CS-BIGS 1(2): 164-172 http://www.bentley.edu/csbigs/vol1-2/germain.pdf

Pampering the Client: Calibrating Vehicle Parts to Satisfy Customers

Jean-François Germain
Ecole Nationale Supérieure des Télécommunications, France

We present in this paper a statistical methodology to address the following industrial problem. Car manufacturers have to calibrate their vehicles in order to reach a level of quality which is acceptable to the customer. We consider here the specific case of a gear-box. Our study relies on a dataset consisting of evaluations by 507 testers of 28 configurations, each described by 12 physical parameters. We suggest a procedure for selecting and calibrating the physical parameters which have an impact on the evaluations. Our procedure consists of two steps. We first compute the regularization path of an L_1 – penalized logistic likelihood from which we extract an increasing sequence of models. In the second step of our procedure, we apply the BIC criterion to select a model in the sequence obtained in the first step. We provide a simple numerical procedure for this approach and discuss its application to the data. This article is accessible to readers with at least an intermediate knowledge of statistics; previous exposure to logistic regression and the principles of model selection would be useful, although not strictly necessary.

Description of our industrial case

One of the most important activities for a car manufacturer is to calibrate vehicle parts. The calibration of some components is driven by levels of customer satisfaction regarding issues such as drivability, habitability, acoustics, ergonomics, etc.

However, this implies that engineers know how to calibrate physical parameters in order to reach a given quality level which is satisfactory enough for the customer. The question is of how to link those qualitative customer evaluations and the quantitative physical design characteristics of the vehicle.

To address this question, we will present a new complete methodology, which handles subjective evaluation and design parameters together. Our modeling of customers' appreciation yields:

- the selection of design parameters that are explanatory, which means that they have an impact on customers' subjective evaluation.
- Some help with the calibration of the selected design parameters.

The industrial application field is the evaluation of gear-boxes. The dataset we tested the methodology on is described below.

Data

The dataset consists of two types of data:

Subjective data are subjective evaluations. Testers, selected for their high sensitivity to issues related to shifting gears, evaluated several different gear-boxes. The goal of the study is to determine what is or is not acceptable for the customer. For the sake of confidentiality we simplify this evaluation procedure as follows: each evaluation is one or zero.

Objective data are physical design parameters. Twenty-eight different gear-boxes are chosen: they are representative enough to allow for a robust analysis of subjective evaluations. All testers evaluate each gear-box several times, so that we globally handle a dataset of 507 observations.

Gear-boxes for which testers proceed with subjective evaluations are all measured in the same way. As they attempt to link subjective evaluations and physical parameters together, engineering experts know or have an idea of which design parameters are of interest. Those experts selected and extracted twelve potentially explanatory parameters from each measured signal.

The dataset used in the analysis is available in the accompanying MS Excel file. A row is one of the 507 samples. A column is one of the twelve design parameters. The data have been centered and scaled.

Methodology based on logistic regression

Let us now introduce a statistical model. Let $\mathbf{Y}$ be the vector of subjective evaluations. As seen above, each response y_i is one or zero. We have:

$$\mathbf{Y} = (y_i)_{i=1\ldots n} \text{ with } y_i \in \{0,1\}, \forall i \in [1,n].$$

Our goal is the modeling of $\mathbf{Y}$ on the basis of the p measured explanatory variables X_1, X_2, …, X_p where $\mathbf{X_j} = [x_{1j}, x_{2j}, ..., x_{nj}]^T$. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_p]$ be the design matrix. The response has only two different ratings: we are in the field of binary classification.

As mentioned above, engineers first want to know which variables explain the separation between the two classes most efficiently. A model must then be provided to engineers, so that they can set target values on the selected explanatory variables in order to reach a given quality level. The model must be simple enough. Therefore we consider generalized linear models, excluding the use of kernels.

Our first task is to select a set of explanatory variables. We will keep in mind the importance of the variable selection aspects in the methodology we want to develop.

Popular approaches to binary classifications are for instance classification and regression trees (CART), discriminant analysis (DA) or support vector machines (SVM). CART, introduced by Breiman et al. (1984), can help with variable selection. However, in practice, this procedure can be sensitive to noise on data. Model selection in the context of a discriminant analysis is not easy because of the large number of models to consider. Finally, support vector machines are not tailored to model selection.

For purposes of model selection, it is crucial to rely on a likelihood criterion. We will base our method on the logistic regression likelihood, which, in the context of binary classification, has the following expression:

$$L_n(\boldsymbol{\beta}) = \exp\left\{\sum_{i=1}^{n} y_i x_i \boldsymbol{\beta} - \log\left(1 + e^{x_i \boldsymbol{\beta}}\right)\right\} \qquad (1)$$

where $\boldsymbol{\beta}$ is the regression parameter.

Model selection

A two step approach

Since in our context a model can be considered as a set of variables, the number of models grows exponentially with respect to the number of explanatory variables. To deal with model selection, one often relies on penalization functions. First contributors in this direction were Akaike (1973) who introduced a penalized log-likelihood for density estimation and Mallow (1973) who proposed a penalized least square regression. We can also quote the work of Birgé and Massart (2004) where powerful theoretical results are derived in a general Gaussian framework. In this work the collection of models is also quite general. An important conclusion of these theoretical results is that the larger the collection of models, the stronger the penalization term. Moreover, the stronger the penalization term, the larger the prediction error of the estimated model.

It is more favorable to perform model selection with a smaller number of models. Therefore we suggest a two steps approach. We propose to:

a) organize explanatory variables into a hierarchy
b) apply a penalized likelihood method to the sequence of nested models.

The sequence of nested models comes from step a). The first model contains only the first explanatory variable in the hierarchy and each following model consists of the current model with the addition of the next variable in the hierarchy.

Regularization path

We want to organize explanatory variables into a hierarchy, in order to know which model with only one variable is best, which model with two variables is best, etc. Maintaining a control on the model size is a goal we keep in mind in addition to the other goal of the performance of the binary classification.

There is indeed a trade-off between the model size and the error rate. The idea is to obtain a model which describes the data well enough while having a suitably low number of explanatory variables. Those two goals are indeed opposite: the best predictor with all the variables leads to the lowest error rate, but this best predictor is specific to the data and does not adapt itself well to new observations.

The model size is defined as the number of variables appearing in the linear combination in (1), and can be interpreted as the L_0 norm of the regression parameter. Direct handling of the L_0 norm penalization causes heavy algorithmic difficulties and computation delays. We prefer here to use the L_1 norm, which equals the sum of the absolute values of the regression coefficients, and is a good compromise between the L_0 norm and the L_2 norm – the Euclidian norm. We prefer the L_1 norm to the L_2 norm because the L_1 norm is closer to the L_0 norm than the L_2 norm which defines the Ridge regression. The reason for choosing the L_1 norm rather than the L_0 norm is the convexity of the L_1 – penalized problem, contrary to the L_0 – penalized one which is not convex. In this approach, the "model size" is thus evaluated by the L_1 norm of the regression parameter.

We focus on the trade-off between effective classification and controlled model size. We can describe this trade-off as follows:

- The error rate is controlled by the likelihood $L_n(\boldsymbol{\beta})$
- The model size is controlled by $\|\boldsymbol{\beta}\|_1$

Combining these two terms, we define the expression:

$$\boldsymbol{\beta}(\lambda) = \underset{\boldsymbol{\beta}\in\mathbf{R}^p}{\arg\min}\left\{-\log L_n(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1\right\} \qquad (2)$$

where λ is a regularization parameter. This parameter sets the relative importance of each of the two antagonistic goals.

This penalized approach – in the case of standard linear regression – was introduced as LASSO by Tibshirani (1996) and well-studied in the literature since. We can quote the very interesting contribution by Efron et al. (2004). The authors present an algorithm called LAR (for Least Angle Regression) which computes a LASSO solution (for L_1-penalized least squares regression). He also obtains a Stagewise Regression solution by slightly modifying the LAR algorithm. Rosset et al. (2004) establish conditions on both cost and penalty functions in order to have a piecewise linear regularization path. Under those conditions, the entire path can easily be calculated from only a few points. Keerthi and Shevade (2006) suggest an approximation of the logistic regression loss function by a piecewise quadratic function.

Methods have also been proposed for choosing λ from data. For instance Zou et al. (2004) prove that the number of non-zero coefficients is an unbiased estimator of the number of degrees of freedom.

Asymptotic results have also been established. For a fixed p, Knight and Fu (2004) prove in a more general setting (namely least squares penalized by $\sum_{j=1}^{p}\left|\beta_j\right|^{\gamma}$) than the LASSO formulation that there exists –asymptotically with n – a mass of probability at 0 when the variable is not in the true model. Zhao and Yu (2006) establish that LASSO selects the true model consistently under a condition called the "Irrepresentable Condition". Those statements hold in the large p setting as n gets large.
Our approach is different since we focus on building a hierarchy of explanatory variables.

We are interested in the *order of appearance*, which is the order in which the explanatory variables enter the model as the model size increases. We assume that an explanatory variable is more important if it appears early in the model linear combination. However this requires a lot of care, as will be discussed later in the paper.

One way to obtain the explanatory variable hierarchy is to determine the *regularization path*, defined as the mapping $\lambda \mapsto \boldsymbol{\beta}(\lambda)$, where $\boldsymbol{\beta}(\lambda)$ is defined by (2). It can be computed by the LAR in the case of LASSO (Efron et al., 2004).

In the case of the penalized logistic regression procedure, an algorithm presented by Park and Hastie (2006)

computes an approximation to the path which is inspired by the LAR algorithm.
Since we focus only on the order of appearance, we do not need to compute the whole path. We propose later a different algorithm whose goal is to determine this order.

Model selection with BIC

A regularization path procedure organizes explanatory variables into a hierarchy, as shown in Figure 1. We denote by *active set* the current set of explanatory variables in the model. The sequence of active sets is displayed in Figure 1.

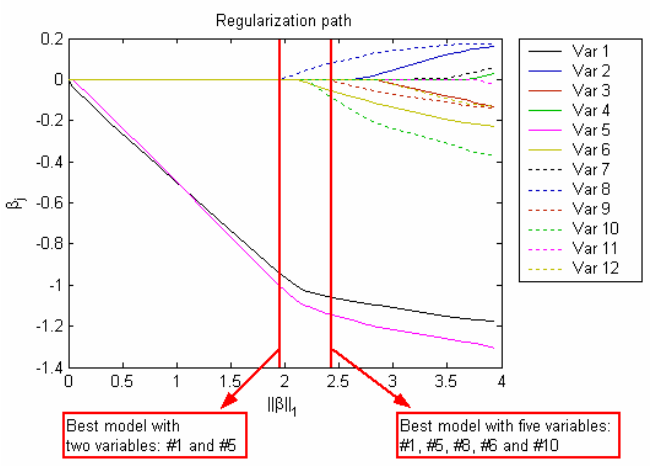

Figure 1. The regularization path identifies the best model with one variable, the best model with two variables, etc. Reading from left to right, explanatory variables are organized by importance into a hierarchy.

The regularization path results in an increasing sequence of models, as follows. Reading the figure vertically, at the very left we have $\|\boldsymbol{\beta}\|_1 = 0$, which corresponds to $\lambda \to \infty$. Then $\|\boldsymbol{\beta}\|_1$ increases as λ decreases until the very right of the figure, where $\lambda = 0$ corresponds to the maximal value of $\|\boldsymbol{\beta}\|_1$. The coefficient vector for $\lambda = 0$ is the result of the non-penalized logistic regression. Between $\lambda = 0$ and $\lambda \to \infty$, we have all the intermediate models.

In a second step, we consider the Bayesian Information Criterion of Schwarz (1978) for selecting the optimal model. If we denote by $\{M_k\}_{k=1...m}$ the sequence of models, this criterion equals:

$$\left[BIC(M_k), \boldsymbol{\beta}_{BIC}(M_k)\right] = \inf_{\boldsymbol{\beta} \in V_k} \{-2\log L_n(\boldsymbol{\beta})\} + \|M_k\|_0 \log(n) \quad (3)$$

where n is the sample size of sample, V_k is the subspace of $\mathbf{R}^p$ with zeros at coordinates that do not appear in the current model M_k, and $\|M_k\|_0$ is the number of variables in the model M_k.

In our procedures, we recommend selecting the model that yields the minimal value of the BIC sequence. This way we ensure good prediction performance in the sense that we avoid over-fitting. Figure 2 displays the BIC sequence corresponding to the models sequence obtained from Figure 1.

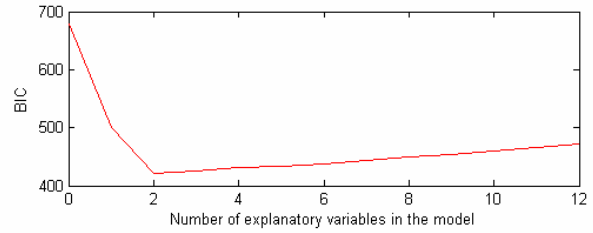

Figure 2. BIC curve.

Numerical aspects

In this section, we present an algorithm for obtaining the sequence of actives sets. An active set A_λ is defined as the set of explanatory variables corresponding to non-vanishing coordinates of β_j for a given regularization parameter $\lambda \geq 0$.

At a given λ, in the setting of L_1-penalized logistic regression, the calculation of a specific active set is a convex optimization problem.

By (1) and (2), for a given $\lambda \geq 0$, we need to compute:

$$\boldsymbol{\beta}(\lambda) = \arg\min_{\boldsymbol{\beta} \in \mathbf{R}^p} \left\{ \sum_{i=1}^{n} [-y_i x_i \boldsymbol{\beta} + \log\left(1 + e^{x_i \boldsymbol{\beta}}\right)] + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \quad (4)$$

This criterion is clearly convex, but the L_1 norm on coefficients is problematic for differentiation near the axes. That is why we rewrite this criterion as the following:

$$\boldsymbol{\beta}(\lambda) = \arg\min_{\boldsymbol{\beta} \in \mathbf{R}^p} \left\{ \sum_{i=1}^{n} [-y_i x_i \boldsymbol{\beta} + \log\left(1 + e^{x_i \boldsymbol{\beta}}\right)] + \lambda \sum_{j=1}^{p} u_j \right\}$$
$$s.t. \quad -u_j \leq \beta_j \leq u_j \quad (5)$$

This new criterion searches for the same solution as (4) and is C^∞, with linear constraints.
Such an optimization can be performed using standard numerical procedures. In our application, we have used the Matlab function *fmincon* to solve this optimization problem.

A very simple way (close to the one presented by Park and Hastie, 2006) to calculate the path consists in a stepwise optimization. Beginning with $\lambda = 0$, λ is increased by $\delta\lambda$ step by step. At each step $\boldsymbol{\beta}(\lambda)$ is calculated along with the optimization function. The algorithm ends when all coordinates of $\boldsymbol{\beta}(\lambda)$ are zero (for λ large enough, all coordinates are indeed zero).

The main problem is the choice of step length, which must be small enough to insure that all interesting points of the path are visited. We present an intuitive algorithm, which is easy to develop and reduces the number of optimizations to perform.

> Algorithm idea: instead of calculating $\boldsymbol{\beta}(\lambda)$ at each step, by increasing λ by steps of $\partial\lambda$, we suggest exploring the range of $\lambda \in [0,+\infty[$ in a dichotomist way.

Considering equation (4), we first note that there exists $\lambda_{\max}$ such that $\forall \lambda > \lambda_{\max} : \boldsymbol{\beta} \equiv 0$. The constraint on the size of the coefficient tends to infinity and thus forces every coefficient to be zero. We consider the value of $\lambda_{\max}$ given in Park and Hastie (2006), namely $\lambda_{\max} = \max_{j\in\{1,\dots,p\}} \left| X'_j (y - \bar{y}) \right|$. We just need an upper bound on λ.

We base our stopping condition on the current active set, that is, we decide to perform new levels of dichotomy if the difference between the two active sets is more than one in terms of cardinality. If we denote the active set for a given λ by A_λ, the stopping condition can be written as:
$COND(A_{\lambda_1}, A_{\lambda_2}) : \exists!\, j \in \{1,\dots,p\}, \exists B \subset A_{\lambda_2} : A_{\lambda_1} = \{j\} \cup B$.
Our algorithm is described in more detail, with comments in italics, in the Appendix.

Application to the industrial case

In this section, we apply our algorithm to the data we presented earlier. We recall that variables are centered and scaled.

Interpretation of the regularization path

Our algorithm performs the regularization path presented in Figure 1. For ease of interpretation, we present the evolution of the coefficients of explanatory variables as functions of $\|\boldsymbol{\beta}\|_1$.

We can summarize the interpretation of this result as follows. As we release the constraint, variables #1 and #5 enter the model nearly together. For an industrial application, we shall say that there is no statistical reason to consider one of those without the second one. This is the first piece of information. The BIC curve presented in Figure 2 contributes to the second main part of the interpretation, that is: both variables #1 and #5 are sufficient to explain the physical phenomenon we model.

Those results concur with the engineers' intuitive analysis of the physical phenomenon: variable #1 was already in the specifications and engineers had the feeling that variable #5 could bring some more information. Our statistical analysis proves them to be right.

Sensitivity analysis

In this step of the interpretation, we have found which variables have an impact on subjective evaluations by customers. Having estimated the model and the regression parameter β, we have in hand the probability $\mathrm{P}(Y = 1 | X = x)$ where *Y* is the gear-box rating and *x* is the value of the physical variables.

Our prediction is a probability between 0 and 1. But the variable *Y* we wish to predict is binary. So in order to decide the value of *Y*, knowing that $X = x$, we have to choose from which threshold probability we decide that *Y=1*.

In statistical words, we say that we build the following test:

$$\hat{Y}(x) = \begin{cases} 1 & \text{if } \mathrm{P}(Y = 1 | X = x) > c \\ 0 & \text{otherwise} \end{cases},$$

where *c* is called a *probability threshold*. This test decides whether the physical parameter *x* calibrates an acceptable gear-box.

Our job is also to determine the value of the probability threshold. Specifications will follow. We can for instance set the threshold to a high value to harden specifications: only very good gear-boxes will be validated.

The choice of the probability threshold is driven by the error rate. There are two types of error rate: the false

negative one and the false positive one. The false positive error rate is $P(\hat{Y}=1|Y=0)$ and the false negative error rate is $P(\hat{Y}=0|Y=1)$.

From an industrial point of view, a false positive observation corresponds for instance to a validated gear-box (a recast-as-"1" sample) which is criticized by the customer (the customer's subjective evaluation is "0"). For a car manufacturer, such a false positive observation causes the most prejudicial type of error.

Figure 3 summarizes these trade-offs graphically. The x-axis plots differences between a given value of the model linear combination and the value of the linear combination which corresponds to an estimated probability $P(\hat{Y}=1)$ of .8. These differences decrease from left to right. The smooth curve in Figure 3 represents the estimated probabilities $P(\hat{Y}=1)$ for each value of the differences (note that a zero difference does correspond to a .8 estimated probability). On the other hand, the jagged line represents the false positive rates for threshold probabilities corresponding to differences on the x-axis. We can see that the .8 threshold probability corresponds to a false positive rate of about 8%.

Note that the choice of a threshold probability of .5 would imply a false positive rate of near 20%, which is too high. Overall we feel that the choice of .8 for a probability threshold is a reasonable compromise.

For another description of the evolution of error rates, we display the ROC (Receiver Operating Characteristic) curve in Figure 4.

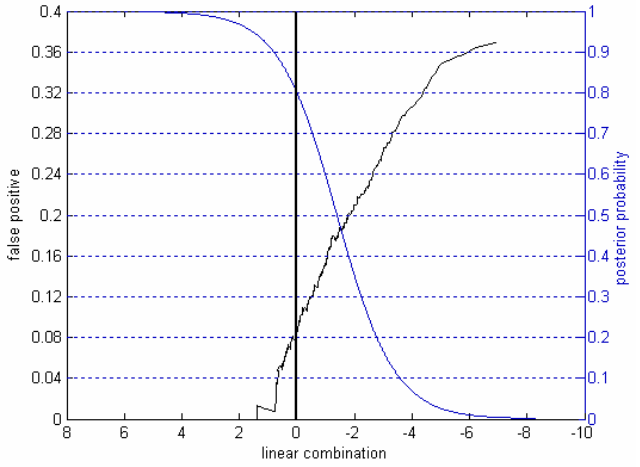

Figure 3. Posteriori probability and false positive error rate as functions of the value of the linear combination (distance from the separating line). Vertical line corresponds to the separating line in Figure 5.

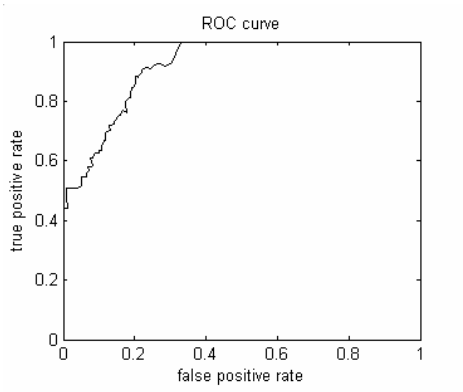

Figure 4. ROC curve.

We also plot the complete sample in the space {variable #1, variable #5} in order to visualize the separating line corresponding to the probability threshold of .8.

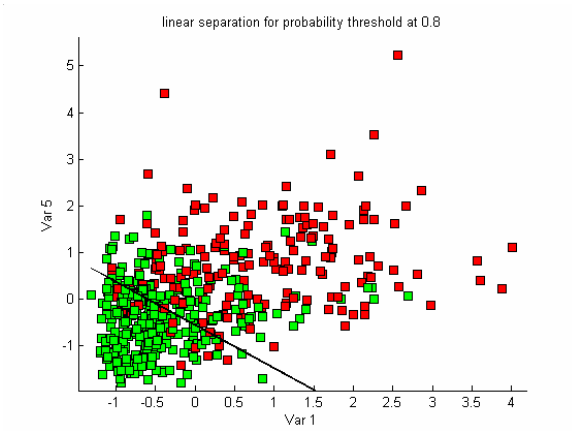

Figure 5. Position of the separating line corresponding to a probability threshold of 0.8. Observations with Y=1 are green. Observations with Y=0 are red. Observations below the line have $\hat{Y}=1$; observations above the line have $\hat{Y}=0$.

For a complete industrial analysis of those results and recommendations, we have to point out that those figures can also help with design. Indeed, the links between values of linear combinations, separating line positions and probability thresholds highlighted above can help to calibrate a new gear-box. Engineers have to decide on a specific target zone in the space {variable #1, variable #5}. Assuming that one of the variables has already been calibrated and cannot be modified anymore, the target zone could nevertheless be reached thanks to the other variable. Moreover, engineers would have information about the gain or loss in probability for each value of the variables.

Discussion

We assumed above that the importance of explanatory variables is related to the order in which the variables

enter the model as the controlled model size gets larger. We assumed that an explanatory variable is more important if it appears early in the model linear combination.

In the figure presented in this paper, the order of the two selected variables changes along the path. This is not a problem because they are selected together: we consider those two variables as a group. Variables maintain a constant order during the path if the following circumstances occur: a variable which enters the model maintains a magnitude

- larger than the following variables
- and at the same time smaller than that of the active variables (variables which are already in the model).

In practice, this is not the case. Links between explanatory variables such as nearly linear relations may have an impact on the stability of the order of variables. This issue becomes even more critical as soon as $p > n$.

We show in Figure 6 the following example. We consider the same dataset as previously, but to which we add only three more explanatory variables. They are measurements of physical parameters similar to the twelve first variables.

The BIC leads us to considering two variables in the model. But which two variables? In this setting, the two first variables to enter the model ('Var 1' and 'Var 5') are not the same two variables as the two maximum absolute magnitude variables at the end of the path. Our approach seems to fail because variable #5 is the second variable into our defined hierarchy but is the third variable in absolute magnitude at the end of the path.

Greenshtein and Ritov (2006) give bounds on the L_1-norm and on the number of variables as n and p simultaneously get large, with $p >> n$. These model selection procedures tend to select much larger variable sets. The first variables that enter the model are obviously of some importance in the explanation of the response but this procedure may have to be moderated: there might be a trade off to discover between those two approaches.

The path presented in Figure 6 is ambiguous. In our industrial application, we discussed this phenomenon with industrial experts. Considering the physical interpretation of each design parameter, variables #5, #13 and #14 appeared deeply linked. Another physical link appeared between variables #1 and #15.

Experts decided to consider only one variable, representative of its group. We kept variables #1 and #5. Experts made their choice according to the physical meaning of each group and have selected the variable which is easiest to interpret.

Once the variables #13, #14 and #15 are removed from data, the path appears to be clearer and the selection of variables set is not ambiguous any more: the two variables first entering the model remain the two variables with the largest absolute magnitude along the path.

This step turned out to be crucial for a good understanding of the problem. Discussions with experts enabled us to find a regularization path that can be easily interpreted. It would be interesting to have an automatic method to clarify the regularization path. This is an open issue.

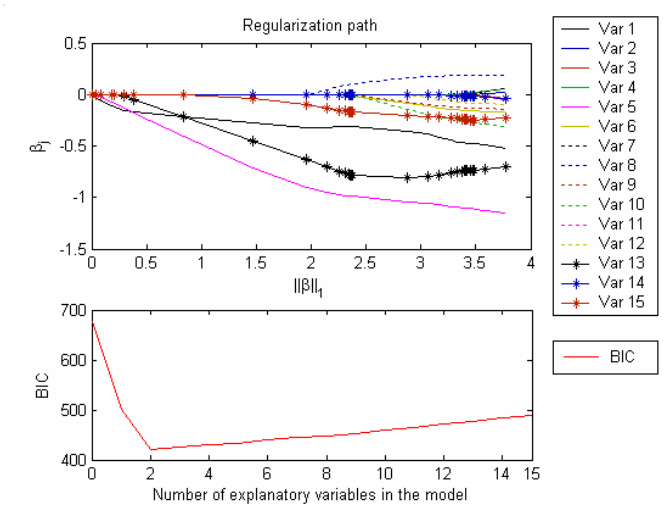

Figure 6. Regularization path and BIC curve on the augmented industrial dataset. We note that variable #1 and variable #5 are the two first design parameters to enter the model but they are not the two with the greatest magnitude at the end of the path (model resulting from the non-penalized logistic regression).

Conclusion

We have considered the problem of binary classification using a logistic regression model. Our main objective was to select a few explanatory variables. We suggested

- first to select a sequence of models by ordering the explanatory variables as an output of the regularization path of the L_1 – penalized likelihood.
- second to apply the BIC criterion to select a model in this sequence.

Our ordering method can be summarized as follows:
"The sooner a variable enters the regularization path as the penalty decreases, the more explanatory it is." This

interpretation of the regularization path is ad hoc and thus needs some care in practice. As we applied this approach to our industrial case, we relied on experts to first clean up the regularization path so that the selected variables also correspond to those having large regression coefficients order of magnitude in the regularization path.

Acknowledgments

This methodology was developed in the framework of a CIFRE-PhD in the Research Center RENAULT. I want to thank François Roueff[1] for his precious help with this work, and Marc Chauvet for his language corrections.

Correspondence: germain@enst.fr

[1] Télécom Paris, CNRS LTCI, URA820, 75634 Paris Cedex 13, France.

REFERENCES

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd International Symposium on Information Theory*, P. N. Petrov and F. Csaki, editors, 267-281.

Birgé, L., and Massart, P. 2001. Gaussian model selection. *Journal of the European Mathematical Society* 3(3):203-268.

Breiman L., Friedman J. H., Olshen R. A., and Stone C. J. 1984. *Classification and regression trees*. Wadsworth International, Belmont, California.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. 2004. Least angle regression. *Annals of Statistics* 32:407-499.

Greenshtein, E., and Ritov, Y. 2004. Persistency in high dimensional linear predictor-selection and the virtue of over-parametrization. *Bernoulli* 10:971-988.

Keerthi, S., and Shevade, S. 2006. A fast tracking algorithm for generalized LARS/LASSO. Submitted to *IEEE transactions on Neural Networks*.

Knight, K., and Fu, W. 2000. Asymptotics for LASSO-type estimators. *Annals of Statistics* 28(5):1356-1378.

Mallows, C. L. 1973. Some comments on Cp. *Technometrics* 15:661-675.

Park, M. Y., and Hastie, T. 2006. L_1 regularization path algorithm for generalized linear models. *Technical report*. Stanford University, Stanford.

Rosset, S., and Zhu, J. 2004. Piecewise linear regularized solution paths. *Annals of Statistics*, to appear.

Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics* 6(2):461-464.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society B* 58:229-243.

Zhao, P., and Yu, B. 2006. On Model Selection Consistency of Lasso. *Technical Report 702*, Statistics Department, UC Berkeley (to appear in *Journal of Machine Learning Research*).

Zou, H., Hastie, T., and Tibshirani, R. 2004. On the degrees of freedom of the lasso. *Technical report*, Department of Statistics, Stanford University.

Appendix: Description of Algorithm:

1) Initialization: calculation of $\lambda_{\max}$.

   $\lambda_{\max} = \max_{j\in\{1,\dots,p\}} \left| X'_j (y - \bar{y}) \right|$

   set $(\lambda_1, \lambda_2) \leftarrow (0, \lambda_{\max})$

   calculate A_{λ_1} and A_{λ_2}

2) While $not\{COND(A_{\lambda_1}, A_{\lambda_2})\}$ We are going to cut the λ range until the two active sets A_{λ_1} and A_{λ_2} are equal or differ by exactly one variable.

   a) $\mu = \frac{\lambda_1 + \lambda_2}{2}$. We split the λ range in its middle.

   b) calculate A_μ

   c) while $not\{COND(A_\mu, A_{\lambda_2})\}$, do

      i) if $A_\mu = A_{\lambda_2}$, do

         (1) $(\lambda_1, \mu, \lambda_2) \leftarrow \left(\lambda_1, \frac{\lambda_1 + \lambda_2}{2}, \mu\right)$ We dichotomize this other side of the λ range.

         (2) calculate the new A_μ

         (3) calculate the new A_{λ_2} In fact, this calculation is not performed because we already have calculated the new A_{λ_2} : it is A_μ .

      ii) end if

      iii) if $\left| A_\mu - A_{\lambda_2} \right| \geq 2$, do

         (1) $(\lambda_1, \mu, \lambda_2) \leftarrow \left(\mu, \frac{\lambda_1 + \lambda_2}{2}, \lambda_2\right)$ We have not split enough. We keep dichotomizing this side.

         (2) calculate the new A_μ

         (3) calculate the new A_{λ_2} Same remark as in 2) c) i) (3): the new A_{λ_2} is already calculated: it is A_{λ_2} .

      iv) end if

   d) end while

   e) $\lambda_1 \leftarrow \max\{\lambda \in \Lambda : \lambda < \mu \text{ and } not\{COND(A_\lambda, A_\mu)\}\}$

   Let Λ be the set of λ already calculated. During the algorithm, calculations of active sets give much information. We can improve the algorithm by looking for the more judicious λ_1 to consider.

   f) $\lambda_2 \leftarrow \min\{\lambda \in \Lambda : A_\lambda = A_\mu\}$ The more judicious λ_2 to consider is the lowest value of λ for which $A_\lambda = A_\mu$.

3) end while

4) end

The algorithm summarized:

1) $\lambda_{\max} = \max_{j\in\{1,\dots,p\}} \left| X'_j (y - \bar{y}) \right|$

set $(\lambda_1, \lambda_2) \leftarrow (0, \lambda_{\max})$

calculate A_{λ_1} and A_{λ_2}

2) while $not\{COND(A_{\lambda_1}, A_{\lambda_2})\}$ do

   a) $\mu = \frac{\lambda_1 + \lambda_2}{2}$

   b) calculate A_μ

   c) while $not\{COND(A_\mu, A_{\lambda_2})\}$, do

      i) if $A_\mu = A_{\lambda_2}$, do

         (1) $(\lambda_1, \mu, \lambda_2) \leftarrow \left(\lambda_1, \frac{\lambda_1 + \lambda_2}{2}, \mu\right)$

         (2) calculate A_μ

      ii) if $\left| A_\mu - A_{\lambda_2} \right| \geq 2$, do

         (1) $(\lambda_1, \mu, \lambda_2) \leftarrow \left(\mu, \frac{\lambda_1 + \lambda_2}{2}, \lambda_2\right)$

         (2) calculate A_μ

   d) $\lambda_1 \leftarrow \max\{\lambda \in \Lambda : \lambda < \mu \text{ and } not\{COND(A_\lambda, A_\mu)\}\}$

   e) $\lambda_2 \leftarrow \min\{\lambda \in \Lambda : A_\lambda = A_\mu\}$