0.001
hous press 0.001
liber conserv 0.001
hate group 0.001
hate speech 0.001
anti govern 0.001
club growth 0.001
donald trump 0.001
white male 0.001
…

Topic 22
(CV AUC: 0.714)
new age 0.004
american polit 0.003
talk radio 0.003
parti candid 0.003
conserv movement 0.002
ayn rand 0.002
welfar state 0.002
thoma jefferson 0.002
southern state 0.002
governor new 0.001
…

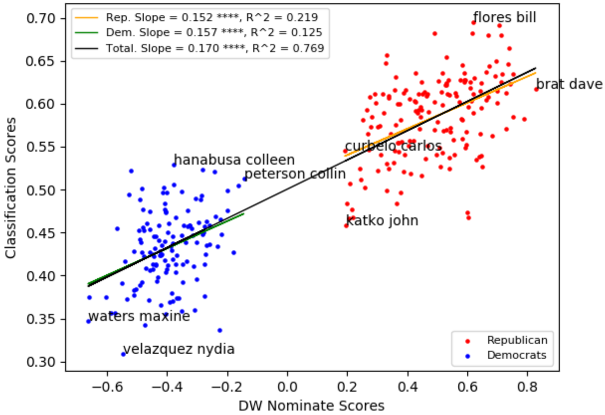
Topic 3
(CV AUC: 0.683)
sarah palin 0.004
look like 0.002
new world 0.002
year ago 0.002
unit state 0.001
world order 0.001
right activist 0.001
human be 0.001
york time 0.001
mani peopl 0.001
…

8	Abortion
2	Procedural phrases
11	Opinions on media
29	Political philosophy
6	Specific people

Direction of Movement

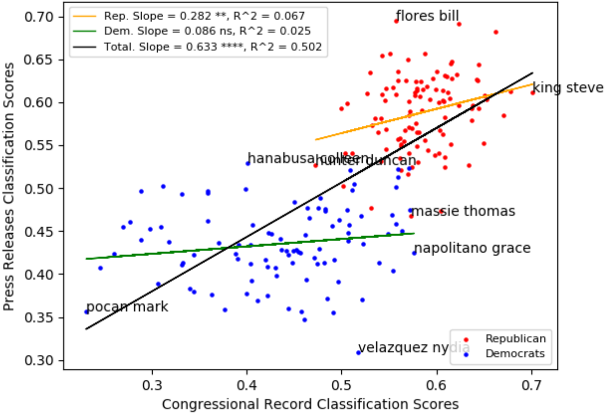
- Lots of suggestive evidence that language moves *from* Congress *to* the media – better performance when using CR as base domain
- Also use our data to show that “new mutual phrases” tend to first appear in the CR and then in ST a few days later
- Consistent with Gentzkow and Shapiro story

Predicting Ideology from Press Releases



Predicting party is easy; predicting ideology *within* party is hard!

Congressional Record vs. Press Releases



Virtually no connection (within party) on ideology as expressed in Congress as opposed to in press releases!

Takeaways

- Text-based methods can be a fascinating complement to existing ways of measuring ideology, but...
- Must be very careful in generalizing across types of text, especially for short texts.
- Topics and temporal change of language can be a big issue.
- Perhaps combining structural and textual information will help?
  - But even semantically rich embeddings can be problematic in the political context (e.g. “private accounts” vs. “personal accounts” for social security).

In Conclusion

- Machine learning is very good with:
  - Out-of-sample prediction.
  - Big data.
  - Nonlinear models.
  - Designing new measures
  - “Not immediately quantitative” data (text, images)
- When these strengths speak well to a problem, it’s silly not to use the best techniques we have available!

References I

Abu-Mostafa, Yaser S, Malik Magdon-Ismail, and Hsuan-Tien Lin (2012). *Learning From Data*. AMLBook.

Butaru, Florentin, Qingqing Chen, Brian J. Clark, Sanmay Das, Andrew W. Lo, and Akhtar R. Siddique (2016). "Risk and Risk Management in the Credit Card Industry". In: *Journal of Banking and Finance* 72, pp. 218–239.

Danielsson, Jon (2002). "The emperor has no clothes: Limits to risk modelling". In: *Journal of Banking & Finance* 26.7, pp. 1273–1296.

Gentzkow, Matthew and Jesse M Shapiro (2010). "What drives media slant? Evidence from US daily newspapers". In: *Econometrica* 78.1, pp. 35–71.

Giesecke, Kay, J Sirignano, and A Sadhwani (2016). *Deep Learning for Mortgage Risk*. Tech. rep. Working Paper, Stanford University.

References II

Khandani, Amir E, Adlar J Kim, and Andrew W Lo (2010). "Consumer credit-risk models via machine-learning algorithms". In: *Journal of Banking & Finance* 34.11, pp. 2767–2787.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer (2015). "Prediction policy problems". In: *The American Economic Review* 105.5, pp. 491–495.

Kong, Danxia and Maytal Saar-Tsechansky (2014). "Collaborative information acquisition for data-driven decisions". In: *Machine Learning* 95.1, pp. 71–86.

Manela, Asaf and Alan Moreira (2016). "News implied volatility and disaster concerns". In: *Journal of Financial Economics*. Forthcoming.

Osborne, Michael A, Roman Garnett, and Stephen J Roberts (2009). "Gaussian processes for global optimization". In: *3rd International Conference on Learning and Intelligent Optimization*, pp. 1–15.

References III

Rajan, Uday, Amit Seru, and Vikrant Vig (2015). "The failure of models that predict failure: Distance, incentives, and defaults". In: *Journal of Financial Economics* 115.2, pp. 237–260.

Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando de Freitas (2016). "Taking the human out of the loop: A review of Bayesian optimization". In: *Proceedings of the IEEE* 104.1, pp. 148–175.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams (2012). "Practical Bayesian Optimization of Machine Learning Algorithms". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, pp. 2951–2959.

References IV

- Socher, Richard, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning (2011). "Semi-supervised recursive autoencoders for predicting sentiment distributions". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 151–161.
- Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More than words: Quantifying language to measure firms' fundamentals". In: *The Journal of Finance* 63.3, pp. 1437–1467.
- Yan, Hao, Sanmay Das, Allen Lavoie, Sirui Li, and Betsy Sinclair (2018). "The Congressional Classification Challenge: Domain Specificity and Partisan Intensity". Working paper.