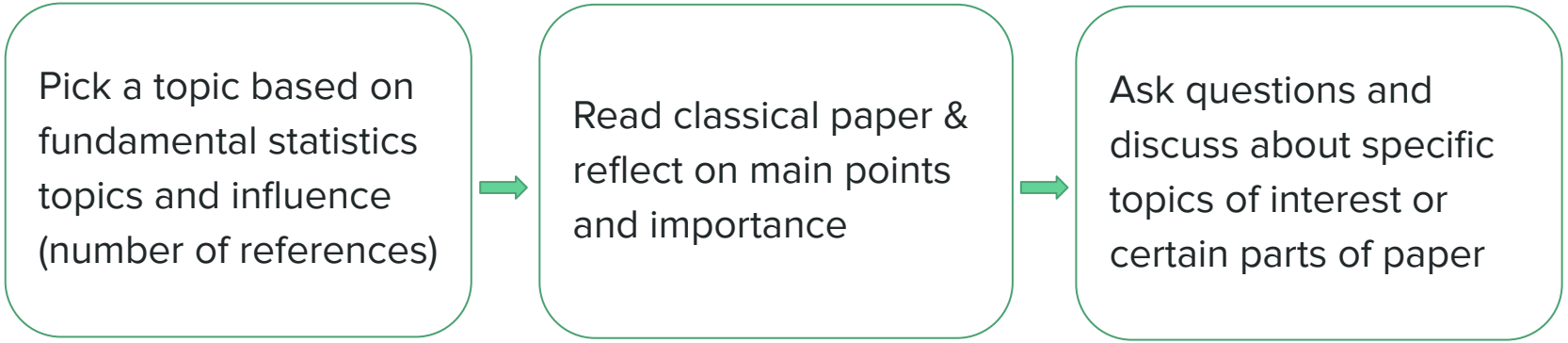
Classical Papers in Statistics

Mentor: Ethan Ancell
Mentee: Janice Kim

Organization of the DRP



- *Our purpose was not an analysis or complete understanding of paper but exploring and talking about interesting topics

Weekly Topics

1) *Regression shrinkage and selection via the lasso* (Tibshirani, R. , 1996) - Lasso Regression
2) *Support-vector networks* (Cortes, C., Vapnik, V., 1995) - Support Vector Machine
3) *Maximum Likelihood Estimation of Misspecified Models* (White, H., 1982)
4) *A new look at the statistical model identification* (H. Akaike, 1974) - Akaike information criterion
5) *Longitudinal data analysis using generalized linear models* (Liang, K.-Y., & Zeger, S. L., 1986)
6) *Nonparametric estimation from incomplete observations* (Kaplan, E. L., & Meier, P., 1958) - Kaplan-Meier Estimator for incomplete observations
7) *Controlling the false discovery rate: A practical and powerful approach to multiple testing* (Benjamini, Y., & Hochberg, Y., 1995)

Lasso Regression

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso

- Main Topics
  - Lasso Regression
  - Complementary Role of Lasso Regression compared to Ridge Regression, Subset Selection, and the garotte function
- Discussion Focus:
  - Two different way of Lasso Regression formula with Lagrange Multiplier
    - Constraint(optimization) or adding penalty
  - Discussion about the p-norms

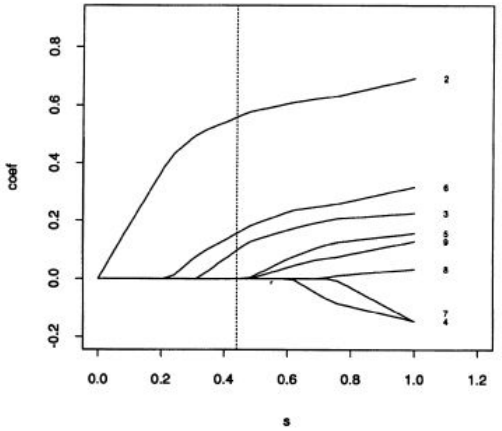


Fig. 2. Estimation picture for (a) the lasso and (b) ridge regression

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}|w_j|$$

$$(\hat{\alpha},\hat{\beta})=\arg\min\left\{\sum_{i=1}^{N}\left(y_i-\alpha-\sum_j\beta_j x_{ij}\right)^2\right\}\quad\text{subject to}\sum_j|\beta_j|\leqslant t.$$

Support Vector Machine

Cortes, C., Vapnik, V. (1995). Support-vector networks.

- Main Topics:
  - Optimal Hyperplane or soft margin in separable and inseparable training data
  - Efficiency and accuracy of SVM compared to other classification methods
- Discussion Topics:
  - Multivariate Normal Distribution
  - Alpha Weight Factor

$$\mathbf{w}_0 = \sum_{i=1}^{\ell} y_i \alpha_i^0 \mathbf{x}_i, \qquad \mathbf{w}_0 \cdot \mathbf{x} + b_0 = 0$$

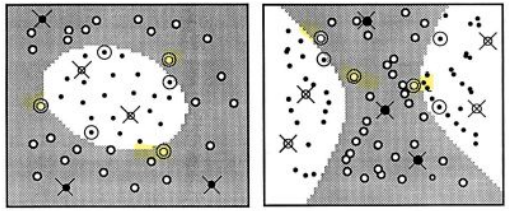


Figure 5. Examples of the dot-product (39) with $d = 2$. Support patterns are indicated with double circles, errors with a cross.

$$K(\mathbf{u},\mathbf{v}) = (\mathbf{u} \cdot \mathbf{v} + 1)^d$$

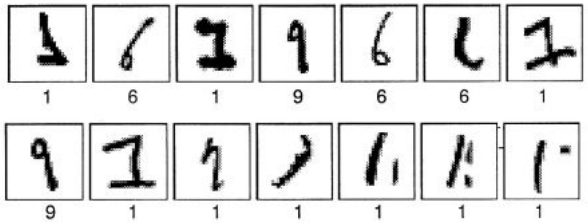


Figure 8. The 14 misclassified test patterns with labels for classifier 1. Patterns with label "1" are false negative. Patterns with other labels are false positive.

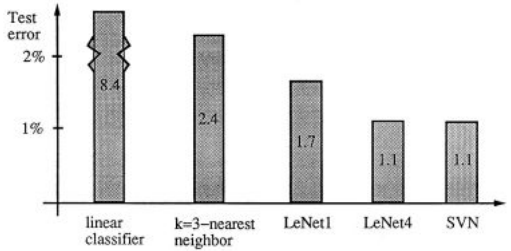
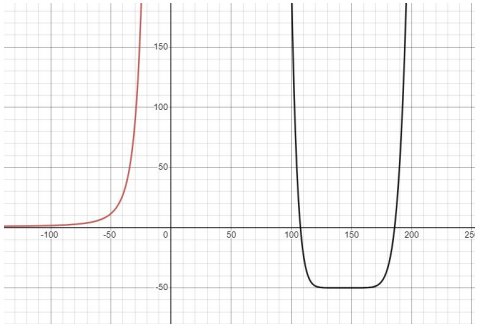


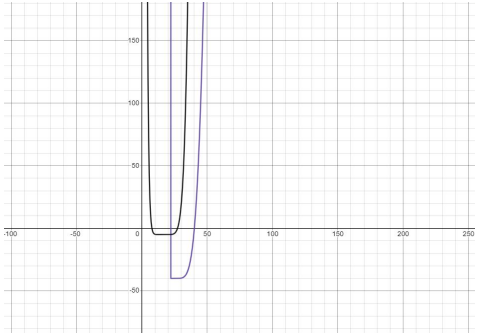
Figure 9. Results from the benchmark study.

Misspecification of the Model

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models

- Main Topic:
  - Misspecification between the assumed distribution(Family of distributions f(θ)) and the unknown true distribution G
- Discussion Topic: Weak Maximum Likelihood Estimator (WMLE)
  - Even when there is misspecification (i.e., g is not an element of f(θ)), the form of the MLE (WMLE) under certain conditions still has properties such as still being asymptotically normal and being consistent for the θ^* which minimizes the Kullback–Leibler divergence between f(θ) and g





Longitudinal Analysis and Generalized Linear Model

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models.

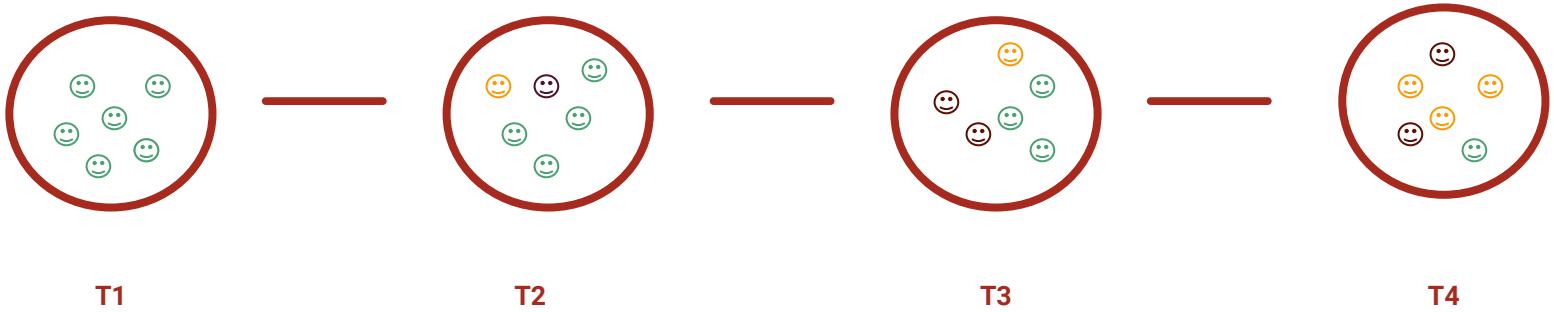
- Main Topic:
  - Generalized Linear Model
  - Longitudinal Data that time is not the primary interest
- Discussion Topic
  - Link Function of Generalized Linear Model (connection to a linear model)
  - Convergence of matrices using different types of norms/metrics

Family	Default Link Function
binomial	(link = "logit") $\mathbf{x}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$
gaussian	(link = "identity")
Gamma	(link = "inverse")
inverse.gaussian	(link = "1/mu^2")
poisson	(link = "log")
quasi	(link = "identity", variance = "constant")
quasibinomial	(link = "logit")
quasipoisson	(link = "log")

Kaplan-Meier Estimator

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association

- Main Topic:
  - Product Limit Estimator, Reduced Sample Method, and Actuarial method for incomplete observation with loss over time
- Discussion Topic
  - Naive Method vs. Product Limit Estimation Method



False Discovery Rate

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing

- Main Topic:
  - Bonferroni, Hochberg, and False Discovery Rate Controlling with formulas
  - Simulations for power testing for different methods
- Discussion Topic
  - Difference between False Discovery Rate and Family-Wise Error Rate

TABLE 1
Number of errors committed when testing m null hypotheses

	Declared non-significant	*Declared significant*	*Total*
True null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	$m - m_0$
	$m - R$	R	m

Key Takeaways

- Great opportunity to review fundamental statistical concepts and know where they came from
- Reading some classical statistics papers is still not easy (can never be?) but became less intimidated through this DRP
- Interesting to find relevant concepts and connections to my past learnings
- Knowledge stacks up

Thank you!

Any Questions?