Building a Robot Judge

15. Measuring Entropy and Complexity in Text

Elliott Ash

Optimal legal complexity

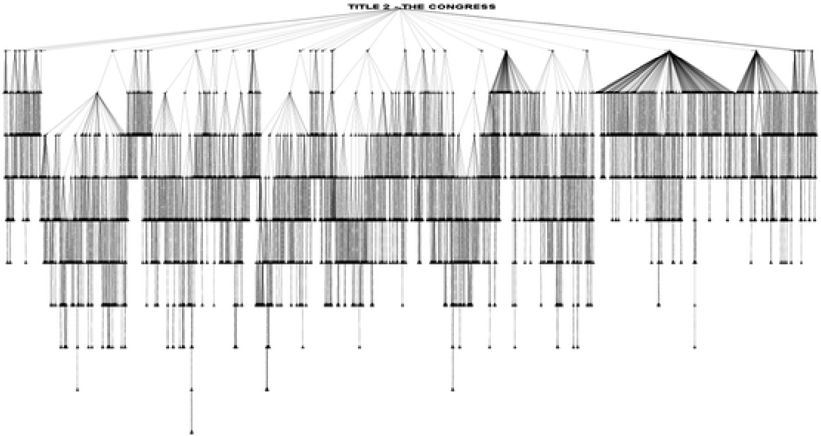
"Everything should be made as simple as possible, but no simpler."

--not Einstein

- More detail is needed in law to properly target incentives to activities and groups.
  - but there are costs to understanding/following complex laws, so there is a trade off.
- Analyzing this issue empirically requires a measure of complexity/detail.

The U.S. Code

Katz and Bommarito (2014)



- The U.S. Code consists of 49 titles, which can be further subdivided into subtitle, chapter, subchapter, part, subpart, section, subsection, paragraph, subparagraph, clause, and subclause.

Number of Elements

Measuring Complexity

Five largest and smallest titles by structural size

Title	V
Public Health and Welfare (Title 42)	110,605
Internal Revenue Code (Title 26)	51,553
Conservation (Title 16)	33,062
Agriculture (Title 7)	29,191
Education (Title 20)	28,096
Arbitration (Title 9)	68
General Provisions (Title 1)	84
Flag and Seal, Seat of Government, and the States (Title 4)	221
Intoxicating Liquors (Title 27)	224
Census (Title 13)	272

Mean Element Depth

Measuring Complexity

Five largest and smallest titles by mean element depth

Title	Avg. depth
Internal Revenue Code (Title 26)	7.80
Govt. Organization and Employees (Title 5)	6.66
Aliens and Nationality (Title 8)	6.51
Education (Title 20)	6.41
Transportation (Title 49)	6.40
Census (Title 13)	3.97
National Guard (Title 32)	3.50
Flag and Seal, Seat of Govt. and the States (Title 4)	3.23
General Provisions (Title 1)	2.85
Arbitration (Title 9)	2.82

Cross-Reference Complexity

Measuring Complexity

Five largest importing and exporting titles

Title	Net flow	Net flow per section
Govt. Organization and Employees (Title 5)	2,654	2.58
Crimes and Criminal Procedure (Title 18)	836	0.62
Money and Finance (Title 31)	751	1.59
Judiciary and Judicial Procedure (Title 28)	659	0.83
Internal Revenue Code (Title 26)	576	0.28
Banks and Banking (Title 12)	−514	−0.28
Conservation (Title 16)	−534	−0.11
War and National Defense (Title 50)	−561	−0.78
Foreign Relations and Intercourse (Title 22)	−719	−0.25
Public Health and Welfare (Title 42)	−846	−0.11

Intra-Title Citation Rates

Measuring Complexity

Table 10

Five highest and lowest proportions of intra-title citation

Title	Citations received
Internal Revenue Code (Title 26)	0.97
Bankruptcy (Title 11)	0.96
Copyrights (Title 17)	0.95
Flag and Seal, Seat of Government, and the States (Title 4)	0.92
General Provisions (Title 1)	0.92
Public Printing and Documents (Title 44)	0.59
Patriotic Societies and Observances (Title 36)	0.59
Public Buildings, Properties, and Works (Title 40)	0.59
National Guard (Title 32)	0.58

Number of Words

Measuring Complexity

Five largest and smallest titles by token count

Title	Tokens	Tokens per section
Public Health and Welfare (Title 42)	2,732,251	369.22
Internal Revenue Code (Title 26)	1,016,995	487.07
Conservation (Title 16)	947,467	200.48
Commerce and Trade (Title 15)	773,819	336.88
Agriculture (Title 7)	751,579	274.00
President (Title 3)	7,564	120.06
Intoxicating Liquors (Title 27)	6,515	144.78
Flag and Seal, Seat of Govt. and the States (Title 4)	5,598	119.11
General Provisions (Title 1)	3,143	80.59
Arbitration (Title 9)	2,489	80.29

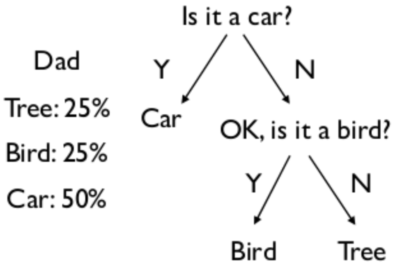
Average Word Length

Measuring Complexity

Ten titles with highest average word size

Title	Avg. word size
Domestic Security (Title 6)	6.90
War and National Defense (Title 50)	6.83
Public Printing and Documents (Title 44)	6.74
Foreign Relations and Intercourse (Title 22)	6.74
Public Contracts (Title 41)	6.73
Crimes and Criminal Procedure (Title 18)	6.16
Intoxicating Liquors (Title 27)	6.15
Internal Revenue Code (Title 26)	6.10
Flag and Seal, Seat of Govt. and the States (Title 4)	6.10
Bankruptcy (Title 11)	6.07

Digression: Twenty Questions (DeDeo 2018)



- The optimal set of questions minimizes the expected number of turns until you guess the answer.
  - The length of the optimal script is closely approximated by the **entropy**

$$H(X) = -\sum_{i=1}^{N} \Pr(x_i) \log_2 \Pr(x_i)$$

where one guesses from words $\{x_1, x_2, \ldots, x_N\}$.

Axioms for Shannon Entropy

Shannon (1948) wanted a function $H(\vec{p})$ for a vector of probabilities $\vec{p}$, that would satisfy four axioms:

1. continuity (small changes in $p_i \rightarrow$ small changes in H)
2. symmetry (invariance to re-ordering of $\vec{p}$
3. Condition of maximum information (H maximized when all p_i are equal
4. [next slide]

4. Coarse Graining

- Consider a set $X = \{p_a, p_b, p_c\}$
  - Consider a subset $X' = \{p_a, p_{bc}\}$ where $p_{bc} = p_b + p_c$ (we don't distinguish b and c
- Coarse graining requires

$$H(X) = H(X') + p_{bc} H(G)$$

  where $G = \{\frac{p_b}{p_{bc}}, \frac{p_c}{p_{bc}}\}$ the distribution for making the finer-grained distinction between b and c.
- The unique function satisfying these four assumptions is

$$H(X) = -\sum_{x \in X} \Pr(x) \log \Pr(x)$$

Continuous Distributions and Cross Entropy

- For continuous distribution with pdf $p(x)$:

$$H(p(x)) = -\int_x p(x) \log(p(x)) dx$$

  - broad distributions have higher entropy.

- The cross entropy for distributions $p(x)$ and $q(x)$ is a measure of similarity between distributions:

$$H(p,q) = -\int_x p(x) \log((q(x)) dx$$

  - In machine learning, can be used as a loss function, where p is true and q is the model prediction.
  - It is the standard loss function for logistic regression.

Entropy in Text

- In text data, the probabilities p_i can be interpreted as the probability (frequency) of observing particular tokens: characters, words, n-grams, etc.
  - In general, documents with more diverse vocabularies/topics will have higher entropy.
  - could run compression algorithms on text and measure bytes of compressed data.

Word Entropy in the U.S. Code

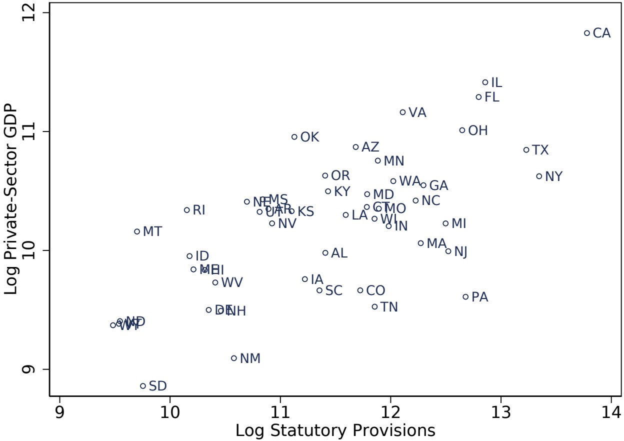
Measuring Complexity

Five highest and lowest titles by word entropy

Title	Word entropy
Commerce and Trade (Title 15)	10.80
Public Health and Welfare (Title 42)	10.79
Conservation (Title 16)	10.75
Navigation and Navigable Waters (Title 33)	10.67
Foreign Relations and Intercourse (Title 22)	10.67
Intoxicating Liquors (Title 27)	9.01
President (Title 3)	8.89
National Guard (Title 32)	8.50
General Provisions (Title 1)	8.49
Arbitration (Title 9)	8.24

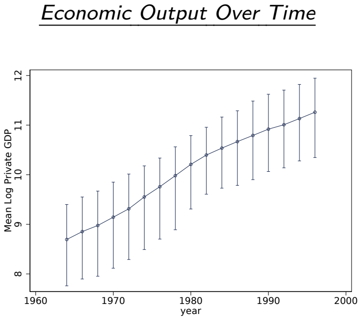
Laws ↔ Growth

Ash, Morelli, and Vannoni (2018)



Log Private-sector GDP and Log Statutory Provisions by state, 1997.

Economic Output

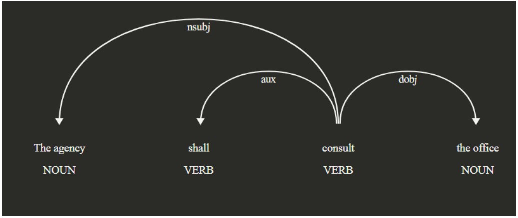


Mean log private-sector GDP by biennium. Error spikes indicated 25th and 75th quantiles.

- Data: sectoral GDP by state from Bureau of Economic Analysis Regional Accounts
  - output by SIC code (57 industries)
  - 1963-1997 (changed to NAICS in 1998)
- Economic growth defined as log change in private-sector GDP by biennium.

Extracting Legal Provisions

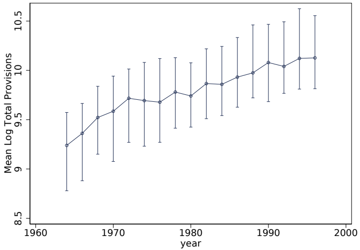
- Pre-processing steps:
  - segment session laws into statutes, segment statutes into sentences
- Extract legal meaning:
  - apply syntactic dependency parser
  - see figure (spaCy example)



- Count legal provisions:
  - obligations/delegations (agent shall/must/will, is required, ...)
  - prohibitions (agent shall/must/may (not), is prohibited, ...)
  - permissions (agent may, is permitted, ...)
  - entitlements (agent has, retains ...)

Legislative Output

Legislative Output over Time



Mean log number of legal provisions by biennium. Error spikes indicated 25th and 75th quantiles.

- Legislative output is the log number of legal provisions for each biennium in a state.
    - More precise measure of volume of legal requirements than word counts

Empirical Model

$$\Delta \log(W_{st}) = \alpha_{st} + \rho_{Y \to W} \Delta \log(Y_{st}) + X'_{st}\beta + \varepsilon_{st}$$

- W_{st}, number of legal provisions enacted in state s at biennium t
- Y_{st}, private-sector GDP in state s at biennium t
- α_{st}, state and time fixed effects
- X'_{st}, other observable factors
- ε_{st}, unobservable factors and random noise
- $\rho_{Y \to W}$, effect of growth on legislation

Instrumental Variables Approach

- Adopt IV method based on Bartik (1991) and Acemoglu et al (2018).
- Classic application (Bartik 1991):
  - Instrument for local growth rates with interaction between pre-treatment local sectoral shares and national growth rates by sector
  - Identification assumption: National growth trends by sector are orthogonal to local pre-treatment sector shares.

No Effect of Economic Growth on Legislative Detail

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	RF	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Growth$\rightarrow$Detail	0.542**	0.0429+	1.091	1.082	0.611	1.132+	1.074	1.018
($\rho_{Y \rightarrow W}$)	(0.155)	(0.0236)	(0.711)	(0.715)	(0.541)	(0.445)	(0.734)	(0.709)
Obs	846	846	846	846	812	846	846	846
First-Stage F			66.59	67.89	70.34	53.01	63.55	67.56
State FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
State Trends				X				
Pre-Treat X					X			
Pop/Income						X		
Govt Expend							X	
Politics								X

Robust standard errors (clustered by state) in parentheses. ** p<0.01, * p<0.05, + p<0.1.

Effect of Legislative Detail on Economic Growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Detail→Growth	0.016*	0.097**	0.11**	0.07**	0.13**	0.088**	0.095**	0.1**
($\rho_{W \to Y}$)	(0.0069)	(0.032)	(0.042)	(0.02)	(0.04)	(0.033)	(0.03)	(0.04)
Obs	843	843	843	809	843	843	843	789
First-Stage F		15.64	14.76	11.91	16.24	15.61	15.66	11.12
State FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
State Trends			X					
Pre-Treat X				X				
Pop/Income					X			
Govt Expend						X		
Politics							X	
Lagged DV								X

Robust standard errors (clustered by state) in parentheses. ** p<0.01, * p<0.05, + p<0.1.

Notes

- A 1% increase in legislative growth rate increases economic growth rate by ~0.1%.
- Alternative "detail" variables – strong first stage and significant 2SLS treatment effects:
  - log number of words ($F = 16.15$, $p = .032$)
  - log number-of-provisions / number-of-words ($F = 11.48$, $p = .029$)
- Placebo test for reduced form:
  - outcome variable not significantly related to leads of instruments.

Mechanisms? Other Outcomes

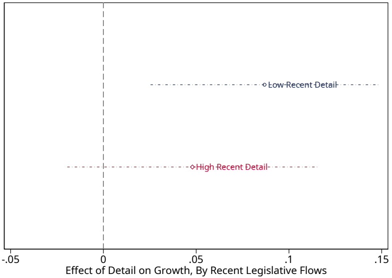
- Log GDP growth (baseline effect):
  - $\beta = .11$, $p = .006$
- Log population growth:
  - $\beta = .001$, $p = .82$
- Log income per capita growth:
  - $\beta = -.007$, $p = .57$
- Log establishment count growth:
  - $\beta = .016$, $p = .33$
- Log firm profits growth (gross surplus):
  - $\beta = .049$, $p = .38$
- Log employment growth:
  - $\beta = .08$, $p = .2$

Heterogeneous Effects by Previous Level of Detail

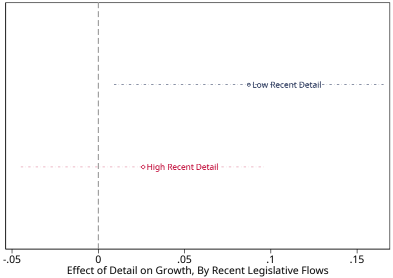
- In incomplete contracts models like Battigalli and Maggi (2002), there is an optimal level of detail that is increasing with the surplus (here, economy size)
  - if detail is too low, increasing detail should improve economic performance
  - if detail is optimal or already high, increasing detail should reduce economic performance.
- For each state/year, we compute the volume of legislation in the previous ten years.
  - we then split up the sample by whether this is above or below the median (or in top/bottom quantile)
  - (results similar to various ways of splitting sample, including computing median by year.)

Heterogeneous Effects by Previous Level of Detail

Top Half vs. Bottom Half Detail



Top Third vs. Bottom Third Detail



Coefficient plots from baseline 2SLS estimates, with sample split by volume of legislation in previous ten years.

Classifying Tax Statutes

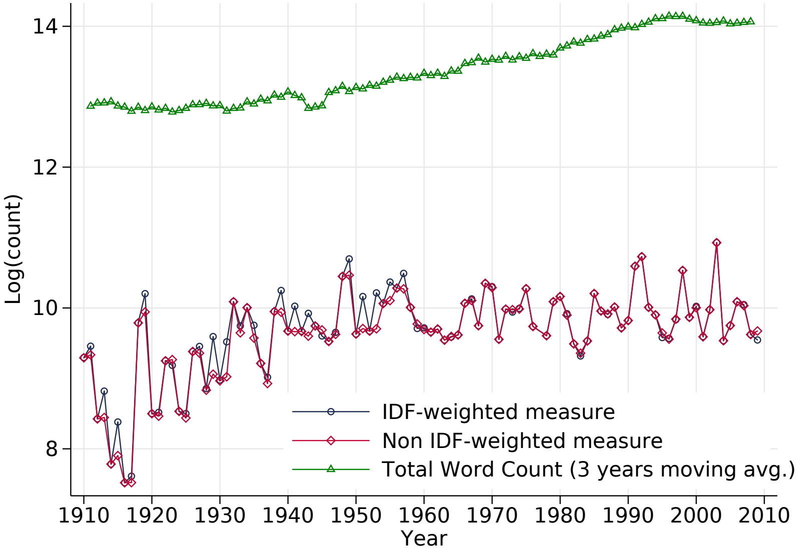
- Usual **pre-processing** steps of the tax codes:
  - Remove punctuation, stopwords
  - Text represented as Tf-idf weighted (1, 2, 3)-grams
- Model trained on corpus of labeled laws:
  - Logistic regression classifier with L2 regularization
  - Predict "is tax code" using text features.
  - Accuracy in held-out text sample: 98.5%, F1 score$= 0.82$
  - Most predictive phrases: “tax”, “revenue”, “commissioner”, “funds produced”, “filing financing statement”

Measure of state-year complexity is

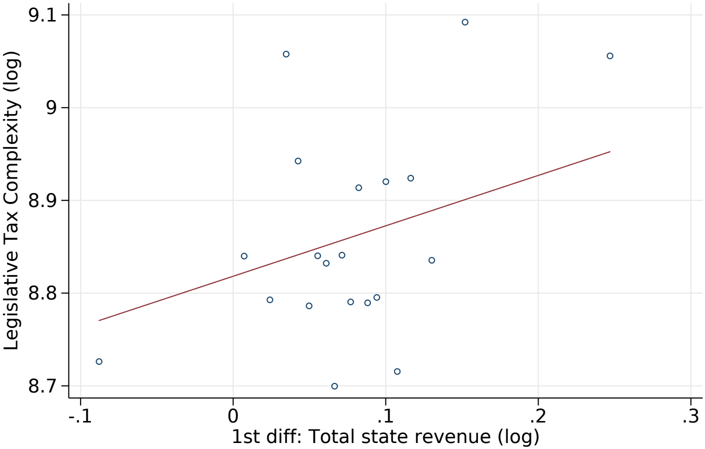
$$Y = \sum_{\text{s} \in \text{statutes}} \mathrm{Prob}(s \text{ is tax}) \times \text{Length}(s)$$

for statutes enacted by state and year.

Tax Law Complexity Over Time

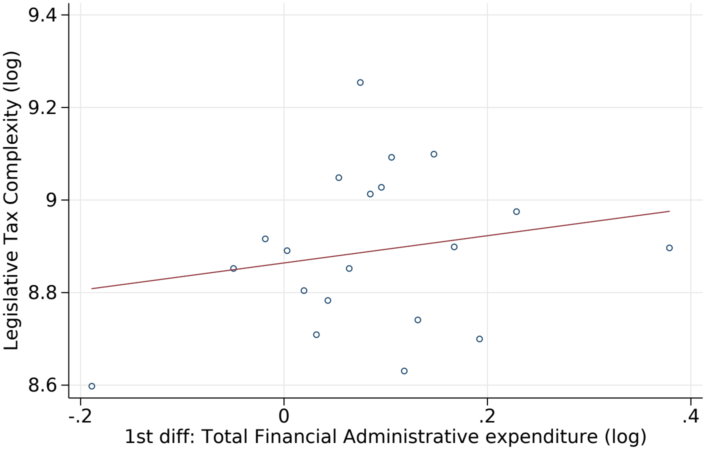


Text Complexity Increases with Revenues



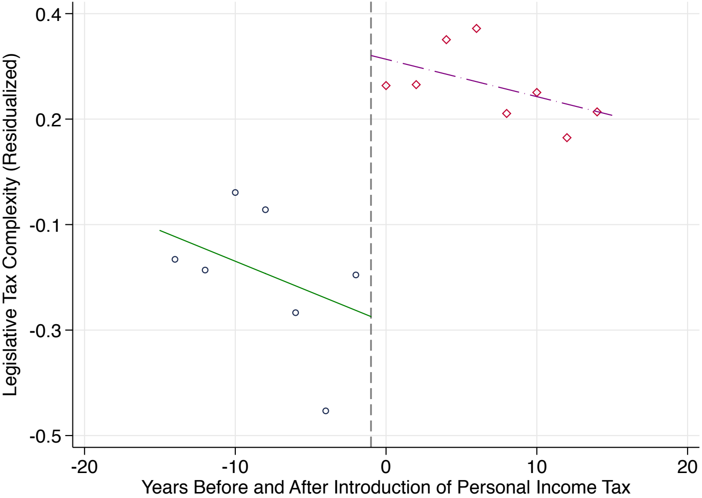
Slope: 0.54
S.e. (clustered at the state level): 0.24
Including year FE and controlling for states

Text Complexity Increases with Financial Admin Costs

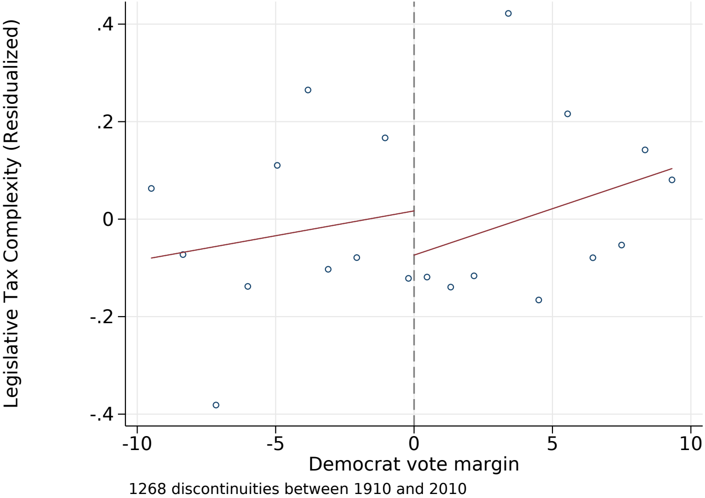


Slope: 0.28
S.e. (clustered at the state level): 0.15
Including year FE and controlling for states

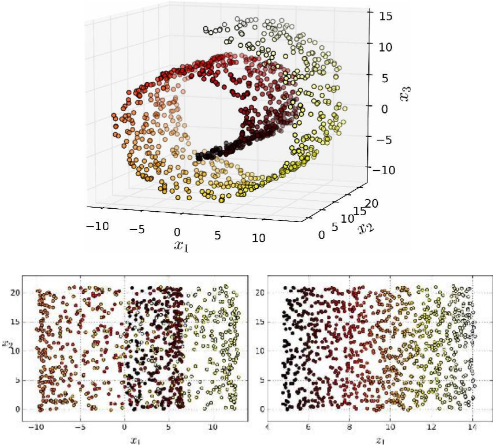
Event Study: Introduction of the Income Tax



No difference in complexity between political parties

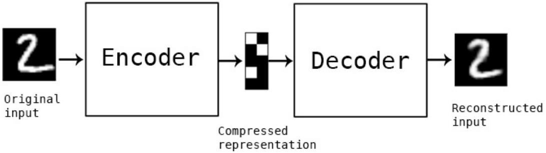


Remember the Swiss role?



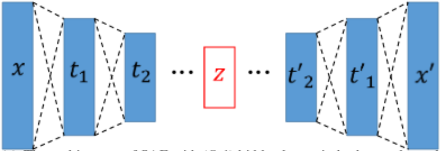
- The dimension reduction process matters: projecting down to two dimensions directly (left panel) might not isolate the variation we are interested in (as done in the right panel, which unrolls the Swiss Roll)

Autoencoders: Domain-specific dimension reduction

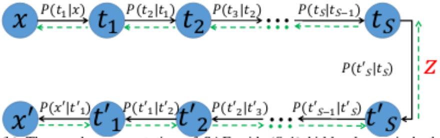


- "Autoencoder" refers to a class of deep neural network that performs domain-specific dimension reduction.
  - They learn efficient encodings of the data, which can then be decoded back to a (minimally) lossy representation of the original data.
  - Can also randomly generate new data that looks like the training data.

Stacked Autoencoder



(a) The architecture of SAE with (S-1) hidden layers in both encoder and decoder.



(b) The graph representation of SAE with (S-1) hidden layers in both encoder and decoder.

- Autoencoders work by stacking layers that gradually decrease in dimensionality to create the compressed representation (Z), and then gradually increase in dimensionality to try to reconstruct the input.
  - the autoencoder is implicitly solving the problem of maximizing entropy in the bottleneck layer.

Autoencoding for data visualization

- For 2D visualization, t-SNE is probably the best algorithm
  - but quite slow, and typically requires relatively low-dimensional data.
- Good strategy:
  - use an autoencoder to compress your data to relatively low dimension (e.g. 32 dimensions)
  - then use t-SNE for mapping the compressed data to a 2D plane.

Recap: Legal Complexity

- Policymakers and social scientists agree that "complexity" is an important concept/issue/
- NLP and information theory have provided some interesting measures of complexity for text.
  - but still no standard or consensus about how to do this.
  - using information content in neural nets (e.g. dimensionality in autoencoders) seems promising but mostly unexplored.
- Perhaps best approach for now is to try out different measures of complexity to see how they compare in downstream analysis.