Day 1: Foundations, Regression Extensions & Variable Selection

Machine Learning for Management Research

Istanbul Bilgi University · Python & Google Colab

Bruce A. Desmarais

Introduction

Everyday encounters with machine learning

Routine examples:

- Personalized online advertisements
- Social media photo tagging
- GPS route optimization
- Email sorting for spam and important messages
- Voice recognition in personal assistants

Higher stake applications:

- Self-driving cars
- Robotic surgery
- Automated astronomical research

Classes of Machine Learning Methods

Learning Types in Machine Learning:

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning

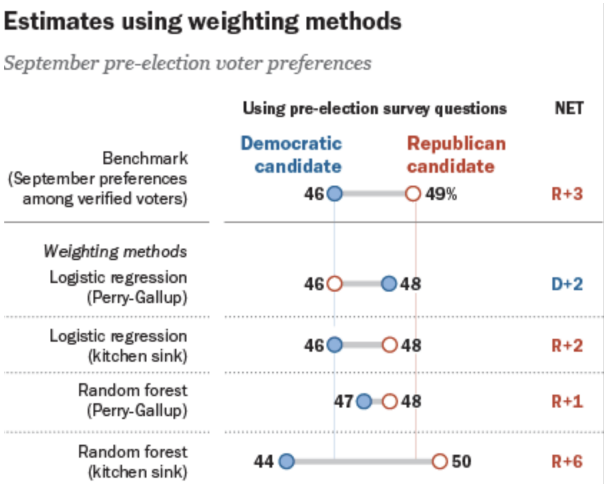
Subcategories in Supervised Learning:

1. Classification: Categorizing data into predefined classes.
2. Regression: Predicting continuous values.

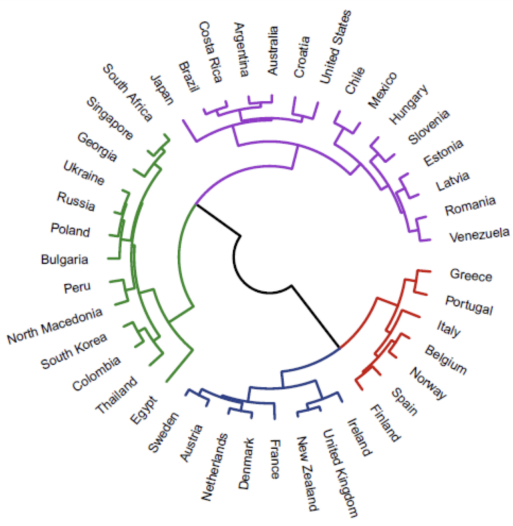
Subcategories in Unsupervised Learning:

1. Dimension Reduction: Simplifying the complexity of data.
2. Clustering: Grouping similar data without predefined classes.

Supervised


Estimates using weighting methods
September pre-election voter preferences
Using pre-election survey questions
NET
Democratic candidate
Republican candidate
Benchmark (September preferences among verified voters)
46
49%
R+3
Weighting methods
Logistic regression (Perry-Gallup)
46
48
D+2
Logistic regression (kitchen sink)
46
48
R+2
Random forest (Perry-Gallup)
47
48
R+1
Random forest (kitchen sink)
44
50
R+6
Source: 2014 American Trends Panel September and November waves. Based on registered voters who participated in both waves and were matched to a national voter.
PEW RESEARCH CENTER

Unsupervised



The dendrogram represents the cultural values of 43 countries. Shorter heights in between branches indicate greater similarity.

Source: Wormley et al., 2022 (Scientific Data)

Deep Learning

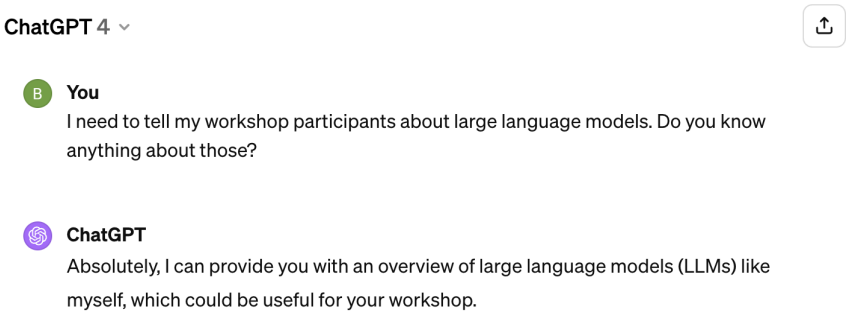
Context:

- A subfield of machine learning.
- Relationship: All deep learning is machine learning, but not vice versa.

Rise in Popularity:

- Exceptional model performance with very big data.
- Increased computational power (GPU) enabling broader application.

Large Language Models



Ethical Concerns in Machine Learning

Machine Learning: Potential for both beneficial and harmful applications.

1. **Predicting Sexual Orientation (2017):**
   - Study by Yilun Wang and Michal Kosinski.
   - Risk of misuse for persecution in certain countries.
   - Source: `https://osf.io/zn79k`

2. **Google's Image Recognition Bias (2015):**
   - Misclassification of people of color.
   - Result of training data bias towards images of white people.

Reflective Questions:

1. Are my intentions ethical?
2. Could my model be misused for harmful purposes?
3. Is there any bias in my model that could cause harm or discrimination?
4. Has the data been collected ethically and responsibly?
5. How will human decision-making integrate with the model once deployed?

Prediction

Objectives in modeling data

Three objectives in developing statistical models:

1. To describe/explore
2. To explain
3. To predict

Research design, and modeling choices, should be adapted to the objective.

The Differences Between Explanatory and Predictive Modeling

Explanatory modeling:

- Statistical model is careful operationalization of theory
- Parameter estimates are the objects of interest
- Statistical model itself is the object of interest.
- Variance less important than bias

Predictive modeling:

- Objects of interest are the variable values.
- Statistical model chosen to produce best predictions.
- Model does not necessarily correspond to any theory
- Prediction is typically prospective
- Variance possibly more important than bias

The Place of Prediction in Explanation

Explanation and prediction are distinct but complementary. Predictive modeling and evaluation can...

- Shed light on new relationships that require explanatory treatment.
- Help in comparing alternative measures of the same construct.
- Be used to compare theoretical claims.
- Be used to assess distance between model and reality.
- Quantify the "predictability" of a phenomenon.

Prediction as standalone objective

- Predictive enforcement of residential housing codes.

  Ye, Teng, Rebecca Johnson, Samantha Fu, Jerica Copeny, Bridgit Donnelly, Alex Freeman, Mirian Lima, Joe Walsh, and Rayid Ghani. "Using machine learning to help vulnerable tenants in New York city" *In Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 248-258. 2019.

- Forecasting civil unrest for defense/intelligence policy

  Bowlsby, Drew, Erica Chenoweth, Cullen Hendrix, and Jonathan D. Moyer. "The future is a moving target: Predicting political instability." *British Journal of Political Science* 50, no. 4 (2020): 1405-1417.

Bias-variance tradeoff

Estimation error: reflects the variation in model estimates from one sample to another. Estimation error

- Can be reduced by dropping weak predictors.
- Can be reduced through non-linear structures that focus on predictive coverage.

Reducing estimation error can increase predictive performance even if it means departing from theoretical model.

True model not always best for prediction

Conditions from linear modeling with OLS...

- Large error term variance.
- Effects are small in magnitude.
- Highly correlated predictors.
- Small sample size.

Trivial case: sample size smaller than true number of predictors.

Key Takeaways

- Prediction and Explanation are two related, but distinct, statistical modeling objectives.
- Prediction is a worthy objective in much scientific research.
- Researchers should be explicit about modeling objectives.
- Research design choices depend on objectives.
- Theoretically correct models can be sub-par at predicting.

Model Evaluation

Why do we care about evaluating a model?

- Better models make for better hypothesis tests.
- Key to data-driven modeling (e.g., new data/topic).
- Necessary step if research objective is predictive.

In-sample bias

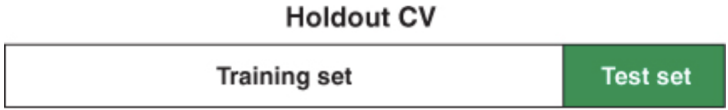
- Prediction error on the training sample is usually smaller than on out-of-sample data.
- The difference between in-sample and out-of-sample error grows with model complexity.

Example: With linear regression, in expectation

$$\text{in-sample-error} - \text{out-of-sample-error} = 2\frac{d}{N}\sigma_\epsilon^2,$$

where d is the number of covariates, N is the sample size, and σ_ϵ^2 is the variance of the error term.

Holding out data



1. The data is randomly split into a training and test set.
2. A model is trained using only the training set.
3. Predictions are made on the test set.
4. The predictions are compared to the true values.

Rhys, Hefin. Machine Learning with R, the tidyverse, and mlr. Simon and Schuster, 2020.

Cross-validation

- Iteratively omitting some observations from estimation, to be used as a validation set.
- With a small validation set (e.g., LOOCV), bias is low, but variance is high.
- With a large validation set (e.g., 5-fold-CV), variance is low, but biased towards methods that do well on small data.
- Cross-validation can only estimate performance of a model type (expected error), not a specific set of estimates (conditional error).

What are Hyperparameters?

- Parameters not learned from the data.
  - Number of variables in a regression model
  - Whether to take the log of one or more independent variables
- Set before training.
- Impact model training and predictive power.
- Not one-size-fits-all; require tuning for each dataset.

Exhaustive Grid Search

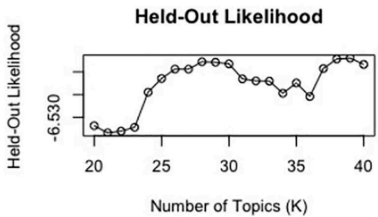
- Define a grid of hyperparameter values.
- Systematically explore all combinations.
- Pros:
  - Comprehensive.
  - Guaranteed to find the best combination within the grid.
- Cons:
  - Computationally expensive.
  - Fineness of the grid affects quality and efficiency.

Random Search

- Randomly sample the hyperparameter space.
- Pros:
  - More efficient than grid search, so can search broader space.
  - Can explore a wider range of values.
- Cons:
  - No guarantee to find the optimal combination.
  - Relies on randomness; results can vary between runs.

What are Validation Curves?

- Show training and validation performance as a function of model complexity (e.g., hyperparameters).
- Assist in:
  - Identifying the best model complexity.
  - Diagnosing issues with model training.



McDonald, Maura, Rachel Porter, and Sarah A. Treul. "Running as a woman? Candidate presentation in the 2018 midterms." *Political Research Quarterly* 73, no. 4 (2020): 967-987.

Evaluating binary predictions

- **Accuracy**:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$$

- **Precision**:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall** (Sensitivity):

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1 Score**:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Area Under Curves: ROC and Precision-Recall

- **Area Under the ROC Curve (AUC-ROC)**:
  - The ROC curve plots True Positive Rate (TPR) against False Positive Rate (FPR) at various threshold settings.
  - Mathematically, $\text{TPR} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}}$ and $\text{FPR} = \frac{\text{False Positives}}{\text{False Positives+True Negatives}}$.
  - AUC-ROC is the area under this ROC curve ($0 \leq \text{AUC-ROC} \leq 1$).
  - 0.5 is random

- **Area Under the Precision-Recall Curve (AUC-PR)**:
  - The Precision-Recall curve shows the trade-off between precision and recall for different thresholds.
  - AUC-PR summarizes the curve's shape, providing a measure of a model's ability to distinguish between classes.
  - ($0 \leq \text{AUC-PR} \leq 1$). No absolute benchmark

Evaluating Continuous predictions

- **Mean Absolute Error (MAE)**:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Mean Squared Error (MSE)**:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **Root Mean Squared Error (RMSE)**:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- **R-squared (Coefficient of Determination)**:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Takeaways

- Model evaluation is an integral parts of data-driven inference.
- In-sample fitting criteria reflect biased estimates of prediction error.
- Well-designed hold-out experiments can estimate error under very general modeling conditions.
- Hyperparameter tuning, via predictive experiments, is a key step in machine learning

Linear model review

A Line

The equation of a line is

$$y = b_0 + b_1 x$$

- y is the **dependent variable**
- b_0 is the y-**intercept**
- b_1 is the **slope**
- x is the **independent variable**

Simple Regression - Setup

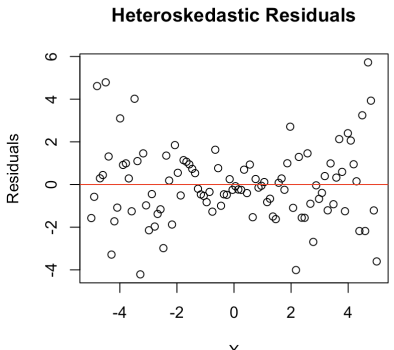
$$Y_i = \alpha + \beta X_i + \epsilon_i$$

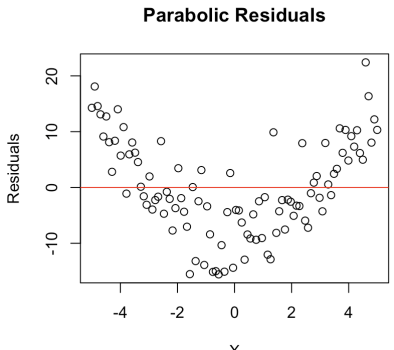
$$\epsilon_i \sim N(0, \sigma^2)$$

- Parameters: α, β, σ^2
- Estimates: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$

Residual Plots and Their Importance

- Residuals: Differences between observed and predicted values.
- Residual plots: Display residuals on the vertical axis and the independent variable on the horizontal axis.





Why do we (still) care about OLS?

- Forms foundation for most other models
- May just outperform more complex learners.

Predicting Driver Behavior from Stress

Technique	Metrics		
	MSE	MAE	R^2
Simple Linear Regression	0.0103	0.0771	0.8123
Linear SVR	0.0128	0.0971	0.7184
SVR (RBF kernel)	0.0131	0.0985	0.7118
Decision Tree Regression	0.0110	0.0826	0.7861

Verma, Rohit, Bivas Mitra, and Sandip Chakraborty. "Avoiding stress driving: Online trip recommendation from driving behavior prediction." In 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, pp. 1-10. IEEE, 2019.

Logistic regression

Extension to dichotomous Y

Problem: $Y_i \in \{0, 1\}$, and

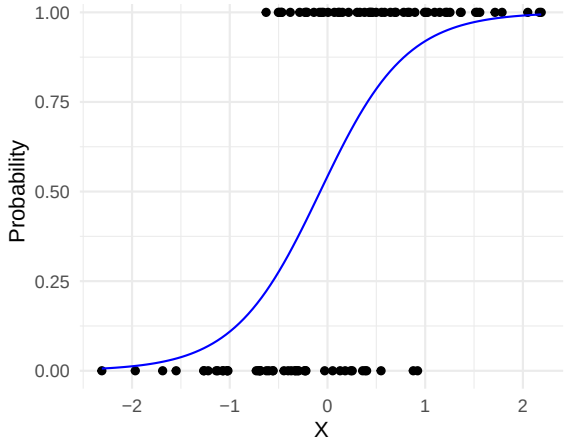
$$0 < E[Y_i] < 1$$

Solution: Logistic regression, a special case of the generalized linear model (GLM)

- $\pi_i = Pr(Y_i = 1)$
- $Y_i \sim \text{Bernoulli}(\pi_i)$
- $\frac{\pi_i}{1-pi_i}$ is the "odds"
- $\ln\left[\frac{\pi_i}{1-\pi_i}\right] = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik}$
- $Pr(Y_i = 1) = \frac{1}{1+\exp[-(\beta_0+\beta_1 X_{i1}+...+\beta_k X_{ik})]}$

The logistic curve

$$Pr(Y_i = 1) = \frac{1}{1 + \exp\left[-(3X)\right]}$$



Maximum Likelihood Estimation

Estimate given by

$$\max_{\theta} f_n(\boldsymbol{x}|\theta)$$

$\hat{\theta}$ is the value of theta under which the data would have been most likely

Can say nothing about the likelihood of the parameters

MLE: Example—Bernoulli distribution

$f_n(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \theta_i^x (1-\theta)^{1-x_i}$

$= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}$

$\ln(f_n(\boldsymbol{x}|\theta)) = \ln(\theta) \sum_{i=1}^{n} x_i + \ln((1-\theta))(n - \sum_{i=1}^{n} x_i)$

$\frac{d \ln(f_n(\boldsymbol{x}|\theta))}{d\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} + \frac{(n-\sum_{i=1}^{n} x_i)}{\theta - 1}$

to maximize, set equal to 0, and solve for theta

$-(1 - \frac{1}{\theta}) = \frac{(n-\sum_{i=1}^{n} x_i)}{\sum_{i=1}^{n} x_i}$

$(\frac{1}{\theta} - 1) = \frac{n}{\sum_{i=1}^{n} x_i} - 1$

$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$

Wrap-up

- Logistic regression represents a straightforward and interpretable classification method.
- MLE permits global optimization of in-sample fit.
- Preview: Foundation for neural nets. LR is a one-layer, fully visible neural net.

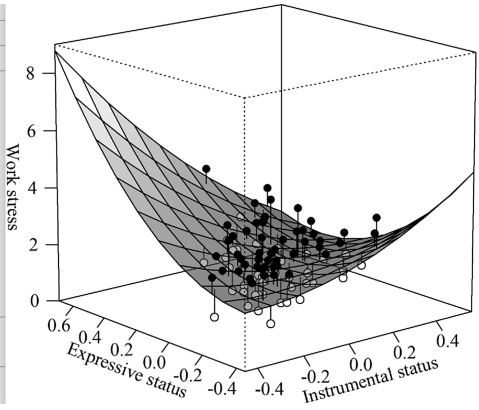
Polynomial regression

Introduction to Polynomial Regression

- Polynomial regression is a type of regression analysis that models the relationship between a dependent variable and one or more independent variables using a polynomial function.
- It can be represented as:
  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \epsilon$
- Used when the data shows curvilinear relationships.

Variable	Model 1		Model 2		Model 3	
	b	*SE*	*b*	*SE*	*b*	*SE*
Intercept	2.01*	0.38	1.78*	0.40	1.49*	0.39
Organizational tenure	0.00	0.01	0.00	0.01	0.01	0.01
Formal position	−0.04	0.05	0.03	0.05	0.01	0.05
Hours per week	0.01	0.01	0.01	0.01	0.01	0.01
Gender (M=0, F=1)	−0.21	0.25	−0.27	0.25	−0.02	0.25
Instrumental Status (b_1)			−1.15*	0.52	−1.42*	0.55
Expressive Status (b_2)			0.68	0.45	0.75	0.52
Instrumental Status^2 (b_3)					2.86	2.08
Expressive Status^2 (b_4)					5.03*	1.69
Inst. × Exp. Status (b_5)					−8.88*	3.06
R^2 (adj.)	0.00		0.02		0.10	
ΔR^2 (adj.)			0.02		0.08*	

Note: N=93. Coefficients are unstandardized. *p<0.05 (two-tailed).



van de Brake, H.J., Grow, A. and Dijkstra, J.K., 2017. Status inconsistency in groups: How discrepancies between instrumental and expressive status result in symptoms of stress. Social Science Research, 64, pp.15-24.

Generalized Additive Models

Generalized Additive Models

Generalized Additive Models (GAMs) extend Generalized Linear Models (GLMs) by allowing non-linear relationships. This flexibility is often achieved using splines.

- In the context of GAMs, the model can be expressed as:

$$g(E[Y]) = \beta_0 + \sum_{j=1}^{p} s_j(x_j)$$

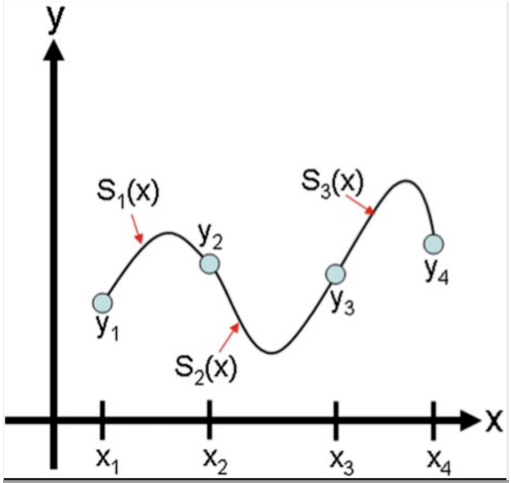
  where $g(\cdot)$ is the link function, $E[Y]$ is the expected value of the response variable, and $s_j(x_j)$ are splines.

- Splines, such as cubic splines or B-splines, are used to model $s_j(x_j)$, providing a smooth and flexible fit.

GAMs in practice

- Will not automatically learn interactions.
- Univariate functions facilitate graphical interpretation.
- Hyperparameters: knots, knot placement, polynomial degree, penalty for adding splines.
- GAMs are prone to overfitting—evaluate model performance before interpreting.

Splines



$$S_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$$

Regularization

Regularization

Problem of overfitting: Noise in the data is spuriously fit as systematic pattern.

Regularization solution: Modify the estimation criterion to penalize explaining variation as systematic.

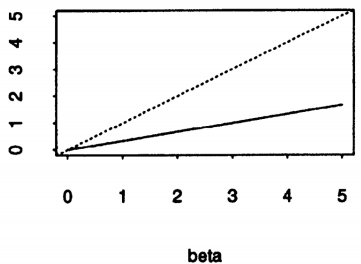
Canonical approaches: Ridge, lasso, and elasticnet methods for regression.

Many extensions/applications: e.g., gene network inference, clustering, topic models, factor analysis.

Regularization: Ridge regression

Ridge: "L2" penalization of coefficient magnitudes.

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \left[\sum_{i=1}^{n} \left(y_i - \boldsymbol{X}_i' \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$$

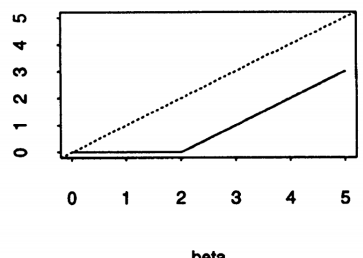


Source: Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288. Vancouver

Regularization: Lasso regression

Lasso: "L1" penalization of coefficient magnitudes.

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \left[\sum_{i=1}^{n} \left(y_i - \boldsymbol{X}_i' \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right]$$



Source: Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288. Vancouver

Regularization: Elasticnet regression

Elasticnet: "L1" and "L2" penalization combined.

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \left[\sum_{i=1}^{n} \left(y_i - \boldsymbol{X}_i' \boldsymbol{\beta} \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right]$$

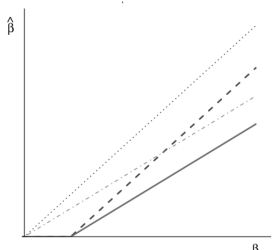


Fig. 2. Exact solutions for the lasso (-------), ridge regression (· · · · · ·) and the naïve elastic net (———) in an orthogonal design (· · · · · · ·, OLS): the shrinkage parameters are $\lambda_1 = 2$ and $\lambda_2 = 1$

Source: Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), pp.301-320.. Vancouver

Considerations

Tuning and selection

- Cross-validation is most common approach for tuning λ's.
- Elasticnet usually best, but takes much longer to tune.

Variable selection

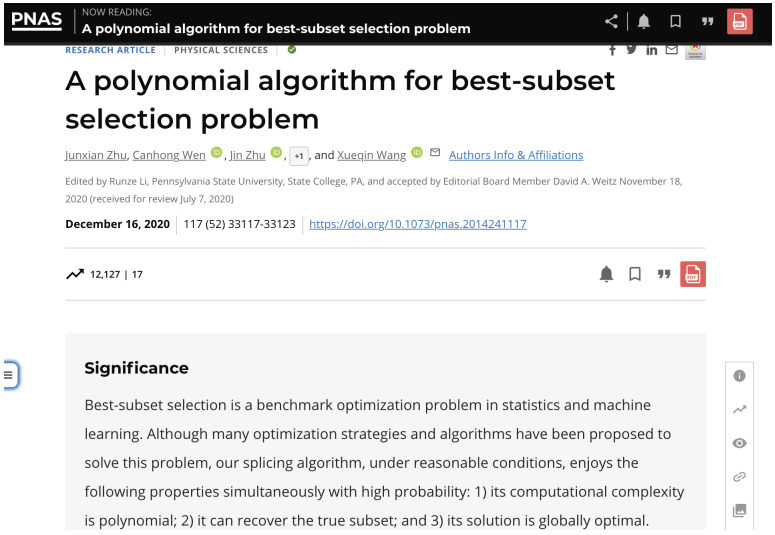
- Lasso's strength is the boundary solution (i.e., $\hat{\beta} = 0$).
- Challenge: Regularization parameters optimized for prediction.
- Key property: One-shot lasso over-selects variables relative to the true model. Two solutions:
  - Double lasso: Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies.
  - Bootstrap lasso: Bach, F.R., 2008, July. Bolasso: model consistent lasso estimation through the bootstrap. ICML 25.

Best Subsets Regression

Overview of Best Subsets Regression

- Examines all possible combinations of predictors to identify the optimal model.
- Aims to find a balance between model complexity and explanatory power.
- Valuable when we have numerous potential predictors.
- Fewer predictors can aid interpretation.

Fast Best Subsets



PNAS | NOW READING: A polynomial algorithm for best-subset selection problem

RESEARCH ARTICLE | PHYSICAL SCIENCES

A polynomial algorithm for best-subset selection problem

Junxian Zhu, Canhong Wen, Jin Zhu, +1, and Xueqin Wang Authors Info & Affiliations

Edited by Runze Li, Pennsylvania State University, State College, PA, and accepted by Editorial Board Member David A. Weitz November 18, 2020 (received for review July 7, 2020)

December 16, 2020 | 117 (52) 33117-33123 | https://doi.org/10.1073/pnas.2014241117

12,127 | 17

Significance

Best-subset selection is a benchmark optimization problem in statistics and machine learning. Although many optimization strategies and algorithms have been proposed to solve this problem, our splicing algorithm, under reasonable conditions, enjoys the following properties simultaneously with high probability: 1) its computational complexity is polynomial; 2) it can recover the true subset; and 3) its solution is globally optimal.

ABESS overview

Let $||\beta||_0 = \sum_{i=1}^{p} I(\beta_p \neq 0)$. Consider the problem

$$\min_{\beta} L_n(\beta) \quad \text{s.t.} ||\beta||_0 \leq s$$

, where $L_n(\beta)$ is $1/(2n)$SSE

1. At a given s, use some screening method to define the active set as the variables for which $\beta_p \neq 0$, and the inactive set as the variables for which $\beta_p = 0$.
2. For values of $k \in (1, 2, ..., s)$, splice the best k inactive variables with the worst k active variables, until unable to improve by τ.
3. Repeat 1 & 2 for all values of $s \leq \frac{n}{\ln(p)\ln[\ln(n)]}$
4. Select subset w.r.t. "special" information criterion

$$\text{SIC} = nL_n(\beta) + ||\beta||_0 \ln(p) \ln[\ln(n)]$$

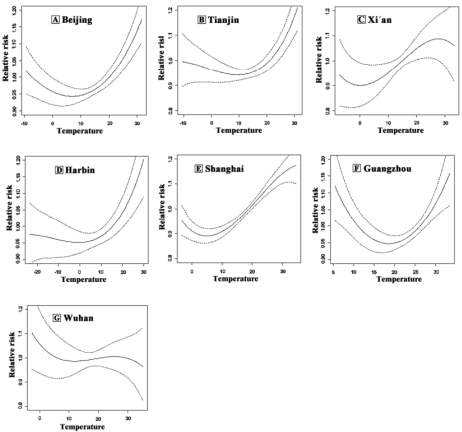
Day 1 wrap-up

- Prediction vs. explanation; the bias–variance tradeoff; honest evaluation by cross-validation.
- Regression extends from linear to logistic, polynomial, and smooth (GAM) models.
- Regularization (ridge/lasso/elastic net) and best subsets select variables for prediction.
- Tomorrow: tree ensembles, SVMs, k-NN, and how to interpret and compare models.

Further reading

Five studies that extend today's examples — with a figure from each

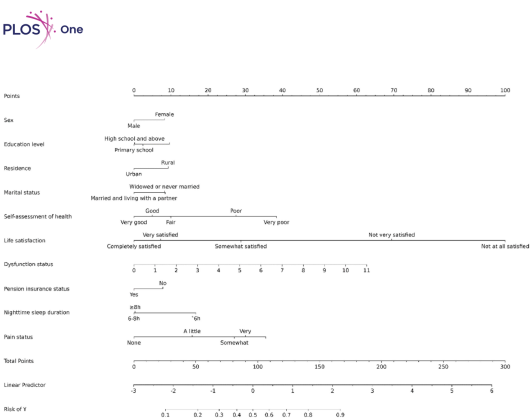
GAMs: temperature and mortality



A public-health application of the smooth (GAM/spline) models from today. Fitting death counts on temperature with penalized splines recovers a *U-shaped* exposure–response curve across Chinese cities: mortality rises at both cold and hot extremes — curvature a linear term would erase.

Zeng, Q., Li, G., Cui, Y., Jiang, G. & Pan, X. (2016). Estimating temperature–mortality exposure–response relationships . . . at the multi-city level of China. *Int. J. Environ. Res. Public Health* 13(3), 279.

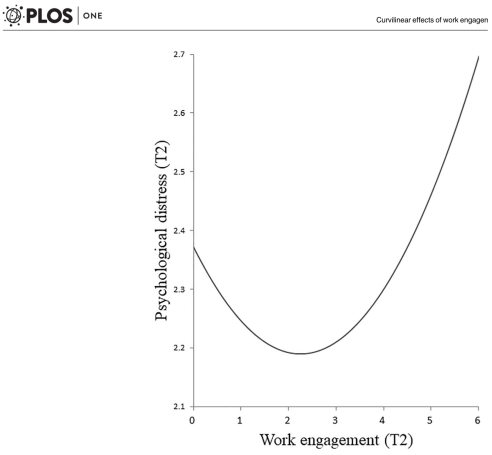
Lasso for a clinical risk model



Uses the **lasso** to pick, from many candidates, the handful of factors that predict depression in older adults with sensory impairment (CHARLS). The selected predictors become a validated risk score (nomogram) — regularized selection turned into a usable clinical tool.

Qin, M., Qiao, M., Dong, Y. & Wang, H. (2025). Development and validation of a depression risk prediction model . . . Evidence from CHARLS. *PLOS ONE* 20(9), e0332907.

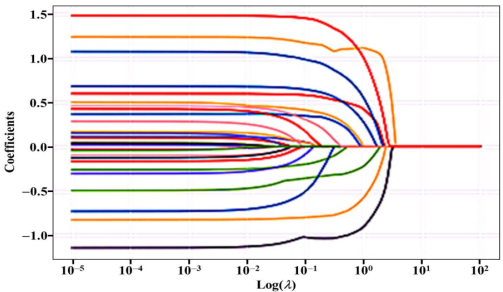
Curvilinear effects: too much engagement?



A management application of **polynomial regression**. Adding a quadratic term reveals an *inverted-U*: job performance rises with work engagement, then declines when engagement gets too high — a curvilinear pattern invisible to a purely linear model.

Shimazu, A., Schaufeli, W. B., Kubota, K., Watanabe, K. & Kawakami, N. (2018). Is too much work engagement detrimental? Linear or curvilinear effects on mental health and job performance. *PLOS ONE* 13(12), e0208684.

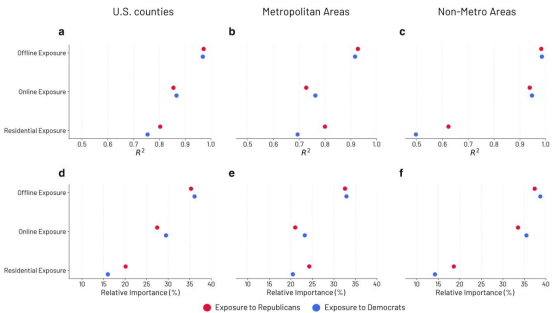
Regularized models of life satisfaction



Compares **lasso**, **ridge**, and other learners to predict older adults' life satisfaction. The coefficient-shrinkage path shows how increasing the penalty λ drives weak predictors to zero, leaving a compact, cross-validated predictive model.

Shen, X., Yin, F. & Jiao, C. (2023). Predictive models of life satisfaction in older people: A machine learning approach. *Int. J. Environ. Res. Public Health* 20(3), 2445.

Prediction as the objective: US voting



A political-science example where *prediction is the goal*. A learner predicts county partisan outcomes from demographic and mobility exposure; the figure ranks each dimension's contribution — physical partisan proximity outweighs online ties.

Tonin, M., Lepri, B. & Tizzoni, M. (2025). Physical partisan proximity outweighs online ties in predicting US voting outcomes. *PNAS Nexus* 4(10), pgaf308.

References

Works cited (1/2)

- Bach, F. R. (2008). Bolasso: Model consistent lasso estimation through the bootstrap. *ICML 25*.
- Belloni, A., Chernozhukov, V. & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies* 81(2), 608–650.
- Bowlsby, D., Chenoweth, E., Hendrix, C. & Moyer, J. D. (2020). The future is a moving target: Predicting political instability. *British Journal of Political Science* 50(4), 1405–1417.
- Kuhlmann, R. & Lewis, D. C. (2017). Legislative term limits and voter turnout. *State Politics & Policy Quarterly*.
- McDonald, M., Porter, R. & Treul, S. A. (2020). Running as a woman? Candidate presentation in the 2018 midterms. *Political Research Quarterly* 73(4), 967–987.
- Qin, M., Qiao, M., Dong, Y. & Wang, H. (2025). Development and validation of a depression risk prediction model for middle-aged and elderly adults with sensory impairment: Evidence from CHARLS. *PLOS ONE* 20(9), e0332907.
- Rhys, H. (2020). *Machine Learning with R, the tidyverse, and mlr*. Manning.
- Shen, X., Yin, F. & Jiao, C. (2023). Predictive models of life satisfaction in older people: A machine learning approach. *Int. J. Environ. Res. Public Health* 20(3), 2445.

Works cited (2/2)

- Shimazu, A., Schaufeli, W. B., Kubota, K., Watanabe, K. & Kawakami, N. (2018). Is too much work engagement detrimental? Linear or curvilinear effects on mental health and job performance. *PLOS ONE* 13(12), e0208684.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *JRSS-B* 58(1), 267–288.
- Tonin, M., Lepri, B. & Tizzoni, M. (2025). Physical partisan proximity outweighs online ties in predicting US voting outcomes. *PNAS Nexus* 4(10), pgaf308.
- van de Brake, H. J., Grow, A. & Dijkstra, J. K. (2017). Status inconsistency in groups. *Social Science Research* 64, 15–24.
- Verma, R., Mitra, B. & Chakraborty, S. (2019). Avoiding stress driving: Online trip recommendation from driving behavior prediction. *IEEE PerCom*, 1–10.
- Wang, Y. & Kosinski, M. (2018). Deep neural networks can detect sexual orientation from faces. *Journal of Personality and Social Psychology* 114(2), 246–257.
- Ye, T., Johnson, R., Fu, S., et al. & Ghani, R. (2019). Using machine learning to help vulnerable tenants in New York City. *ACM SIGCAS COMPASS*, 248–258.
- Zeng, Q., Li, G., Cui, Y., Jiang, G. & Pan, X. (2016). Estimating temperature–mortality exposure–response relationships at the multi-city level of China. *Int. J. Environ. Res. Public Health* 13(3), 279.
- Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *JRSS-B* 67(2), 301–320.