LA-UR- 11-03972

Approved for public release;
distribution is unlimited.

Title: Selecting Linear Models Under The Bayesian Paradigm With Focus On Good Prediction Over A User-Specified Distribution On The Covariate Space

Author(s): A.L. Pintar
C.M. Anderson-Cook
H. Wu

Intended for: Joint Statistical Meetings of the American Statistical Association, Miami, FL August 2, 2011

Los Alamos
NATIONAL LABORATORY
EST. 1943

Selecting Linear Models under the Bayesian Paradigm with Focus on Good Prediction over a User-Specified Distribution on the Covariate Space

Adam L. Pintar[1], Christine M. Anderson-Cook[2], Huaiqing Wu[3]

[1]National Institute of Standards & Technology
[2]Los Alamos National Laboratory
[3]Iowa State University

Model selection is an important part of building linear regression models under the Bayesian paradigm. If unimportant explanatory variables are included, interesting posterior distributions have inflated variance. If important explanatory variable are excluded, interesting posterior distributions can miss their (unknown) target. Several model selection algorithms exist for linear models under the Bayesian paradigm. For instance, one could choose the model with the smallest deviance information criterion, the model with the largest posterior probability, or the model whose terms all have posterior probability greater than 0.5. A common theme to all of these methodologies is that they consider only the observed data. We propose a model selection methodology that focuses on good prediction over a user-specified distribution on the covariate space. Our methodology quantifies the prediction ability of all models under consideration at many covariate points sampled from the user-specified distribution. Then, models are graphically compared based on their distribution of prediction abilities. The methodology is illustrated via an example, and a simulation study highlighting its potential is presented.

Selecting Linear Models Under The Bayesian Paradigm With Focus On Good Prediction Over A User-Specified Distribution On The Covariate Space

Adam L. Pintar[1], Christine M. Anderson-Cook[2], And Huaiqing Wu[3]

[1]National Institute Of Standards And Technology

[2]Los Alamos National Laboratory

[3]Iowa State University

August 3, 2011

Motivation

- The goal of model building should be incorporated into the model selection process
- Consider a population of batteries:
  - What level of performance can we expect in 6 months?
  - Extrapolating inflates variance
  - A model with less terms may be preferred
- Statisticians know well the dangers of extrapolation
- When possible, extrapolation should be based on underlying scientific or engineering understanding
- The model selection method is not restricted to extrapolation

Related Work

- Model selection
  - Deviance information criterion
  - Stochastic search variable selection
  - Median probability model
  - Rank with posterior probabilities
- Model averaging
- Graphical tools used in experiment design literature
  - Boxplots
  - Fraction of design space plots

Procedure Overview

1. Characterize the relationship between covariates, and use that characterization, as well as the study goal, to select the covariate distribution of interest (DI)
2. Randomly sample new points from the DI
3. Calculate a statistic on which comparisons are based, at all newly sampled points for all models considered
4. Compare models numerically and graphically to select a best model

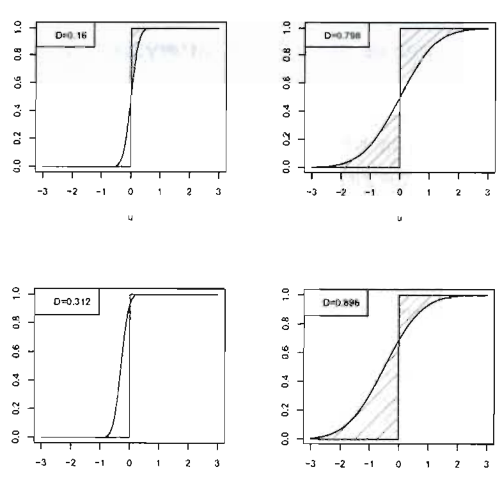
The Measure of Prediction Ability

- The best possible posterior distribution for prediction is a point mass at $\mu(\mathbf{x}_{new})$, the true mean, if it is known
  - $\mathbf{x}_{new}$ is a sampled point from the DI
- Let $F_{\mathbf{x}_{new}}$ be a cumulative distribution function (cdf) representing a point mass at $\mu(\mathbf{x}_{new})$
  - $F_{\mathbf{x}_{new}}$ steps from 0 to 1 at $\mu(\mathbf{x}_{new})$
- Let $F^m_{\mathbf{x}^m_{new}}$ be the posterior cdf of $\mu^m(\mathbf{x}^m_{new})$
- The discrepancy between the cdfs can be quantified by the following expression

$$D^k_m(\mathbf{x}^m_{new}) = \left\{ \int_{-\infty}^{\infty} \left| F^m_{\mathbf{x}^m_{new}}(u) - F_{\mathbf{x}_{new}}(u) \right|^k du \right\}^{\frac{1}{k}}$$

- Need a surrogate for $\mu(\mathbf{x}_{new})$

The Measure of Prediction Ability (continued)



A Surrogate for $\mu(\mathbf{x}_{new})$

- Let $P(M = m|\mathbf{y})$ be the posterior probability of model m
- Let $\hat{\mu}^m(\mathbf{x}^m_{new})$ be a point prediction from the posterior distribution of $\mu^m(\mathbf{x}^m_{new})$
- The weighted average

$$\hat{\mu}(\mathbf{x}_{new}) = \sum_{i=1}^{N_{mod}} P(M = i|\mathbf{y})\hat{\mu}^m(\mathbf{x}^m_{new})$$

  can be used as a surrogate for $\mu(\mathbf{x}_{new})$
- Options for calculating $P(M = m|\mathbf{y})$
  - The BIC approximation

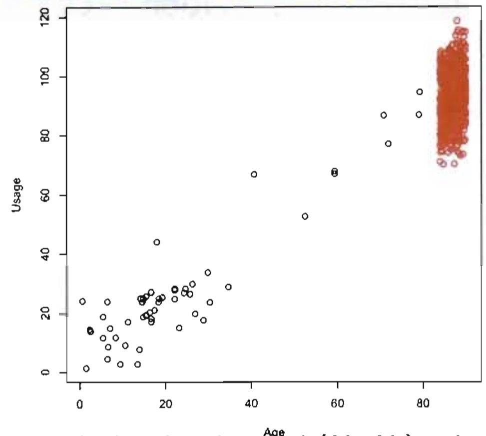
$$P(M = m|\mathbf{y}) \approx \frac{\exp\{-\frac{1}{2}\mathrm{BIC}_m\}}{\sum_{i=1}^{N_{mod}} \exp\{-\frac{1}{2}\mathrm{BIC}_i\}}$$

  - Using an MCMC algorithm
    - Carlin and Chib (1995)
    - Dellaportas et al. (1998)
    - Reversible Jump MCMC

Example Introduction

- Response
  - Y = continuous measure of battery performance
- Two covariates
  - X_1 = Age
  - X_2 = Usage in time in ready mode
- Because of the proprietary nature of the data, they have been rescaled
- Goal
  - The observed X_1 values are between 0.54 and 84.77
  - Predict future reliability for $X_1 \in [25, 30]$
- Full model
  - $Y \sim \mathrm{N}(\mu, \sigma^2)$
  - $\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$
  - Assign a flat prior to the regression coefficients and a Jefferys prior to σ^2
  - Number of models, $N_{mod} = 2^5 = 32$

Covariate Distribution of Interest

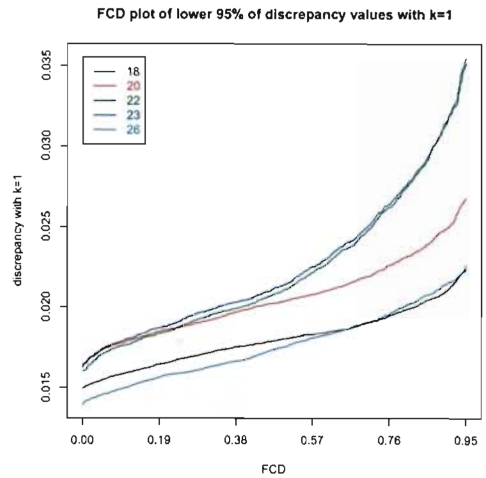


- Black circles depict the observed (X_1, X_2) pairs
- Red circles depict points sampled from the DI
- Positive trend

Table of Summary Statistics

# terms	Model	95th Percentile value	95th Percentile rank	Mean value	Mean rank
0	(1)	0.15294	1(32)	0.12605	1(32)
1	X_1 (17)	0.08074	1(18)	0.05134	2(21)
1	X_2 (9)	0.08341	2(20)	0.03808	1(16)
2	X_1, X_1X_2 (18)	0.02236	1(1)	0.01833	1(2)
2	X_1, X_1^2 (21)	0.03737	2(7)	0.02367	2(7)
2	X_2, X_2^2 (11)	0.04329	3(11)	0.02637	3(12)
2	X_1, X_2^2 (19)	0.04615	4(15)	0.03164	4(15)
3	X_1, X_2, X_1X_2 (26)	0.0226	1(2)	0.01794	1(1)
3	X_1, X_2^2, X_1X_2 (20)	0.02674	2(3)	0.02097	2(3)
3	X_1, X_1^2, X_2^2 (23)	0.03512	3(4)	0.02336	4(5)
3	X_1, X_1^2, X_1X_2 (22)	0.03553	4(5)	0.02316	3(4)
4	X_1, X_2, X_2^2, X_1X_2 (28)	0.03582	1(6)	0.02379	1(8)
4	X_1, X_2, X_1^2, X_2^2 (31)	0.0392	2(8)	0.0247	3(10)
5	X_1, X_2, X_1^2, X_2^2, X_1X_2 (32)	0.04451	1(14)	0.02695	1(14)

Fraction of Covariate Distribution (FCD) Plot



- Choose model 26 (X_1, X_2, and X_1X_2) as the best (small lues desirable)

Results From Other Selection Procedures

# terms	model	DIC
3	X_1, X_2, X_2^2 (27)	-208.2
4	X_1, X_2, X_1X_2 (26)	-207.8
2	X_1, X_1^2 (21)	-207.2
4	X_1, X_1X_2 (18)	-206.9
3	X_1, X_2, X_1^2, X_2^2 (31)	-206.5
		Posterior Probability
2	X_1, X_1^2 (21)	0.24
3	X_1, X_1X_2 (18)	0.21
2	X_1, X_2, X_2^2 (27)	0.16
3	X_1, X_2, X_1X_2 (26)	0.13
3	X_1, X_2^2, X_1X_2 (20)	0.04

term	X_1	X_2	X_1^2	X_2^2	X_1X_2
Posterior Probability	1.0	0.40	0.40	0.32	0.47

- Model 27 leads to the smallest DIC
- Model 21 has the highest posterior probability
- The median probability model is model 17

Conclusion

- The focus of the new method is good prediction over a user-specified distribution of interest (DI) on the covariate space
- General four-step algorithm
  - Select the DI
  - Randomly sample points from the DI
  - Calculate the measure of prediction ability at each sampled location for all models under consideration
  - Compare models numerically and graphically based on the measures of prediction ability
- The DI was chosen to match the study goal of good prediction of future battery performance
- Different models may be preferred for different DI's