PROFESSOR BENJAMIN NOBLE (UCSD)

HOW TO TRAIN YOUR MODEL

The Washington Post

OPINIONS Editorials Columns Guest opinions Cartoons

Opinion | Get real and read some history. The past was worse.

By Jennifer Rubin
Columnist | + Follow
December 31, 2023 at 7:45 a.m. EST

Nostalgia is a powerful political tool. Wielding nostalgia for a bygone era — one that is invariably mischaracterized — is a favorite weapon for fascist movements (Make America Great *Again*), harking back to a time before their nation was "polluted" by malign forces. In the United States, such nostalgia none-too-subtlety appeals to white Christian nationalism. Even in a more benign form (e.g., "*Politics didn't used to be so mean,*" "*Remember the days of bipartisanship?*") plays on faulty memories. If you really go back to study U.S. history, you would find two things: The past was worse, and conflict has always been the norm.

Giorgia Meloni may be no fascist. But she evokes grim memories of Italy's past

John Foot

An election win for her Brothers of Italy would be a threat to democracy across Europe

Sun 14 Aug 2022 04.30 EDT

Share

via the Guardian

WHAT IS “NOSTALGIA?”

- Collective nostalgia: a “predominately positive emotion that is associated with recalling memories of important or momentous events, usually experienced with close others” (Lammers and Baldwin 2018; Müller and Proksch 2023).

British Journal of Political Science (2023), page 1 of 13
doi:10.1017/S0007123423000571

British Journal of Political Science

LETTER

Nostalgia in European Party Politics: A Text-Based Measurement Approach

Stefan Müller[1] and Sven-Oliver Proksch[2]

[1]School of Politics and International Relations, University College Dublin, Dublin, Ireland and [2]Cologne Center for Comparative Politics, Institute of Political Science and European Affairs, University of Cologne, Cologne, Germany
Corresponding author: Stefan Müller; Email: stefan.mueller@ucd.ie

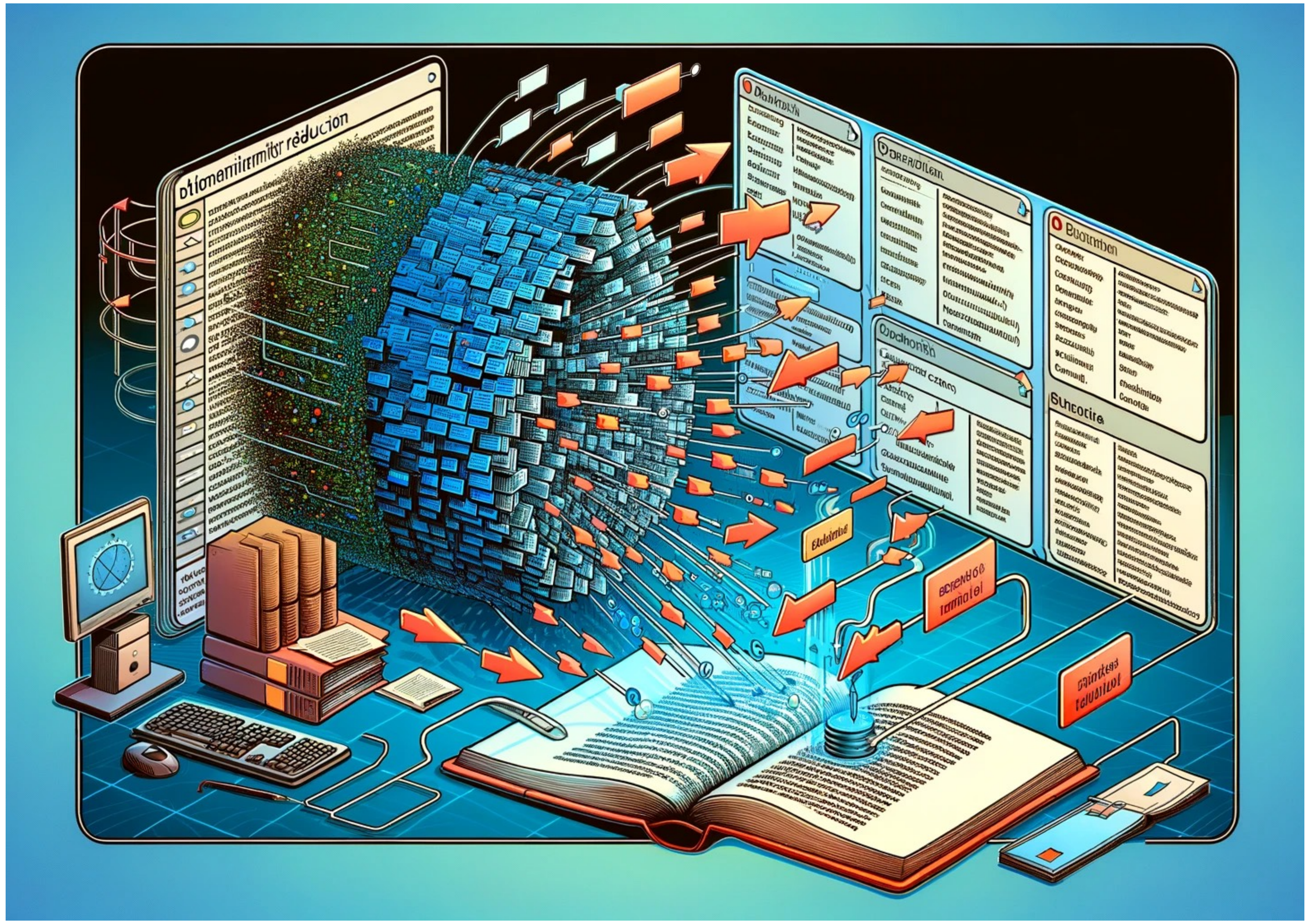
(Received 19 October 2021; revised 20 July 2023; accepted 10 September 2023)

Abstract
Traditional research on political parties pays little attention to the temporal focus of communication. It usually concentrates on promises, issue attention, and policy positions. This lack of scholarly attention is surprising, given that voters respond to nostalgic rhetoric and may even adjust issue positions when policy is framed in nostalgic terms. This article presents a novel dataset, PolNos, which contains six text-based measures of nostalgic rhetoric in 1,648 party manifestos across 24 European democracies from 1946 to 2018. The measures combine dictionaries, word embeddings, sentiment approaches, and supervised machine learning. Our analysis yields a consistent result: nostalgia is most prevalent in manifestos of culturally conservative parties, notably Christian democratic, nationalist, and radical right parties. However, substantial variation remains regarding regional differences and whether nostalgia concerns the economy or culture. We discuss the implications and use of our dataset for studying political parties, party competition, and elections.

Keywords: nostalgia; party competition; elections; quantitative text analysis

RECALL: MAPPING W → Q

- Our goal in text analysis is to try to learn $f(W_i)$, a function that takes our representation of the text, **W**, and maps it to **Q.**
  - Keyword methods map **W** to **Q** using a pre-specified list of words (and weights).
  - But, in many cases, it is difficult (or impossible) to create a good dictionary.



via DALL-E

WHAT WOULD GO IN A DICTIONARY FOR “NOSTALGIA?”

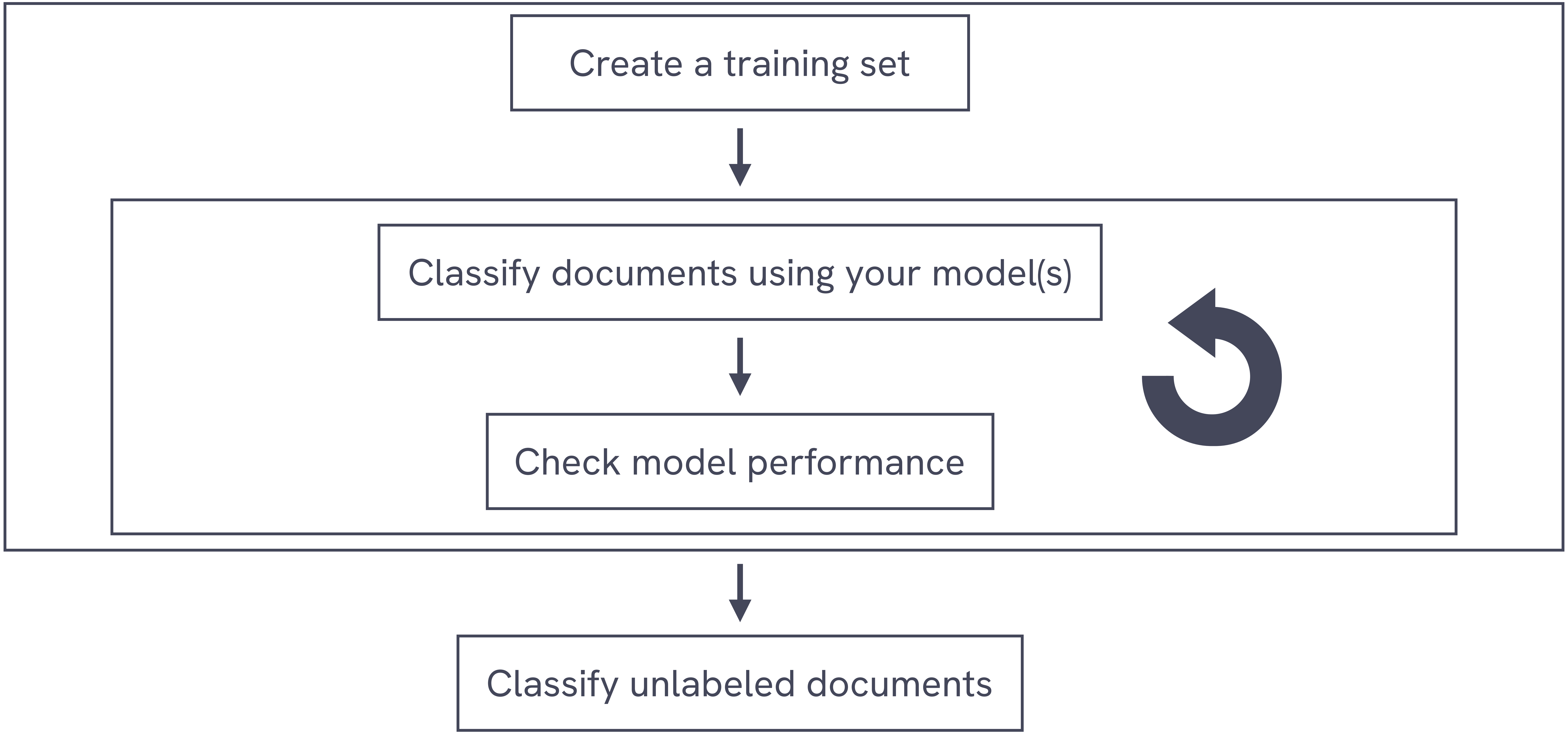
HUMANS ARE GOOD AT GUESSING Q

- Sometimes, it is easier for a human to recognize **Q** directly than define $f(W_i) \to Q$.
  - *"Only if we keep our cities alive and viable, if we impose the duties and guarantee the rights of citizens, if we preserve and enhance the treasures that nature and a great historical past we have left, we will ensure our future."*
  - *"In the past various changes to the day of polling were advocated as a spur to turnout."*

SUPERVISED MACHINE LEARNING

- **Manual processing** (e.g., reading, dictionaries): human does almost all the work (but it's a lot of work)
- **Unsupervised learning** (e.g., LDA): little human input, machine does the work (but better for discovery than finding what we're looking for).
- **Supervised learning**: a synthesis.
  - The human provides a **training set**, a sample of human-labeled **W-Q** pairs.
  - The model uses information in **W** to predict **Q**.
  - Together, we use the rules discovered by the model to predict **Q** for all unlabeled **W**.

STEPS IN SUPERVISED CLASSIFICATION



CREATING A TRAINING SET

- Labeling allows you to incorporate substantive knowledge and context into downstream automated labeling.
  - …but labeling documents is actually very hard!
    - Consider: “In August, the ‘Blank Space’ artist [Taylor Swift] criticized President Donald Trump and the ongoing threat of Covid-19” (NBC News).

SUPPOSE YOU DECIDE TO HANDCODE SOME DOCUMENTS...

- If you are going to develop a codebook...
  - Your first draft will probably stink!
  - Revise and continue coding until you have resolved major ambiguities.

C Human Coding and Validation

C.1 Instructions for Coding of Nostalgic Rhetoric

In this research project, we aim to measure and explain nostalgia in parties' campaign communication. Lammers and Baldwin (2018: 599) define nostalgia as "a predominately positive emotion that is associated with recalling memories of important or momentous events, usually experienced with close others."

You will be asked to annotate sentences from party manifestos. The coding proceeds in five steps:

1. You read the 'target' sentence (highlighted in bold), which should be coded along several dimensions. For a better context, we also provide the sentences before and after the target sentence. Make sure to read the context too.
2. You code whether a sentence is nostalgic or not.
3. For non-English text, code if the translation is inaccurate or very inaccurate. Leave the column blank if the sentence is understandable. Some sentences may be very short or so-called quasi sentences. Only code these sentences as inaccurate translations if they are incomprehensible after reading the context too (`text_pre` and `text_post`).
4. If the sentence is *not* nostalgic, proceed to the next sentence.

Müller and Proksch 2023

SUPPOSE YOU DECIDE TO HANDCODE SOME DOCUMENTS...

- If you are going to develop a codebook...
  - Your first draft will probably stink!
  - Revise and continue coding until you have resolved major ambiguities.
- You can also use an existing codebook.

- **1. Macroeconomics**
  - 100: General
    Description: Includes issues related to general domestic macroeconomic policy.
  - 101: Interest Rates
    Description: Includes issues related to inflation, cost of living, prices, and interest rates
  - 103: Unemployment Rate
    Description: Includes issues related to the unemployment rate, impact of unemployment
    See also: 502 training and retraining; 503 unemployment benefits.
  - 104: Monetary Policy
    Description: Includes issues related to the monetary policy, central bank, and the treasury
  - 105: National Budget
    Description: Issues related to public debt, budgeting, and efforts to reduce deficits
  - 107: Tax Code
    Description: Includes issues related to tax policy, the impact of taxes, and tax enforcement
    Note: Specific tax changes should be coded based upon the substantive area. For instance, airfare taxes should go in Air Travel (1003). General tax issues including changes to multiple substantive areas should be coded as 107.
  - 108: Industrial Policy
    Description: Includes issues related to manufacturing policy, industrial revitalization and growth
  - 110: Price Control
    Description: Includes issues related to wage or price control, emergency price controls
  - 199: Other
    Description: Includes issues related to other macroeconomics subtopics

Comparative Agendas Project

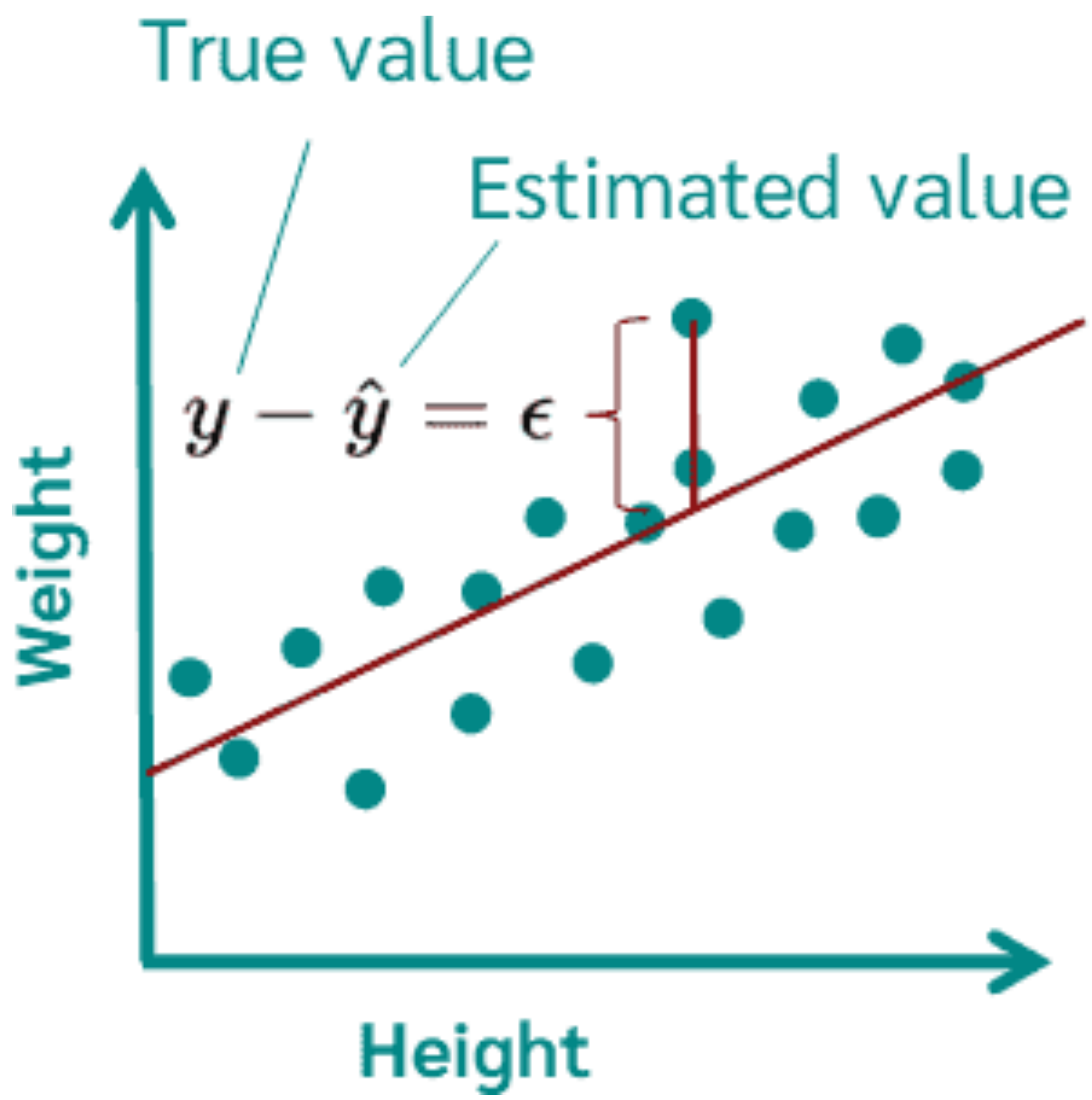
SOME CODING CONSIDERATIONS

- Who? And how many coders?
  - You, trained research assistants, crowdsourced workers.
- Which, and how many, documents?
  - A random/generalizable sample, consider the task.
- What about "found" or "cross domain" labels?
  - Caveats: unknown labeling process, non-random sample.

via DALL-E

TRAINING THE MODEL

- Recall: our goal is to use a model to approximate $f(W_i)$.
  - Not so different from linear regression: $y_i \sim \beta_0 + X_i\beta_1 + \epsilon$.



via datatab.net

A TOY EXAMPLE USING LINEAR REGRESSION

class reviews

```
# A tibble: 6 × 3
  doc_id text                                   positive_review
   <int> <chr>                                            <dbl>
1      1 This class is good.                                  1
2      2 Good class!                                          1
3      3 This class is bad.                                   0
4      4 Bad teacher.                                         0
5      5 A good teacher Not bad!                              1
6      6 Bad class and bad teacher. Not good.                 0
```

A TOY EXAMPLE USING LINEAR REGRESSION

review dfm

```
Document-feature matrix of: 6 documents, 4 features (37.50% sparse) and 1 docvar.
    features
docs class good bad teacher
   1     1    1   0       0
   2     1    1   0       0
   3     1    0   1       0
   4     0    0   1       1
   5     0    1   1       1
   6     1    1   2       1
```

A TOY EXAMPLE USING LINEAR REGRESSION

- "Ben is a good, not a bad, teacher!"
  - 0.47 + good x 0.6 - bad x 0.6 + teacher x 0.27 = 0.73

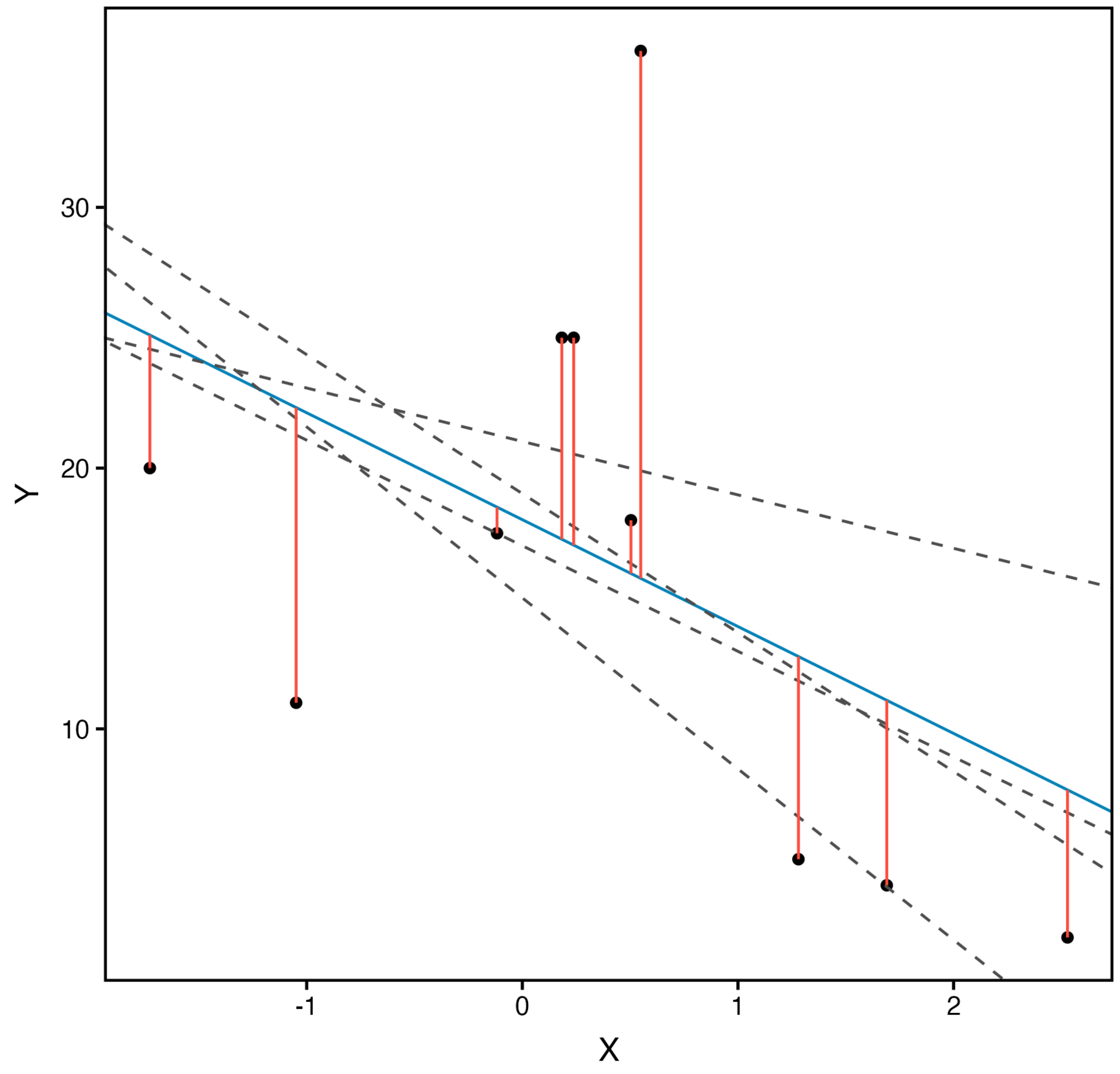
regression output

```
Call:
lm(formula = positive_review ~ good + bad + teacher, data = ct_df)

Residuals:
       1        2        3        4        5        6 
-0.06667 -0.06667  0.13333 -0.13333  0.26667 -0.13333 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4667     0.2404   1.941    0.192
good          0.6000     0.2309   2.598    0.122
bad          -0.6000     0.2309  -2.598    0.122
teacher       0.2667     0.3127   0.853    0.484

Residual standard error: 0.2582 on 2 degrees of freedom
Multiple R-squared:  0.9111,    Adjusted R-squared:  0.7778 
F-statistic: 6.833 on 3 and 2 DF,  p-value: 0.1303
```

INTRODUCING LASSO (AND OTHER PENALIZED REGRESSION ESTIMATORS)

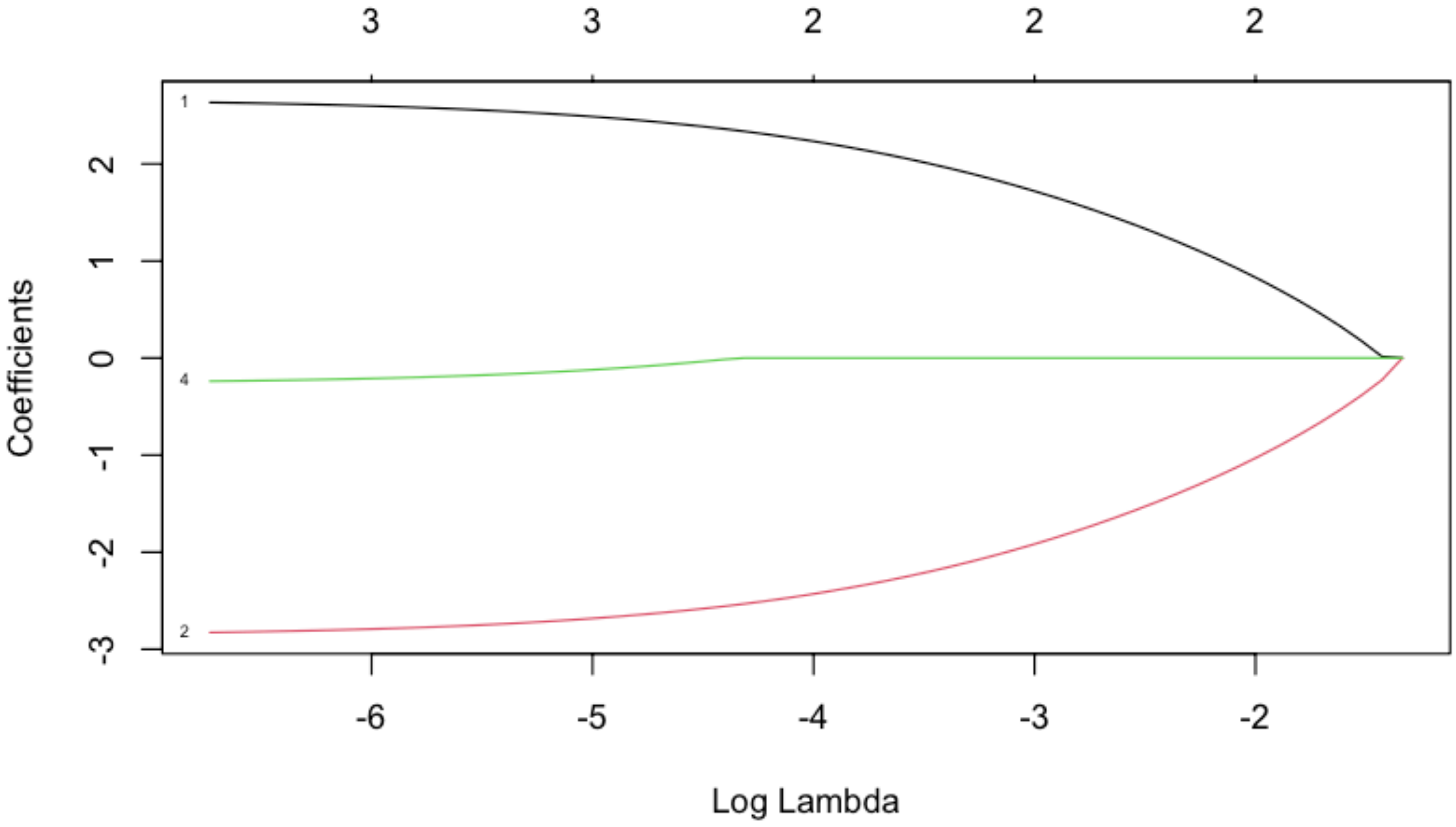
LASSO (LEAST ABSOLUTE SHRINKAGE SELECTOR OPERATOR)

- Recall: linear regression minimizes the sum of squared errors $\min_{\beta} \sum_{i=1}^{n} (y_i - X_i\beta)^2$
- LASSO minimizes $\min_{\beta} \left(\sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$ where the second term is the sum of the absolute value of the coefficients times the hyperparameter.



OK, BUT WHAT DOES THAT ACTUALLY MEAN?

- LASSO "shrinks" many coefficients to 0.
  - $\text{ssr}_{\text{ols}} = y_i - (0.47 + 0.6 \times \text{good}_i - 0.6 \times \text{bad}_i + 0.27 \times \text{teacher}_i) = 0.13$
  - $\text{ssr}'_{\text{ols}} = y_i - (0.47 + 0.6 \times \text{good}_i - 0.6 \times \text{bad}_i + 0 \times \text{teacher}_i) = 0.34$
  - $$\begin{aligned}\text{ssr}_{\text{lasso}} &= y_i - (0.47 + 0.6 \times \text{good}_i - 0.6 \times \text{bad}_i + 0.27 \times \text{teacher}_i) + \\ &\quad 1 \times (0.47 + 0.6 + 0.6 + 0.27) = 2.07\end{aligned}$$
  - $$\begin{aligned}\text{ssr}'_{\text{lasso}} &= y_i - (0.47 + 0.6 \times \text{good}_i - 0.6 \times \text{bad}_i + 0 \times \text{teacher}_i) + \\ &\quad 1 \times (0.47 + 0.6 + 0.6 + 0) = 2.01\end{aligned}$$


3
3
2
2
2
1
4
2
Coefficients
2
1
0
-1
-2
-3
-6
-5
-4
-3
-2
Log Lambda

LASSO, AND RIDGE, AND RANDOM FOREST, OH MY!

- So many models...
- Model averaging.

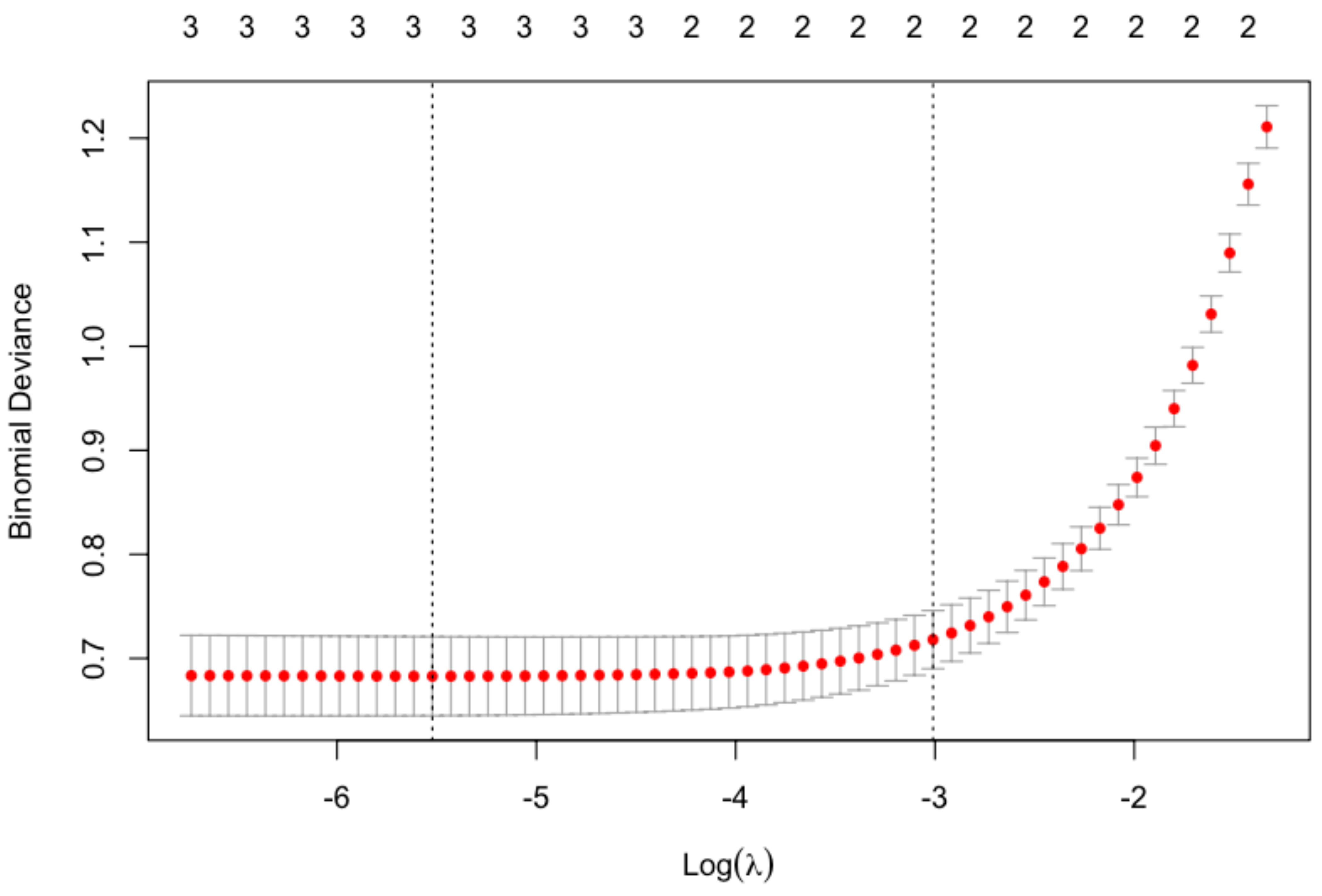
via DALL-E

VALIDATING YOUR CLASSIFIER

- Train/test/validation split:
  - Label 1,000 documents.
  - Train model on 700 documents.
  - Test model on 200 documents.
  - Iterate.
  - Retrain model on all 900 documents.
  - Test with held out validation set.

VALIDATING YOUR...HYPERPARAMER?

- We can split our training data further and use it to select a value for λ.
- Cross validation:
  - Split your training data into V "folds," e.g., $V_i \in (v_1, v_2, \ldots v_5)$.
  - Train/test your model rotating out which is the test set.
  - Average the results.



STEPS IN SUPERVISED CLASSIFICATION