Imprecise Regularization

Tathagata Basu

supervised by
J. Einbeck and M. C. M. Troffaes

WPMSIIP 2018, July 30 – August 1, Oviedo, Spain

Durham University
THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION HORIZON 2020
MARIE CURIE ACTIONS

Objective and Framework

- To formulate an imprecise regularization technique
- Use of cross–validation as a link between regularization methods and credal classification.
- Proposal of other possible approach.
  - Use of Gaussian assumption
  - Use of weights.

Outline

Linear Models

Let $\boldsymbol{X}$ be a set of predictors (attributes) and $\boldsymbol{Y}$ be the corresponding response (classes). The linear model is given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where

$$\boldsymbol{Y} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} := \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \quad \boldsymbol{\beta} := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\epsilon} := \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{2}$$

$\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ are error terms, $\boldsymbol{\beta}$ are regression coefficient.

Regression Models

- **Ordinary Least Squares**

$$\hat{\beta}^{\text{OLS}} := \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{Y} - \boldsymbol{X}\beta\|_2^2 = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad (3)$$

Regression Models

- **Ordinary Least Squares**

$$\hat{\beta}^{\text{OLS}} := \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{Y} - \boldsymbol{X}\beta\|_2^2 = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad (3)$$

- **Issues with OLS**
  - Overfitting Problem
  - $p > n$

Regression Models

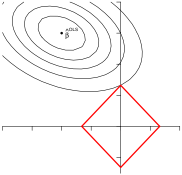
- **Ordinary Least Squares**

$$\hat{\beta}^{\text{OLS}} := \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{Y} - \boldsymbol{X}\beta\|_2^2 = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \quad (3)$$

- **Issues with OLS**
  - Overfitting Problem
  - $p > n$
- **Regularization $\rightarrow$ LASSO**

$$\hat{\boldsymbol{\beta}}_\lambda = \arg\min_{\boldsymbol{\beta}} \left(\frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right) \quad (4)$$

Graphical Interpretation



- LASSO as constrained optimization problem
- other penalty terms
  - non-convex for $q < 1$
  - $q = 1$ is smallest value for convex region

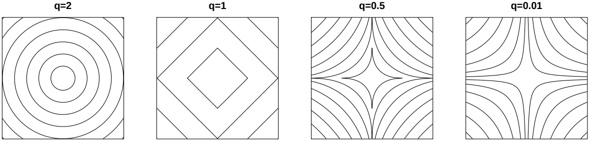


Figure: different penalty terms

A Basic Example

Gaia dataset to formulate 3–d mapping of space.

- number of observation, $n = 8286$
- number of predictors (wavelength bands), $p = 16$
- steller temperature as response

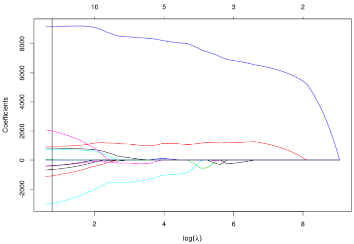


Figure: Coefficient path

- LASSO estimates around black vertical line

Classification

- Let $C = (c_1, c_2, \cdots, c_m)$ be a classification variable defined on $\mathcal{C}$
- $A_1, A_2, \cdots, A_n$ be set of attributes having values $a_1, a_2, \cdots, a_n$ defined on $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$.

We calculate the joint probability $P[C, A_1, A_2, \cdots, A_n]$.

Naive Bayes Classifier

- Naive Bayes Classifier

$$P[A_1, A_2, \cdots, A_n|C] = \prod_{i=1}^{n} P[A_i|C] \quad (5)$$

- Modified joint probability

$$P[C, A_1, A_2, \cdots, A_n] = P[C]. \prod_{i=1}^{n} P[A_i|C] \quad (6)$$

Imprecise Dirichlet Model

For hyperparameter t and a constant $s > 0$, we have

$$f(x|s,t) \propto \prod_{c \in \mathcal{C}} \left[x_c^{st(c)-1} \prod_{i=1}^{n} \prod_{a_i \in \mathcal{A}_i} x_{a_i|c}^{st(a_i|c)-1} \right] \quad (7)$$

subject to the following constraints

$$\sum_{c} t(c) = 1 \quad (8)$$

$$\sum_{a_i \in \mathcal{A}_i} t(a_i|c) = t(c) \quad \forall (i,c) \quad (9)$$

$$t(a_i|c) > 0 \quad (i, a_i, c) \quad (10)$$

Naive Credal Classifier

Naive credal classifier is based on the assumptions of NBC and use of IDM as prior which gives us –

$$E[x_{c,\boldsymbol{a}}|n,t] = P[c,\boldsymbol{a}|n,t] = P[c|n,t]\prod_{i=1}^{n} P[a_i|c,n,t] \qquad (11)$$

where,

$$P[c|n,t] = E[x_c|n,t] = \frac{n(c)+st(c)}{N+s} \qquad (12)$$

$$P[a_i|c,n,t] = E[x_{a_i|c}|n,t] = \frac{n(a_i|c)+st(a_i|c)}{n(c)+st(c)} \qquad (13)$$

Credal Dominance

A class c' dominates c'' ($c', c'' \in \mathcal{C}$) iff $P[c'|\boldsymbol{a}, n, t] > P[c''|\boldsymbol{a}, n, t]$ for all values of t.

$$\inf_{t} \frac{P[c'|\boldsymbol{a}, n, t]}{P[c''|\boldsymbol{a}, n, t]} \tag{14}$$

subject to

$$\sum_{c} t(c) = 1 \tag{15}$$

$$0 < t(a_i|c) < t(c) \quad \forall (i, c) \tag{16}$$

Cross–validation

LASSO

- λ as tuning parameter
- mean-squared error as measure of accuracy

NCC

- s as tuning parameter
- different accuracies
  - Determinacy
  - Single Accuracy
  - Indeterminate Set–Size
  - Set Accuracy

Example

Sonar Dataset

- Binary Classification Problem
- 60 attributes
- 208 observations

Example

Sonar Dataset

- Binary Classification Problem
- 60 attributes
- 208 observations

Naive approach – Feature Selection

- Apply LASSO for feature selection
- Credal classification on the selected features

Example

Feature selection using LASSO

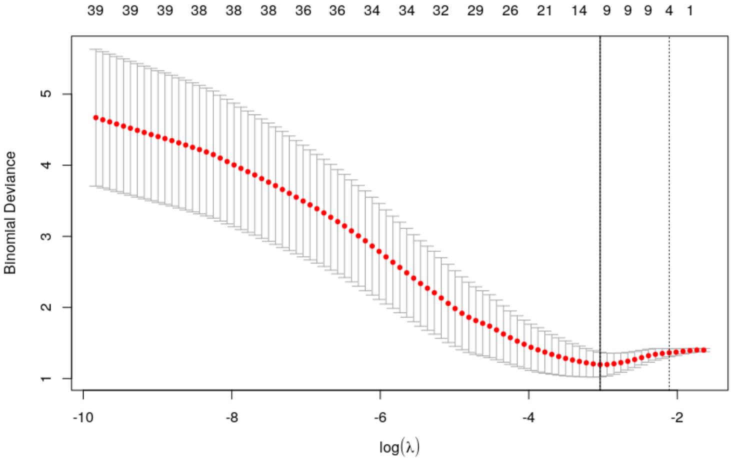


Figure: Cross–validation Curve

Example

S as tuning parameter

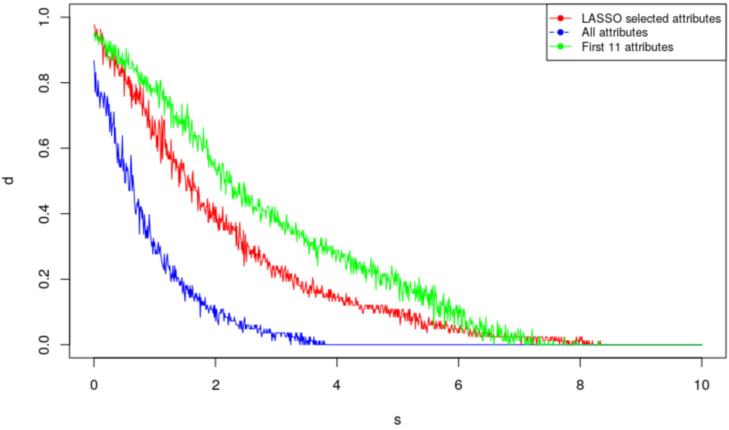


Figure: Cross–validation Curve for Classification

Example

S as tuning parameter

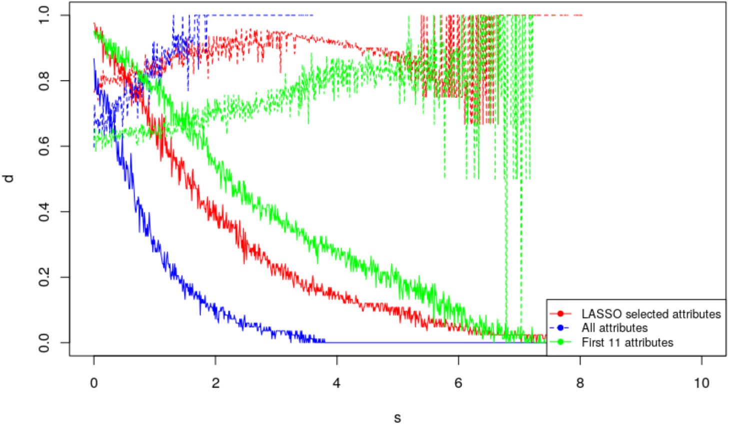


Figure: Cross–validation Curve for Classification

Example

S as tuning parameter

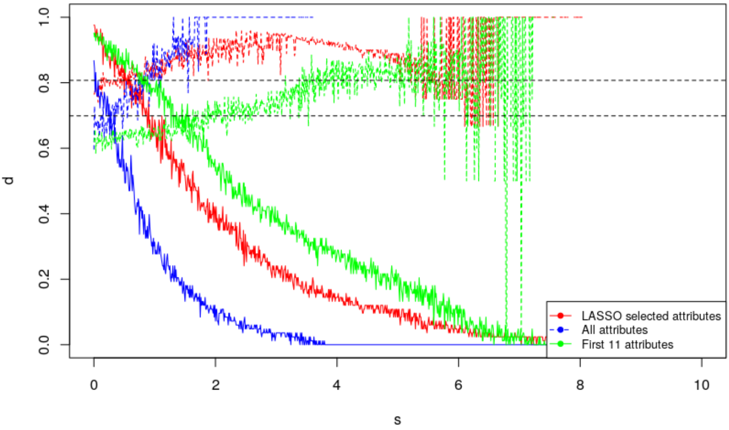


Figure: Cross–validation Curve for Classification

Example

S as tuning parameter

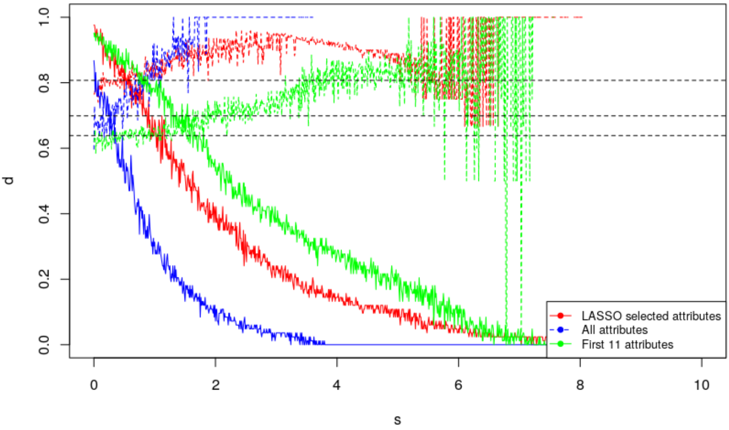


Figure: Cross–validation Curve for Classification

Possible Approaches

- Gaussian Naive Bayes assumption
  - logistic regression as classification problem
  - use of credal classification in logistic–LASSO setting
  - simultaneous cross–validation

Possible Approaches

- Gaussian Naive Bayes assumption
  - logistic regression as classification problem
  - use of credal classification in logistic–LASSO setting
  - simultaneous cross–validation
- Hierarchical Bayes
  - imprecise weights on the hyper parameter of penalty term

Conclusions and Future Work

Conclusion

- Cross–validation as a tool
- Possible Approaches

Questions

- Shrinking regression co-efficients in GNB setting
- Simultaneous cross–validation : (Chicken and egg)

Conclusions and Future Work

Conclusion

- Cross–validation as a tool
- Possible Approaches

Questions

- Shrinking regression co-efficients in GNB setting
- Simultaneous cross–validation : (Chicken and egg)

Thank You