Mean Squared Error in Model Selection

Adam L. Pintar\textsuperscript{1,2}, Christine M. Anderson-Cook\textsuperscript{2}, Huaiqing Wu\textsuperscript{1}

\textsuperscript{1}Department of Statistics, Iowa State University

\textsuperscript{2}Statistical Sciences, Los Alamos National Laboratory

June 3, 2009

LA-UR: 09-03286
Some existing work in variable selection
 Goal of methodology
 Our algorithm via an example
 A real data example
 Discussion and conclusions
Many Methods of Variable selection

- Some of the most popular methods
  - AIC (Akaike (1974))
  - BIC (Schwarz (1978))
  - Cross Validation (Shao (1993))
  - Mallows $C_p$ (Mallows (1973))
  - Adjusted $R^2$
  - Stepwise Selection
  - Stochastic Search Variable Selection (George and McCulloch (1993))

- All consider only the observed data
Goal

- Consider probit regression models
- \( p \) - True system reliability
- \( \hat{p}^m \) - Estimated system reliability under model \( m \)
- Provide a variable selection algorithm, focused on prediction, in a user defined region of the covariate space
Overview of Our Algorithm

- Select and characterize the user-specified region of interest in the covariate space
- Randomly sample new locations from the region of interest
- Estimate prediction bias, prediction variance, and prediction MSE at all sampled locations for all models to be compared
- Compare models graphically, based on the estimated values in the previous step to select a best model
  - Focus is on MSE
Why Focus on MSE?

- Ideal to simultaneously minimize prediction variance and prediction bias
- \[ \text{MSE}(\hat{p}^m) = \text{variance}(\hat{p}^m) + \text{bias}(\hat{p}^m)^2 \]
- MSE balances variance and bias, which is a compromise to minimizing both
- One issue
  - \[ \text{bias}(\hat{p}^m) = E(\hat{p}^m) - p \]
  - Need a surrogate for \( p \)
  - The estimated reliability from the full model is used
Example 1 Introduction

- Two covariates $X_1$ and $X_2$
  - $(X_2 \mid X_1) \sim N(10 \times X_1, 100)$

![Diagram showing the relationship between $X_1$ and $X_2$]

- The response $Y_i \sim Bernoulli(p_i)$
  - $\Phi^{-1}(p_i) = 2.3 - 0.1 \times X_1 - 0.02 \times X_2$
Selecting a Region of the Covariate Space

- Suppose we are interested in predicting reliability for $X_1 \in [0, 10]$
Characterizing the Relationship Between Covariates and Sampling

- Simple linear regression for characterization
- Sampling
  - Regress $X_2$ on $X_1$ using the observed points.
  - Sample $X_1 \in [0, 10]$ uniformly.
  - For every sampled $X_1$ value, sample $X_2 \sim N(b_0 + b_1 \cdot X_1, \hat{\sigma}^2)$
  - Here, $b_0 = 0.76$, $b_1 = 9.97$, and $\hat{\sigma}^2 = 97.73$
Calculation Details

- For each model under consideration, at each sampled point calculate bias\(^2\), variance, and MSE
- \(\hat{\beta}^f\) - estimated regression coefficients for the full model
- \(\hat{\beta}^m\) - estimated regression coefficients for model \(m\)
- \(\hat{p} = \Phi(x'\hat{\beta}^f)\) - estimated true system reliability under the full model at covariate location \(x\)
- \(\hat{p}^m = \Phi(x'\hat{\beta}^m)\) - estimated reliability under model \(m\) at covariate location \(x\)
- \(\hat{\text{Var}}(\hat{p}^m) = \left[(\frac{\partial p^m}{\partial \beta})'\right]\hat{\beta}^m \hat{\Sigma}_{\beta} \left[(\frac{\partial p^m}{\partial \beta})\right] \beta^m = \hat{\beta}^m\)
- \(\hat{\text{bias}}^m = \hat{p}^m - \hat{p}\)
- \(\hat{\text{MSE}}^m = (\hat{\text{bias}}^m)^2 + \hat{\text{Var}}(\hat{p}^m)\)
- Note that bias is estimated as zero for the full model
## Naming the Models

<table>
<thead>
<tr>
<th>Model</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_1 \times X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure: Boxplots of MSE, bias$^2$, and variance
Fraction of Design Space (FDS) Plots

**Figure:** FDS curves of the four best models with respect to MSE. (Zahran et al. 2003).
Quantile-Quantile Plots

Q–Q plot of MSE distributions for models 3 and 7
Results from AIC and Cross Validation

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Model Terms</th>
<th>AIC Value</th>
<th>Cross validation $\hat{\Gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>none</td>
<td>1369.741</td>
<td>0.2470</td>
</tr>
<tr>
<td>2</td>
<td>$X_1 \times X_2$</td>
<td>802.9641</td>
<td>0.1278</td>
</tr>
<tr>
<td>3</td>
<td>$X_2$</td>
<td>784.1065</td>
<td>0.1261</td>
</tr>
<tr>
<td>4</td>
<td>$X_2, X_1 \times X_2$</td>
<td>785.8046</td>
<td>0.1269</td>
</tr>
<tr>
<td>5</td>
<td>$X_1$</td>
<td>807.8247</td>
<td>0.1305</td>
</tr>
<tr>
<td>6</td>
<td>$X_1, X_1 \times X_2$</td>
<td>797.5526</td>
<td>0.1287</td>
</tr>
<tr>
<td>7</td>
<td>$X_1, X_2$</td>
<td>784.2477</td>
<td>0.1271</td>
</tr>
<tr>
<td>8</td>
<td>$X_1, X_2, X_1 \times X_2$</td>
<td>786.2079</td>
<td>0.1272</td>
</tr>
</tbody>
</table>

- True model
  - $\Phi^{-1}(p_i) = 2.3 - 0.1 \times X_1 - 0.02 \times X_2$

- Our method identifies correct model

- Standard methods emphasize a model without $X_1$
  - $X_1$ is only observed at three distinct values
Example 2 Introduction

► Background
  ▶ The responses are pass/fail results collected from a missile system
  ▶ The available covariates are age in years, and usage in hours in ready mode
  ▶ Due to the proprietary nature of the full systems, the actual pass/fail results for individual systems has been adjusted
  ▶ Use a probit regression model to describe the data
► Characterizing the relationship between age and usage
  ▶ Start with a scatter plot
The Relationship Between age and usage

(a) Scatter Plot of Age Versus Usage, with Regression Line
(b) Scatter Plot of Age Versus $\sqrt{\text{Usage}}$, with Regression Line
Characterizing the Covariate Space, Choosing the Region, and Sampling

- Linear regression is an appropriate description
- Specifying a range for age describes the region
- Decisions are made based on the prediction of the future
  - Extrapolation is required
  - Scientific and engineering understanding
- The observed range of age is about 2 to 25 years.
- Suppose interest is in making prediction for age in 24 to 30 years
- Sampling
  - Sample age randomly in 24 to 30 years
  - Sample $\sqrt{usage}$ according to the linear regression for each sampled value of age
Figure: Boxplots of MSE, bias$^2$, and variance.
Results from AIC and Cross Validation

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Model Terms</th>
<th>AIC Value</th>
<th>Cross validation $\hat{\jmath}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>none</td>
<td>181.9922</td>
<td>0.0854</td>
</tr>
<tr>
<td>2</td>
<td>$X_1 \times X_2$</td>
<td>183.6414</td>
<td>0.0864</td>
</tr>
<tr>
<td>3</td>
<td>$X_2$</td>
<td>170.1779</td>
<td>0.0823</td>
</tr>
<tr>
<td>4</td>
<td>$X_2$, $X_1 \times X_2$</td>
<td>156.3253</td>
<td>0.0793</td>
</tr>
<tr>
<td>5</td>
<td>$X_1$</td>
<td>176.2935</td>
<td>0.0849</td>
</tr>
<tr>
<td>6</td>
<td>$X_1$, $X_1 \times X_2$</td>
<td>168.5032</td>
<td>0.0843</td>
</tr>
<tr>
<td>7</td>
<td>$X_1$, $X_2$</td>
<td>159.7092</td>
<td>0.0826</td>
</tr>
<tr>
<td>8</td>
<td>$X_1$, $X_2$, $X_1 \times X_2$</td>
<td>157.9798</td>
<td>0.0825</td>
</tr>
</tbody>
</table>

- AIC and cross validation highlight the same best model
  - These do not consider extrapolation
- Our method chooses a smaller model
  - The extra variance caused by including correlated terms overtakes the reduction in bias in the extrapolation case
Discussion & Conclusions

- Recap of Algorithm
  - Select and characterize region of interest
  - Randomly sample new locations from the region
  - Calculate MSE, $\text{bias}^2$, and variance
  - Compare models graphically
- How model will be used should influence selection procedure
- Characterization of the covariate space is key
- Able to deal with correlation between explanatory variables
- Extendable to other model forms


