1 Supervised topics with ideal point text regression

We can generalize ridge regression on ideal points to provide a fully supervised topic model. Recall that ridge regression on ideal points corresponds to the model

1. Fit the ideal point model

   $$\arg\max_{a_d, b_d, x_u} \sum_{u \in U, d \in D} \log p(v_{ud}|a_d, b_d, x_u) - \sum_{d \in D} (\lambda_1 a_d^T a_d + \lambda_2 b_d^T b_d) - \sum_{u \in U} \lambda_3 x_u^T x_u,$$

   where $p(v_{ud}|a_d, b_d, x_u)$ is the logistic function $\sigma(a_d^T x_u + b_d)$ and $\lambda_1, \lambda_2, \lambda_3$ are regularization parameters.

2. Fit the ridge regression

   $$\begin{aligned} \beta_a, \beta_b &\sim N(0, 1/\lambda_4), \\ a_d &\sim N(\boldsymbol{w}_d^T \beta_a, \sigma_a), \\ b_d &\sim N(\boldsymbol{w}_d^T \beta_b, \sigma_b), \end{aligned}$$

   where $\boldsymbol{w}_d$ are word counts and λ_4 is a ridge penalty.

Although this model performed well in the prediction task, it suffers from a critical limitation: the ideal points are informed only by votes. The regression coefficients β_a, β_b are interpretable, but only in the sense that they relate words to ideal points.

Ideal point text regression We can solve this problem by forcing $\sigma_a, \sigma_b \to 0$. We can accomplish this by modeling the document parameters only implicitly:

$$\arg\max_{\boldsymbol{B}, \boldsymbol{b}, x_u} \sum_{u \in U, d \in D} \log p(v_{ud}|\boldsymbol{w}_d, \boldsymbol{B}, \boldsymbol{b}, x_u) - \lambda_1 \sum_{ij} \boldsymbol{B}_{ij}^2 - \lambda_2 x_u^T x_u - \lambda_3 \boldsymbol{b}^T \boldsymbol{b}, \quad (1)$$

where $\boldsymbol{B}$ is a $r \times V$ matrix, x_u is again the $r \times 1$ user ideal point, $\boldsymbol{b}$ is a $V \times 1$ vector, λ_i regularize, and

$$p(v_{ud}|\boldsymbol{w}_d, \boldsymbol{B}, \boldsymbol{b}, x_u) = \sigma(\boldsymbol{w}_d^T (\boldsymbol{B} x_u + \boldsymbol{b}))$$

We call this model *ideal point text regression*, as it infers an ideal point $x_u \in \mathbb{R}^r$ for each individual while implicitly performing regression on word counts. The role of $\boldsymbol{b}$ is a per-word intercept term, serving the same purpose as the per-document difficulty terms b_d.

The columns of $\boldsymbol{B}$ define a r-dimensional subspace of $\mathbb{R}^V$ and can be interpreted as topics. [1] A document's topic vector is given by $\boldsymbol{w}^T \boldsymbol{B} \in \mathbb{R}^r$, and users' ideal points interact with documents' topics as with traditional ideal point models.

Setting r to 1, legislators' ideal points turn out similar to traditional ideal points (see Figure 2 for examples). Table 1 provides examples of words which are extreme in discrimination, difficulty, or indifference point.

[1] These topics are much more similar to classic LSA topics than LDA topics [2]

	Most	Least
Liked	national defense	dod
	industrial base	subtitle
	army navy	defense
	defense acquisition	mean
	develop and implement	directly
	procure	eligible
	clinical	item relate to section
	provisions subtitle	commodity
	national guard	tracking
	executive agency	limitation
Liberal	weapons	prepare
	testing	offset
	architect	desire
	reasonable	enrich
	acquire real property	stop
	passport	fha
	noncompliance	excludes
	energy technology	minimum standard
	research development	education and training
	maximum extent practicable	candidate
Polarizing (preferred by liberals)	thereon	waive except those arise
	witness	unemployment
	coalition	innovation
	distance	previous question
	revises	borrower
	bracket	fraud and abuse
	peer review	recipient
	core	chair
	asset	enrollee
	lieu	mean the secretary

Figure 1: Phrases in a 1-dimensional ideal point text regression. “Polarizing” is discrimination, “liberal” is indifference, and “liked” is difficulty.

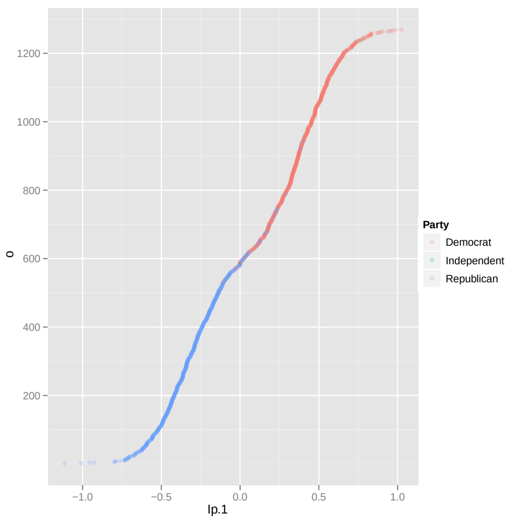

Figure 2: Legislator ideal points from a 1-dimensional ideal point text regression model.

Semiometrie interpretation A particularly nice interpretation of ideal point text regression comes about when we compare it with semiometrie. Semiometrie is a methodology used in marketing and politics for placing users in a latent space of preferences [1]. To find this latent space, each user $u \in U$ is asked to provide a rating $r_{ud} \in \{1, \ldots, 7\}$ to 210 words, summarizing their personal warmth toward these words. Principle component analysis is then applied to the resulting ratings, optimizing the PCA objective

$$\arg \max_{Z \in \mathbb{R}^r, X} \sum_{u \in U, d \in D} \log p(r_{ud} | \boldsymbol{Z}, X),$$

where

$$r_{ud} | \boldsymbol{Z}, X \sim N((\boldsymbol{Z} x_u)_d, \sigma_d), \tag{2}$$

subject to several orthogonalization constraints on Z, X, and σ_d: they are fit to maximize explained variance. The principle component $\boldsymbol{z}_1 := \boldsymbol{Z}_{1*}$ explains the positive ratings shared by everyone: users will share nearly identical loadings on this dimension. We therefore can interpret $\boldsymbol{z}_1$ as an intercept column and, for model simplicity, set $x_{*1} = 1$ (we can scale the component $\boldsymbol{z}_1$ of $\boldsymbol{Z}$ to do this). Rewriting Equation 2 with the intercept $\boldsymbol{z}_1$, we have

$$r_{ud} | \boldsymbol{Z}, X \sim N((\boldsymbol{Z}_{\backslash 1*} x_u + \boldsymbol{z}_1)_d, \sigma_d).$$

The similarity of ideal point text regression to this scenario is evident; akin to generalized linear models, we simply replace the *identity* (normal) link function with the *logit* link function:

$$\arg\max_{\boldsymbol{B},\boldsymbol{b},X} \sum_{u\in U,d\in D} \log p(v_{ud}|\boldsymbol{B},\boldsymbol{b},X),$$

where

$$v_{ud}|\boldsymbol{B},\boldsymbol{b},X \sim \sigma(\boldsymbol{w}_d^T(\boldsymbol{B}x_u+\boldsymbol{b})).$$

In short: the goal of ideal point text regression is to best explain observed votes by finding an intercept $\boldsymbol{b}$ and subspace $\boldsymbol{B}$ sensitive to user preferences which best explain observed votes.

In semiometrie, the principle component z_1 is typically discarded because it is not informative about individual preferences. The intercept $\boldsymbol{b}$ is likewise noninformative about individuals' preferences, although we keep it to investigate individual words. At the same time, the subspace described by B plays the same role as that of semiometrie: it summarizes legislators by their latent preferences.

References

[1] F. Camillo, M. Tosi, and T. Traldi. Semiometric approach, qualitative research and text mining techniques for modelling the material culture of happiness. 185:79–92, 2005.

[2] S. Deerwester, S. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 1990.