Variable Selection for Modeling the Absolute Magnitude at Maximum of Type Ia Supernovae

Makoto Uemura, Koji, S. Kawabata (Hiroshima Univ.),
Shiro Ikeda (The Institute of Statistical Mathematics),
Maeda Keiichi (Kyoto Univ.)
2015, PASJ, submitted.

Abstract

To date, many parameters have been proposed for the explanatory variables of the absolute magnitude at maximum of Type Ia supernovae (SNIa). Here, we present a novel method in which we use the cross-validation to control the generalization error, and a LASSO-type estimator to choose the set of variables. Our variable selection approach does not support to add any other variables to the classical model with the color and decay rate.

Introduction

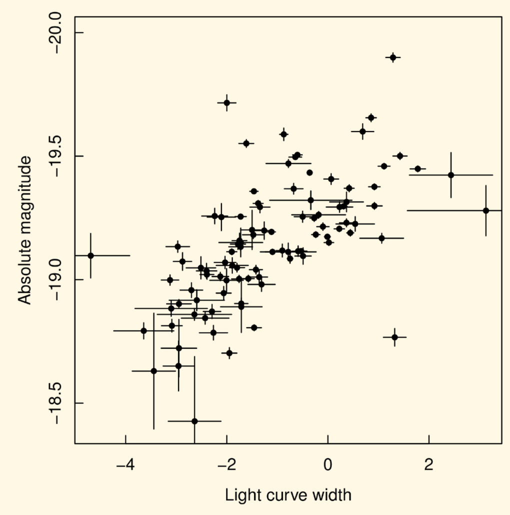
$M_B = \beta_1(\text{color}) + \beta_2(\text{decay rate}) + \text{any other???}$

Searching for explanatory variables of the absolute magnitude (M_B) of Type Ia supernovae

- ✓ Previously proposed variables:
  - ➢ Classical parameters = color (B-V or "c") & decay rate (Δm_{15} or "x1") [1]
  - ➢ Equivalent width (EW). Depth of absorption lines
  - ➢ Ratio of EWs. Ratio of depths of different lines
  - ➢ Arbitrary flux-ratios [2][3][4]

→ More than the number of samples!
(, but the number of significant variables should be small.)

→ This is a problem of "variable selections" in statistics and machine learning.



Clear correlation between M_B and the light-curve width, "x1", which is believed to be a "classical" explanatory variable of M_B.

Our method

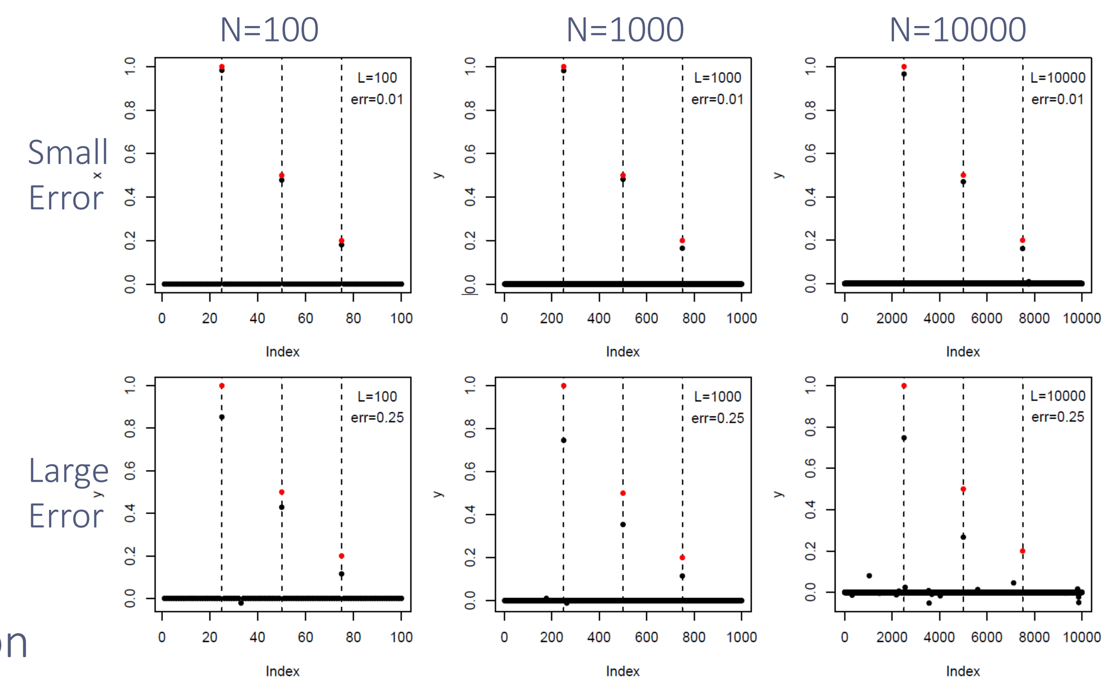
LASSO (Least Absolute Shrinkage and Selection Operator [5])

$$\hat{\beta}_\lambda = \arg\min_{\beta} \{\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\}$$

- ✓ y : Target variable, having M elements. The absolute magnitude, MB.
- ✓ X : Explanatory variables. A MxN matrix. The color, decay rate, and any other measurements proposed to date.
- ✓ β : Coefficients of explanatory variables, having N elements.

l1 norm ($||\beta||_1 = \sum|\beta|$) minimization

→ Reconstruction of sparse vectors (even in the case of N>M)



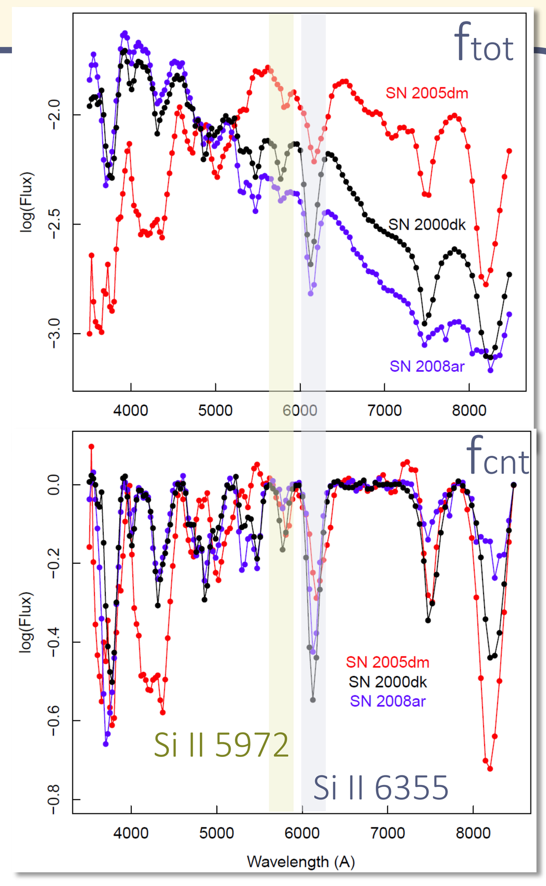
Demonstration of LASSO. We assumed three non-zero signals (red) among N=10^2,3,4 variables. We estimated them from 50 target variables which were generated with a random matrix.

Data & Explanatory variables

- ✓ All data were from "Berkeley Supernova Database".[6]
- ✓ 78 samples (the absolute magnitude, M_B)
- ✓ 276 candidates of explanatory variables
  - ➢ Color (c), light-curve width (x1)
  - ➢ Spectra normalized by the total flux (f_{tot})
  - ➢ Spectra normalized by the continuum (f_{cnt})
  - ➢ Previously proposed flux-ratios (R) [2][3][4]

Our model of the absolute magnitude

$$\begin{aligned} M_B &= \beta_1 c + \beta_2 x_1 \\ &+ \beta_3 f_{\rm tot}(3512) + \beta_4 f_{\rm tot}(3534) + \cdots + \beta_{136} f_{\rm tot}(8472) \\ &+ \beta_{137} f_{\rm cnt}(3512) + \beta_{138} f_{\rm cnt}(3534) + \cdots + \beta_{270} f_{\rm cnt}(8472) \\ &+ \beta_{271}\mathcal{R}(3780/4580) + \beta_{272}\mathcal{R}(4610/4260) \\ &+ \beta_{273}\mathcal{R}(5690/5360) + \beta_{274}\mathcal{R}(6420/4430) \\ &+ \beta_{275}\mathcal{R}(6420/5290) + \beta_{276}\mathcal{R}(6630/4400) + e. \end{aligned}$$



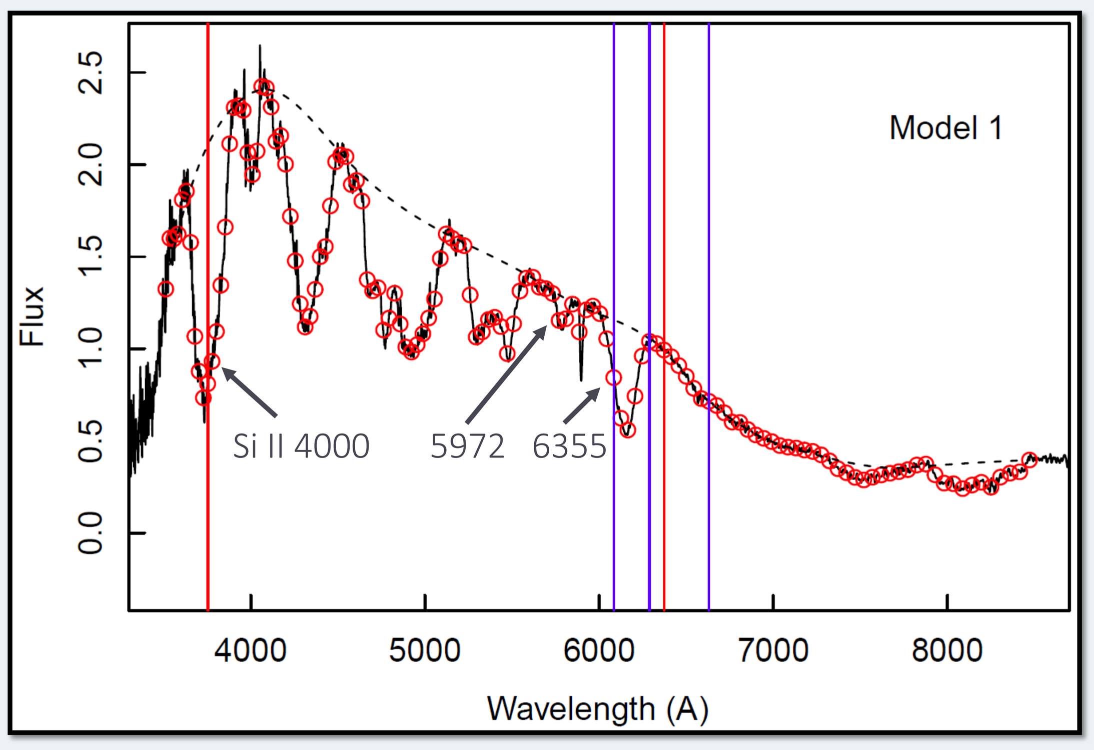
Examples of (up) total-flux normalized spectra, and (low) continuum-normalized spectra.

Results

Model 1

(Using all candidates of variables)

Target variable: M_B
Explanatory variables: c, x1, f_{tot}, f_{cnt}, R



Selected variables in Model 1 are indicated with the vertical lines with the typical spectra (SN 2006et). Red: f_{tot}. Blue: f_{cnt}.

Selected variables (in order of large coefficients):

- ➢ c (color)
- ➢ f_{tot}(6373) → correlation with "c"?
- ➢ x1 (light-curve width)
- ➢ f_{cnt}(6084) → Si II (6355) ?
- ➢ f_{cnt}(6289) → Si II (6355) ?
- ➢ f_{cnt}(6631) → continuum (false signal)
- ➢ R(3780/4580) → consistent with [4]
- ➢ f_{tot}(3752) → Si II (4000) ?

✓ f_{tot}(6373) & R(3780/4580) may be significant variables. But, they have a positive correlation with the color, "c".
→ Using color-corrected M_B → Model 2

✓ Several variables are related to the Si II lines. The Si II lines may provide good variables, while it is well known that the light curve width, "x1" shows a correlation with those lines.[7]
→ Model for "x1" → Model 3

Model 2

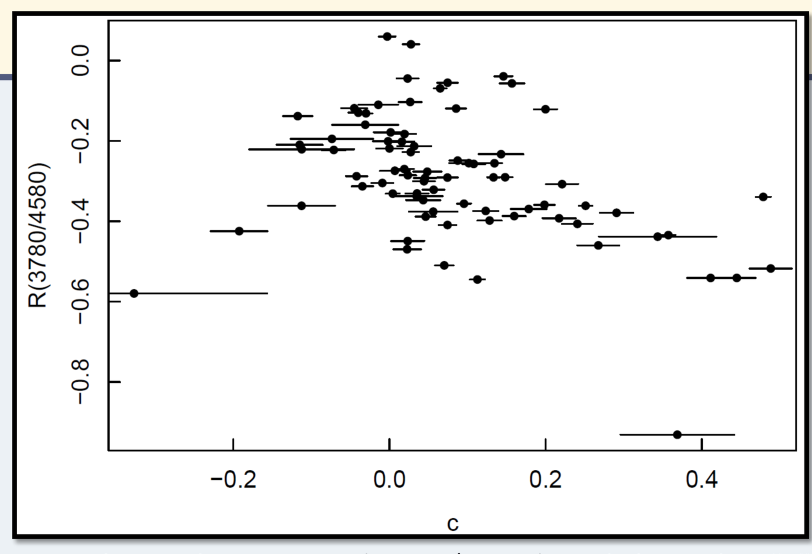
(Color correction for M_B)

Target variable: ($M_B - \beta c$)
Explanatory variables: x1, f_{tot}, f_{cnt}, R

Selected variables:

- ✓ x1 (light-curve width)
- ✓ No other variables were selected.

→ f_{tot}(6373) and R(3780/4580) do not significantly revise the model.
Their selection in Model 1 was due to their high correlations with "c".



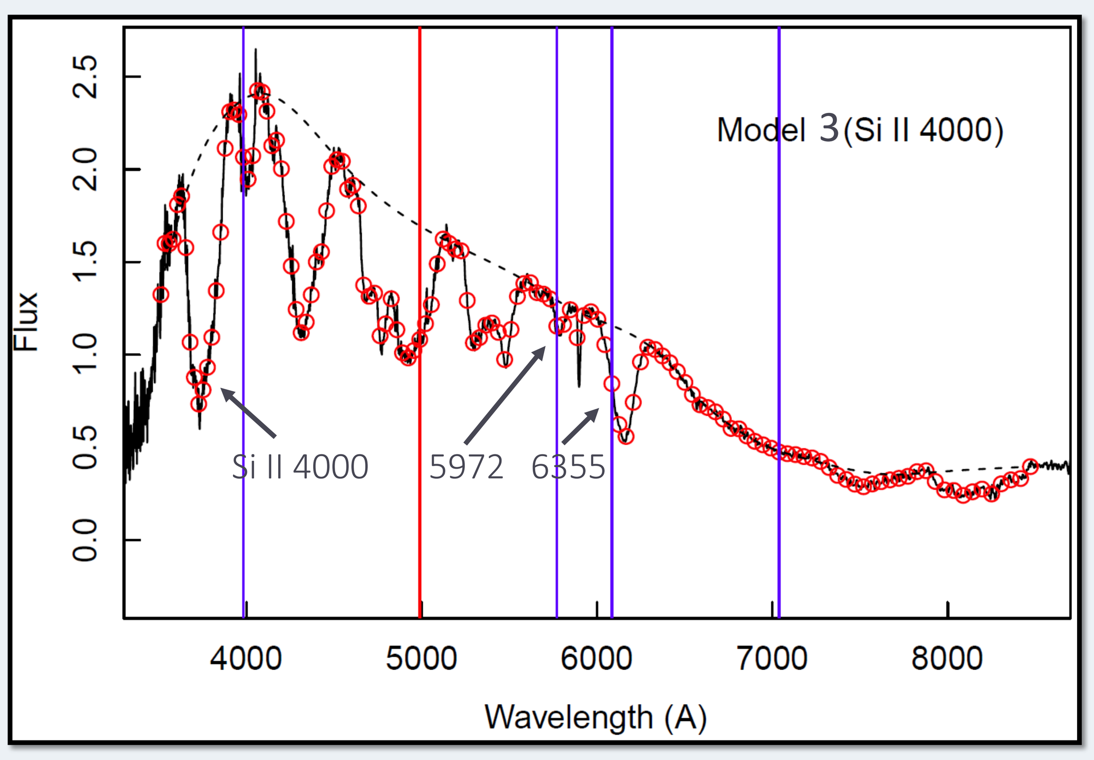
Correlation between R(3780/4580) and the color "c".

Model 3

(Model of x1)

Target variable: x1
Explanatory variables: c, f_{tot}, f_{cnt}, R
Selected variables (see below) → Si II lines

→ The strength of the Si II lines is a good explanatory variables of "x1".
→ Using color & light-curve width correction to M_B
→ Model 4



Selected variables in Model 3 are indicated with the vertical lines with the typical spectra (SN 2006et). Red: f_{tot}. Blue: f_{cnt}.

Model 4

(Color & light-curve width corrections for MB)

Target variable: ($M_B - \beta_1 c - \beta_2$ x1)
Explanatory variables: f_{tot}, f_{cnt}, R

Selected variables:

- ✓ No variables were selected.
- ✓ The variables about the Si II lines do not significantly revise the model. Their selection in Model 1 was due to their high correlations with "x1".

→ The best set of variables is (c, x1) for the sample we used.
→ Our analysis does not support to add any other variables.

Discussion & Future plan

- ✓ Larger and better-calibrated data may allow us to find additional variables. The proposing method can treat one-order larger size of the data set in a laptop.
- ✓ To include the variables about the host galaxies would be interesting.
- ✓ LASSO did not use the measurement errors nor correlations of the explanatory variables. It might cause false negatives.

Reference:
[1] Phillips 1993, ApJL, 413, L105, [2] Bailey, et al., 2009, A&A, 500, L17, [3] Blondin, et al., 2011, A&A, 526, A81, [4] Silverman, et al., 2012, MNRAS, 425, 1889, [5] Tibshirani 1996, Journal of Royal Statistical Society. Series B (Methodological), 58, 267, [6] http://hercules.berkeley.edu/database/index_public.html, [7] Hachinger, et al. 2006, MNRAS, 370, 299