NTNU
Norwegian University of
Science and Technology

Variable selection in Cox-models using the L1-regularization path algorithm

Mahder Assefa Tuji

Master of Science in Mathematics (for international students)
Submission date: August 2010
Supervisor: Bo Henry Lindqvist, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Variable Selection In Cox-models Using the L_1-regularization Path Algorithm

Mahder Assefa Tuji
Department of Mathematics,
Norwegian Science and Technology University,
Trondheim, Norway

Abstract

In gene expression data usually the dimension of covariates are much bigger than the total number of observation, in such a case the Least Square Estimator are not possible to compute so taking dimension reduction either by variable selection or shrinkage method, which make it possible to estimate the most significant variable connected to the response(time to event). Therefore, In the paper have seen three examples; which shows using Glmpath package with Cox-models to select variables. Among these one example lung cancer patients when $p < n$ while the other two examples have looks anonymsed data but unclear when $p > n$. Afterall the treatment is highly significant variables in the response on the time of event. So I have done special analysis by using the only covariates term treatment. However choosing AIC as one of the model selection criteria to select the most significant variables on the survival data set. Finally selecting the variable which included for those 121 covariates among 17910 in the coxpath model.

Contents

Problem Description

The purpose of the thesis is:

(i) to give a theoretical overview of survival modeling and inference in Cox-models with high-dimensional covariates

(ii) to use the R-package 'glmpath' to analyse real data sets

Assignment given: September 1, 2009 Supervisor: Bo Henry Lindqvist

1 Introduction

In regression analysis having the number of covariates much bigger than the sample size will not be in proportion to the assumptions of the regression model. So glmpath algorithm is a very good algorithm to identify variables which is the most significant on survival time for such kinds of a problem. The objective of this thesis is to select a model from high dimensional data in the regression setting when the number of covariates are large and the sample size is small. And the algorithm which could be a shrinkage method with L_1- penalties of the lasso type. The L_1 regularization procedure is useful for selects variables according to the amount of penalization on the lasso coefficients. Therefore to select covariates according to the algorithm, AIC of the breast cancer data and BIC of the lung cancer data are calculated or used for model selection method. The picked out to be used in the analysis, it is being the fact that the covariates showing the minimum AIC/BIC were kept outside in the model. The thesis is organized as follows, in the first section briefly discussed about the theory of survival analysis connected with such kinds of data type. In this section define Censored observation, Survival function and Hazard function in general. And look also the general relation ship between hazard and survival time can be used to develop the distribution will indicates the good assumptions of the model of the data. In section three one of the most important statistical models in survival analysis is the Cox-proportional hazards model is included, in addition to this the cox model with the penalty lasso is focused on problems of high dimensionality and then implementing a design for handling such problems is presented in this section of sub section. Further I am interested in determining which model describes the data best, therefore the model with the minimum Akaike-information criterion(AIC) is chosen as the best model. By minimizing or limiting the lasso type of l_1- norm of an unknown variable and using the algorithm glmpath could be maximized the estimated coefficient. Since the non-zero components in the solution correspond to useful features for the L_1 regularization. The discussion part in the thesis devoted to interprate the results and graphs in general. This is organized in three sub section and the first sub section give describtion of lung cancer data, breast cancer data on time to event A are presented in the second sub section and under the last sub section time to event B is presented. The R-package code and its output are presented in the appendix.

2 Survival Analysis

In this chapter I have done the theory of Survival Analysis along with notation and description of it. The theory is gained from different author and books and their names have been put in the reference section. A problem usualy solved by applied statistician is the analysis of survival data. Such data are obtained in different fields like in biology, medicine, engineering and so on. Because of Censoring information the analysis of survival experiments is difficult. The characteristics of such kinds of data sets are they contain censored observation or not. It is important notice that censoring times can appear both as random variables and fixed quantities. Censored data arises when an individuals life length is known to occur only in a certain period of time. The most known censoring schemes are right censoring, where all that is known is that the individual is still alive at a given time, left censoring when all that is known is that the individual has experienced the event of interest prior to the start of the study, or interval censoring, where the only information is that the event occurs within some interval. Denote T as the time from an initiating event and to an event of interest. Therefore T is known as a survival time and is a non-negative random variable. In this part survival time is used for the time until the patients experience a recurrence of cancer related death.

2.1 Survival function

Let T represent survival time, regard T as a random variable with cumulative distribution function

$$F(t) = \Pr(T \leq t)$$

The probability of an individual surviving beyond time t is defined mathematically

$$S(t) = \Pr(T > t)$$

In this paper T consider to be the time until some specified event. This event may be death, the development of some disease, recurrence of disease and so on. T is survival function which is probability of an individual surviving to time t. If T is a continuous random variable, then S(t) is a continuous, strictly decreasing function. When T is a continuous random variable, the survival function is the complement of the cumulative distribution function, that is

$$S(t) = 1 - F(t)$$

Therefore the relationship between S(t) and F(t) makes it possible to obtain the density function as $f(t) = -S'(t)$

2.2 The Hazard function

The hazard function is the conditonal failure rate or the age specific failure rate. It is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

As described in the above if T is continuous random variable, then

$$h(t) = \frac{f(t)}{S(t)}$$

$h(t)\Delta t$ may be viewed as the approximate probability of an individual of age t experiencing the event in the next instant. It is determining the appropriate failure distributions utilizing qualitative information about the mechanism of failure and for desribing the way in which the chance of experiencing the event changes with time. There are many general shapes of hazard rate, but it is restricted that should be non-negative, i.e.,

$$h(t) \geq 0.$$

Lucia Ohno-Machado[5] has been shown modelling of survival data usually employes the hazard function or the log hazard. Moreover, conditional on the value of any covariates in a survival model and on an individual's survival to a particular time, censoring must be independent of the future value of the hazard for the individual. If this condition is not met, then estimates of the survival distribution can be seriously biased. For example if individuals tend to drop out of a clinical trial shortly before they die, and therefore their deaths go unobserved, survival time will be over-estimated. Censoring that meets this requirement is non-informative. A common instance of non-informative censoring occurs when a study terminates at a predetermined date.

3 Cox Proportional Hazard Model

The proportional hazards model relates the hazard function at time t, $h_i(t)$, the instantaneous risk of an event given that the event has not yet occurred, to the risk covariates, X1,X2,....,Xp. Survival analysis typically examines the relationship of the the survival distribution to covariates. Most commonly, this examination entails the specification of a linear-like model for the log hazard. For example, a parametric model based on the exponential distribution may be written as

$$\log h_i(t) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$$
$$h_i(t) = \exp \alpha + \beta_i x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$$

that is, as a linear model for the log-hazard or as a multiplicative model for the hazard. Here, i represents a subscirpt for observation, and the x's are the covariates. The constant α in this model denotes a kind of log-baseline hazard

$$\log h_i(t) = \alpha$$
$$h_i(t) = e^{\alpha}$$

In this mode β's are the proportional hazards regression coefficient which are to be estimated. The quantity α is the underlying hazard rate at time t when all the covariates are zero. The model implies that the ratio of the hazards for two individuals depends on the difference between their linear predictors at any time. Without time-varying covariates this ratio is a constant independent of time. This means that no assumption is made about the distribution of the α with a function of time. But the hazards for the different covariate sets are assumed to be proportional to that of the underlying hazard function α with a function of time. The estimates of Cox's proportional hazards regression coefficient β do not depend on the exact time at which the outcome event occurs, but on the rank ordering of the event times. The Cox-model, in contrast, leaves the baseline hazard function

$$\alpha(t) = \log h_0(t)$$
$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$$

or, again equivalently

$$h_i(t) = h_o(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip})$$

$h_i = h_0(t) \exp x_i \beta$, where $h_i(t)$ is the baseline hazard at t, $\mathbf{x}$ is the vector of explanatory variables and β is the vector of coefficients for each variable.

This model is Semi-parametric because the baseline hazard can take any form which means either parametric or non-parametric, the covariates enter the model linearly. John Fox(2002)[4] presents by comparing two observations i and i' that differ in their x-values, with the corresponding linear predictors

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$$
$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'2} + ... + \beta_p x_{i'p}$$

The hazard ratio for these two observations,

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{h_o(t)e^{\eta_i}}{h_o(t)e^{\eta_{i'}}} = \frac{e^{\eta_i}}{e^{\eta_{i'}}}$$

is independent of the time t. In the Cox-model the baseline is unknown since its model is flexible to introduce time dependent covariates and handle censoring of survival times due to this use the method of partial likelihood can be estimate the Cox model. Like regression method, Cox proportional models are used for a model with distribution are a member of exponential family, however the response variable consists of both Survival time and Censoring information with their covariates. Generally, this model frequently used to study the importance of covariates for survival, but is rarely used to produce survival prognoses. The Cox model is a multivariate regression semi-parametric model that allows modelling of continuous covariates, and involves the assumption that the hazards for the different groups are proportional. The model assumes a baseline hazard and hazards for individuals with certain variable values are multiples of that baseline. In order to provide predictions of survival for individual patients, a baseline hazard that is common to all patients has to be estimated .This estimation represents no trivial task and the choice of the wrong baseline hazard can change the results of predictions in a very impressive manner.

3.1 Cox model with lasso penality

Consider the relationship between the Survival time and the Covariates such as gene expression levels and clinical trials in my case study treatment. The survival time for patients i=1,...,N, observe(t_i,δ,x_{i1},...,x_{ip}), where δ_i is the censoring indicator taking 1 being complete and 0 otherwise, t_i denotes the survival time if $\delta_i = 1$ or censoring time otherwise, and $\mathbf{x_i}$ =(x_{i1},...,x_{ip})^t is the vector of covariates. The hazard function for patient i, the proportional-hazard model for suvival data is given as

$$\lambda_i = \lambda_0(t) \exp x_i^T \beta,$$

where $\lambda_0(t)$ is a arbitrary baseline hazard function and $\beta = (\beta_1,...\beta_p)^T$ is the estimate of the parameter vector. The partial log-likelihood is expressed as

$$l(\beta) = \sum_{i=1}^{N}\{\delta_i x_i^T \beta - \log(\sum_{j \in R_i} \exp x_j^T \beta)\}$$

Denote the log-partial likelihood by $l(\beta) = logL(\beta)$, and assume that x_{ij} are standardized so that the mean and variance is 0 and 1 respectively, where R_i is the set of indices of the patients at risk at time t_i. Usually for microarray data have $p > N$, there exists a serious collinearity problem when applying the partial likelihood estimation to the Cox model directly, Tibshirani(1997)[10] has shown to estimate the parameters in the above model under L_1 constraint:

$$\hat{\beta} = argmaxl(\beta)$$

subject to $||\beta||_1 \leq S$, where $||.||_1$ denote the L_1 norm and $s > 0$. The optimization problem is good for dimension reduction of covariates, but is computationally difficult because the L_1 objective function is not differentiable. The equivalent lagrange multiplier version of this $\hat{\beta} = argmin(-l(\beta) + \lambda||\beta||_1)$ for $\lambda > 0$. To select important variables under the proportional hazard model; Tibshirani(1997)[10] proposed to minimize the penalized log partial likelihood function i.e $\hat{\beta} = argmin\{-l(\beta)+\lambda||\beta||_1\}$ in the same sprite Fan(2002)[2] proposed penalize likelihood approach but it is based on nonconcave. The lasso penality is $||\beta||_1$, which shrinks small coefficients to zero and therefore gives the sparse solution.

4 Model selection criteria

Selection of variables is the most way of performing regularization. As discussed in Zou[1] for any regularization method the main point is to find a good choice of the regularization parameter such that the corresponding model is optimal. There are many types of model selection critera, for example using Cross validation it possible select the regularization parameter λ. And others such as AIC(Akaike Information Criteria) and BIC(Bayesian Information Criteria) posses different asymptotic optimality. It is well known that if the true regression function is not the candidate models, the model selected by AIC asymptotically achieves the smallest average squared error among the candidates, and the AIC estimator of the regression function is in the candidate models or not, Yang(2003)[11]. On the other hand, BIC is well known for its consistency in selecting the true model see Shao 1997[7]. If the true model is in the candidate list, the probability of selecting the true model by BIC approaches one as the sample size $n- > \infty$. Sugiura's(1978)[8] while for problems with small sample sizes AIC might be preferable to other model criterion when high dimensional data exist.

5 Glmpath package

Glmpath is a regression algorithm for high dimensional data, developed by Park and T. Hastie [6] have been introduced a path following algorithm for L1 regularization generalized linear model. Then the Glmpath algorithm provides a means of selecting variables according to the amount of penalization on the L1-norm of the coefficients, in a manner that is less greedy than forward selection-backward deletion. The path computes the exact solution coefficients at particular values λ to other values of λ. This strategy yields a more accurate path in an efficient way than alternative methods and provides the exact order of the active set changes, which is important information in many applications, such as gene selection. This algorithm facilitates model selection by using the prdictor-corrector method and finding the entire regularization path. And also selecting the step length of the regularization parameter is critical in controlling the overall accuracy of the paths.

6 Lasso-path algorithm

When $p > n$, any Least Squares solution has zero residuals, with infinitely many solution coefficients β. The lasso path leads to the unique zero-residual solution having minimum L_1 norm. Because of typically large number of covariates, models like Lasso-path algorithm can be used to approximate the regularization path for emperical error and penalty function. By using this model I can avoided the insignificant covariates before model fitting, both a parameter estimation and selection of covariates is performed together. This is working by putting a penality term on the model. Usually if large number of covariate exist or one of the set covariates will expressed as a linear combination of other covariates in other words collinearity occur, therefore such kind of algorithm is provided. The way of putting a penality on model parameter estimation is choose the most of the estimated parameters will be equal to zero. It is only some steps are required for the full set of solutions. One of the model fitting in statistical model is regularization, specially for the number of covariates are much bigger than observation, used for minimized both emperical error and penality. Lassos's L_1 penalty leads to sparse solutions, that is, there are few non-zero estimates among all possible choices. The parameter $\lambda \geq 0$ controls the amount of regularization applied to the estimate. According to Park and Hastie[6] when $\lambda = 0$ the Lasso as similar property as Ordinary Least Squares which minimizes the unregularized emperical loss. On the other hand, a very large λ will comletely shrinks the coefficient to 0 thus leading to the empty or null model. As discussed on their papers the values of λ will cause shrinkage of the solutions towards 0, and some coefficient may end up being exactly 0. But should be need to reduce the amount of shrinkage on the β estimates that are away from zero and regularize in a more similar fashion as L_o penalty. The vector of covariates are strongly correlated and their coefficient estimates are unstable, so Zou and Hastie(2005)[1] putting a -$\lambda \sum {\beta_j}^s$ to the log-likelihood to be maximize. $\beta(\lambda) = argmin\{-logL(Y;\beta) + \lambda||\beta||1\}$ where $\lambda > 0$ is control the amount of regularization parameter. As they discussed briefly inroducing an algorithm that works the predictor-correcor method to determine the entire path of the coefficient estimates as λ varies which means compute $\hat{\beta} : 0 < \lambda < \infty$. Start with λ is maximum value, the algorithm computes at each step solution set, estimating the coefficient with smaller value of regularization parameter λ based on the previous estimate. At each step optimization consists of three steps, these are determining λ the step size in λ predicting the corresponding change in the coefficients, and correcting the error in the previous prediction. The algorithm computes the coeficient paths in two steps by changing λ and updating the coefficient estimates through a Newton Iteration method. The goal of Lasso path algorithm is a fixed λ is reduced to finding the vectors of the coeffictients with maximized log-likelihoood loss function, it will approximate the L_1 regularized glmpath. Assuming that none of the vectors of the coefficiants is zero and apply differentiating the log-lokelihood with respect to the coefficient that is β. Since a unique continuous and differentiable function i.e $\beta(\lambda)$, such that the differentiating equals to zero exist within each open range of λ that give a certain active set of variables. Like other methods predictor-corrector method finds a serious of solutions by using solutions at one extreme value of the parameter and continuing to find the adjacent solutions based on the current solutions. Park[6] as proved that, when λ exceeds a certain threshold, the intercept is the only non-zero coefficient since in the beginning λ is maximum so the only parameter in the model is the intercept. And while as λ decreased more and more, other variables join the current set. And discussed by Friedman [3] and his colleguals it is similar to Lasso, ridge regression is known to shrink the coefficients of correlated predictors towards each other, allowing them to borrow strength from each other. But the drawbacks of this method in only shrinks the coefficient, does not have selection method but lasso does have both.

7 Discussion and Conclusion

In this section I showed L_1 regularization path algorithm for the cox-model using the lung cancer and breast cancer data set.

7.1 Analysis on Lung Cancer Data

The L1 regularization path algorithm for the cox model using the lung cancer data take out from the Life Time Course Datafile also included in the paper. The survival data looks on the paper, 137 advanced lung cancer patients come from the Veterans Administration Lung Cancer Study Group listed in R.L.Prentice: Exponential survivals with censoring and explanatory variables, Biometrika, 1973. Patients were randomized according to one of two chemotherapeutic agents (Treatment: 1=Standard, 2=test). Of particular interest was the possible differential effects of therapy on tumor Cell type. Tumors are classified into one of four broad groups (Celltype: 1=squamous, 2=smallcell, 3=adeno, 4=large). Further covariates recorded when the patients were taken on study were Performance Status(a measure of general medical status where 10-30 is completely hospitalized; 40-60 is partial confinement to hospital; 70-90 is able to care for oneself), Months from diagnosis to stating on study, Age in years, and Prior therapy(0=no,10=yes). Clearly, the data set consists of 137 samples with seven covariates and Survival time with censored information. In this data the covariate Cell considerd as a factor that have four levels. I used three of the levels of Cell and one counted as reference. From the Cox model the data are fitted very well, and also from the R-output almost all the covariates are significant, but this indicates one of the drawbacks of survival fit on cox-model i.e over estimation occur, so it must be modified by using the Coxpath. Here gives some information about the whole

```
Call:
coxph(formula = Surv(ytime3, ystatus3) ~ Treat + C1 + C2 + C3 +
    PS + Month + Prior, data = slung3)


          coef exp(coef) se(coef)      z       p
Treat  0.1216     1.129   0.1019  1.193 2.3e-01
C1    -0.1590     0.853   0.1254 -1.268 2.0e-01
C2     0.2227     1.249   0.1290  1.726 8.4e-02
C3     0.3275     1.387   0.1231  2.660 7.8e-03
PS    -0.6140     0.541   0.1056 -5.814 6.1e-09
```

```
Month  0.0207     1.021   0.0965  0.215 8.3e-01
Prior  0.0190     1.019   0.1068  0.177 8.6e-01

Likelihood ratio test=60.2  on 7 df, p=1.35e-10  n= 136
```

The exponential coefficients of results are interpretable as multiplicative effects on the hazard or the estimation of the parameter also interpreted the same way as in parametric models, except no shape parameter is estimated because not making have assumptions about the shape of the hazard. Therefore holding the other covariates are constant increase a unit change in PS(Performance Status) reduces the timely hazard of lung cancer by a factor of $\exp\beta_5$=0.541, similar interpretation for Cell 1. And also holding the other covariates constant an increase unit change in Cell 3 increase the timely hazard of lung cancer by a factor of $\exp\beta_4$=1.387.
And using Coxpath algorithm could be simple to interpret the output, therefore figure.1 shows from the coefficient path with step, in the first step the variable PS(Performnce status) is included in the model, Cell 1 in the second step, in third the variable C3 in the model,in the fourth step C2 is included. Generally in the first step there is only one varlable, after two steps 2 variables join the active set, after four steps three variable as in the active set, and finally after 11 steps all the variables included in the model. In this case PS(Performance Status) and C1 appear to be most important, followed by C3, C2, treatment, Month and Prior.
In figure.2 by introduce penalized parameter find the coefficients that minimizes the log-likelihood function of β and λ and also it should be non-zero components of β. As lambda is decreased further and further the path starts and at the same time variables join the active set. From this graph, when λ=0 exists both the algorithm stop and the coefficient path has similar feature with Least Square regression.
From figure.3 can see the graph as λ decreased some amount of the step length the current solution set will be changed. In other words in every next largest value of λ at which the active set reduced. The coefficients were computed at some amount of different grids of λ. From the graph the vertical line indicates where the active set is modified.
In figure.4 and figure.5 based on the minimum value of AIC(Akaike Information Criterion) and BIC(Bayesian Information Criterion) five variables included in the model. The variables PS(Performance Status), C1(Cell1), C3(Cell3), C2(Cell2) and treatment are the most significant variables on the survival of lung cancer.

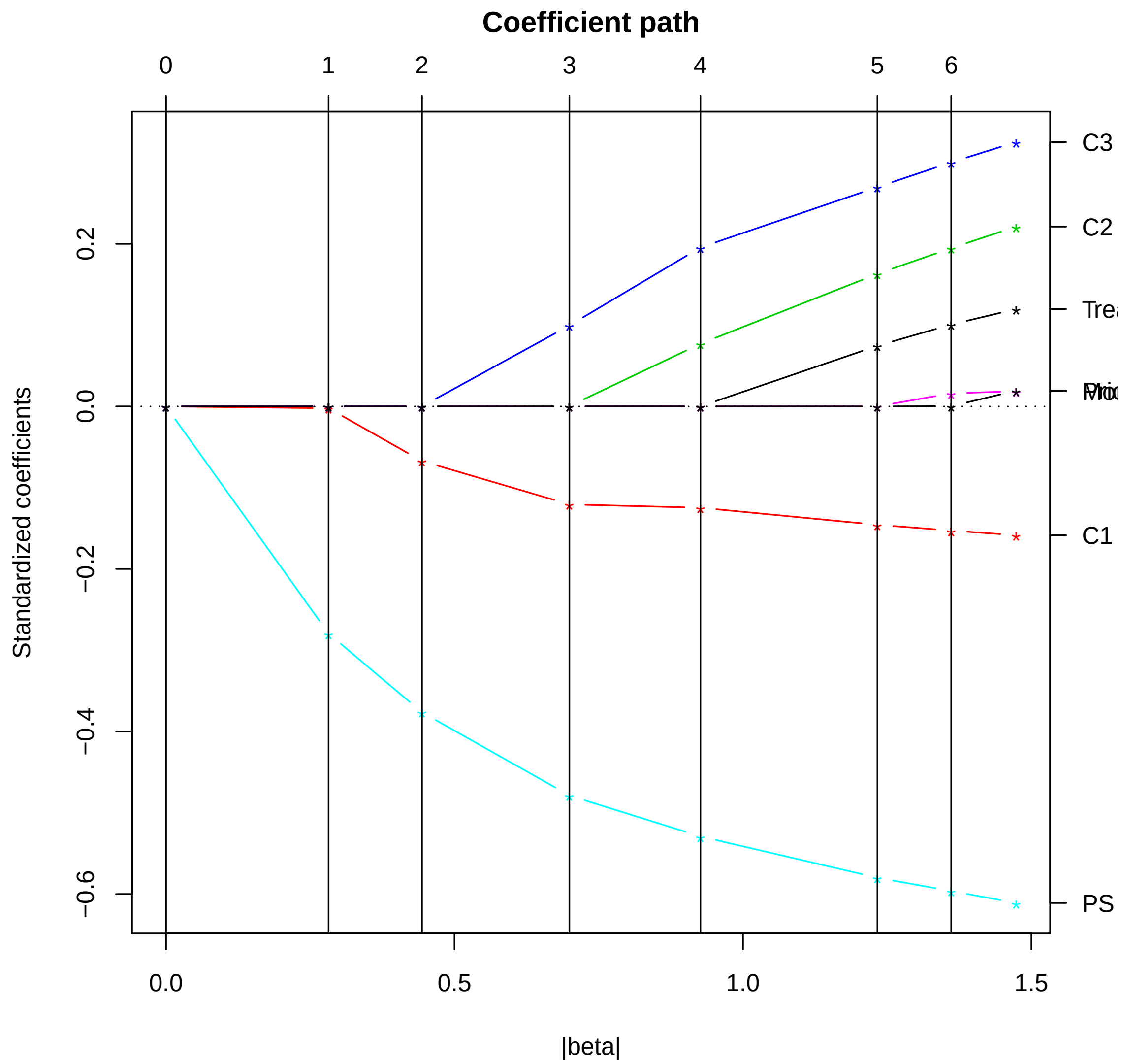


Figure 1: In the first step there is only one varlable, after two steps 2 variables join the active set,after four steps three variable as in the active set, and finally after 11 steps all the variables included in the model.

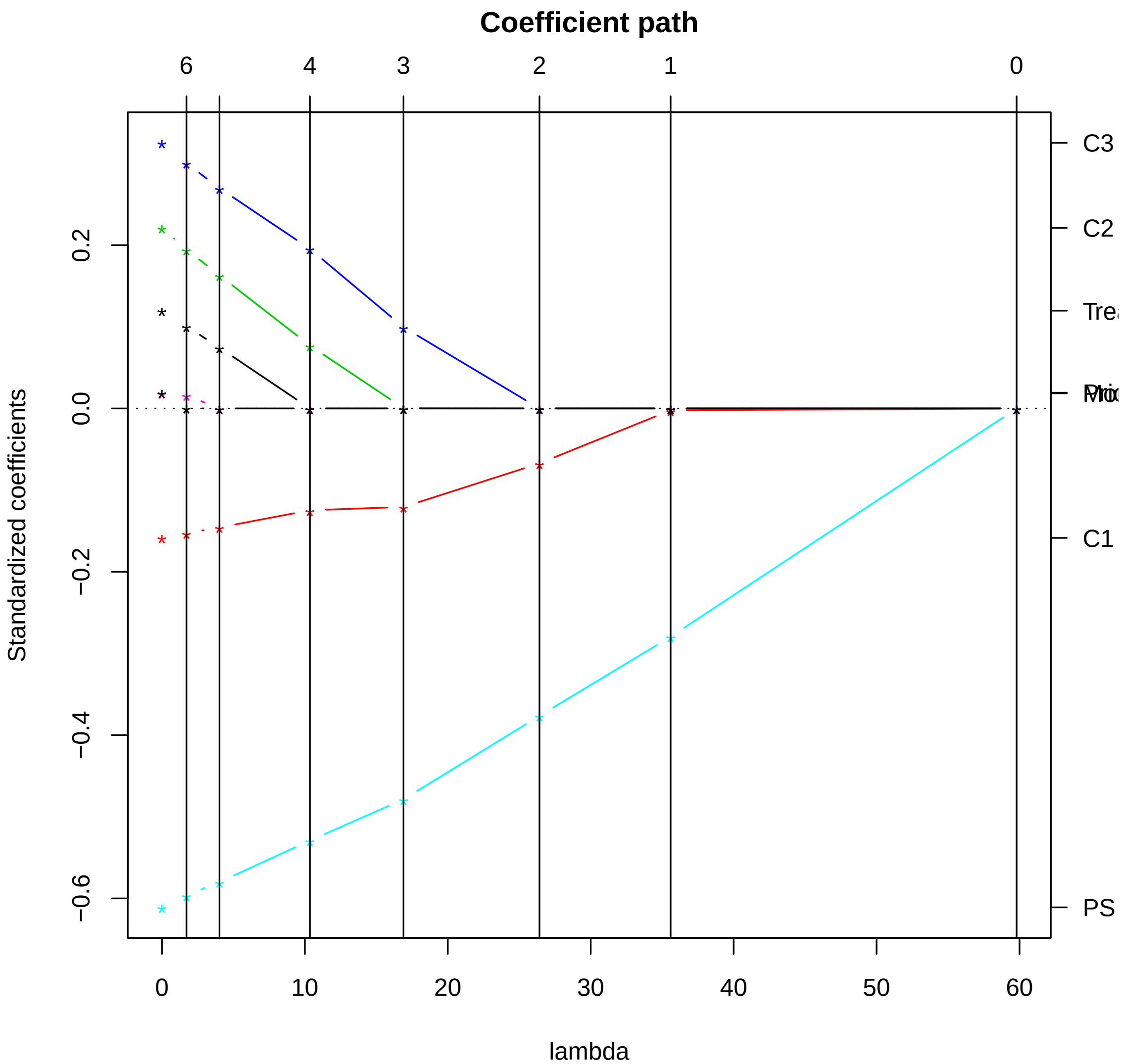


Figure 2: As lambda is decreased further and further the path starts and at the same time variables join the active set.

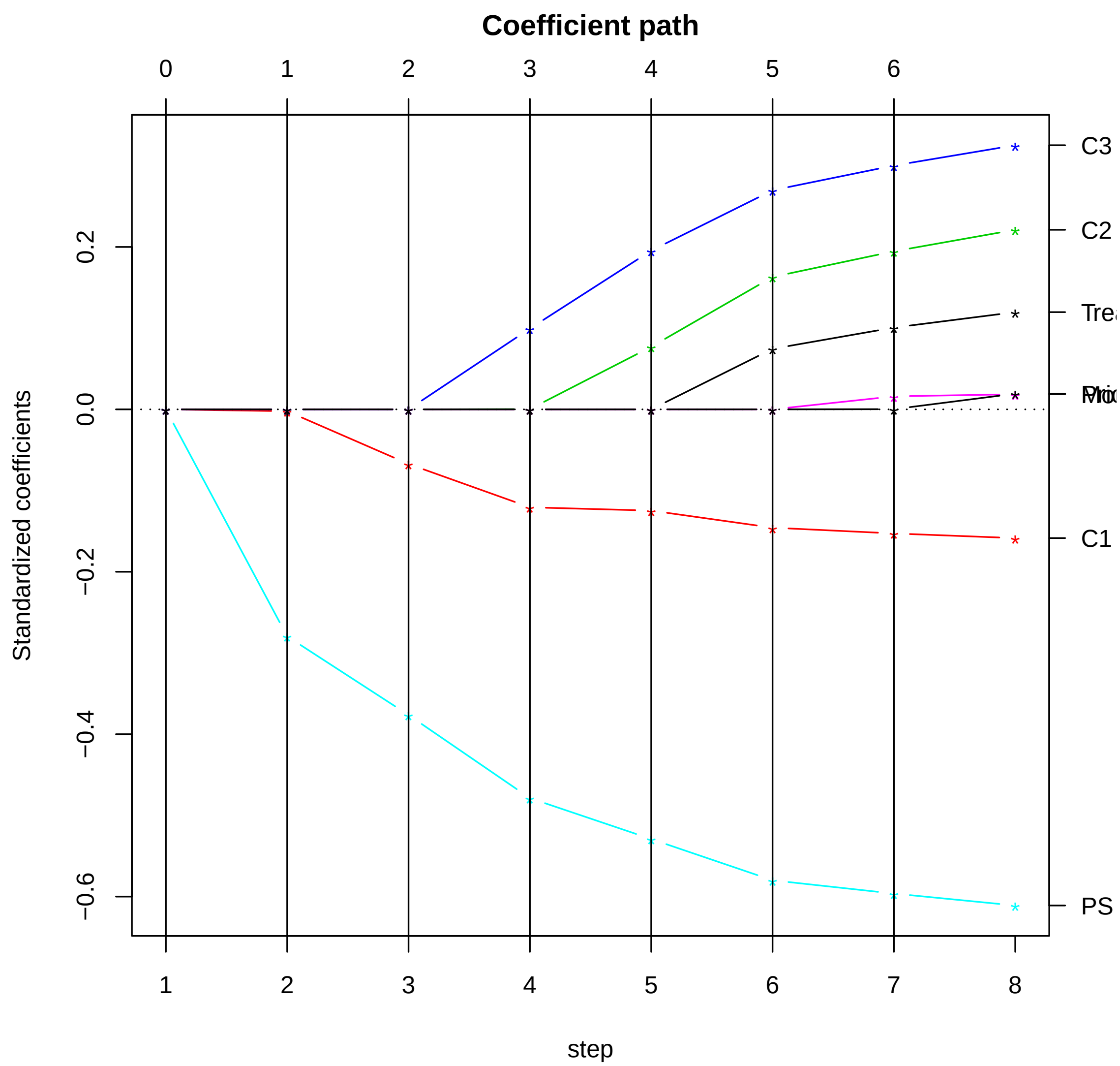


Figure 3: In the graph the vertical line indicates where the active set is modified.

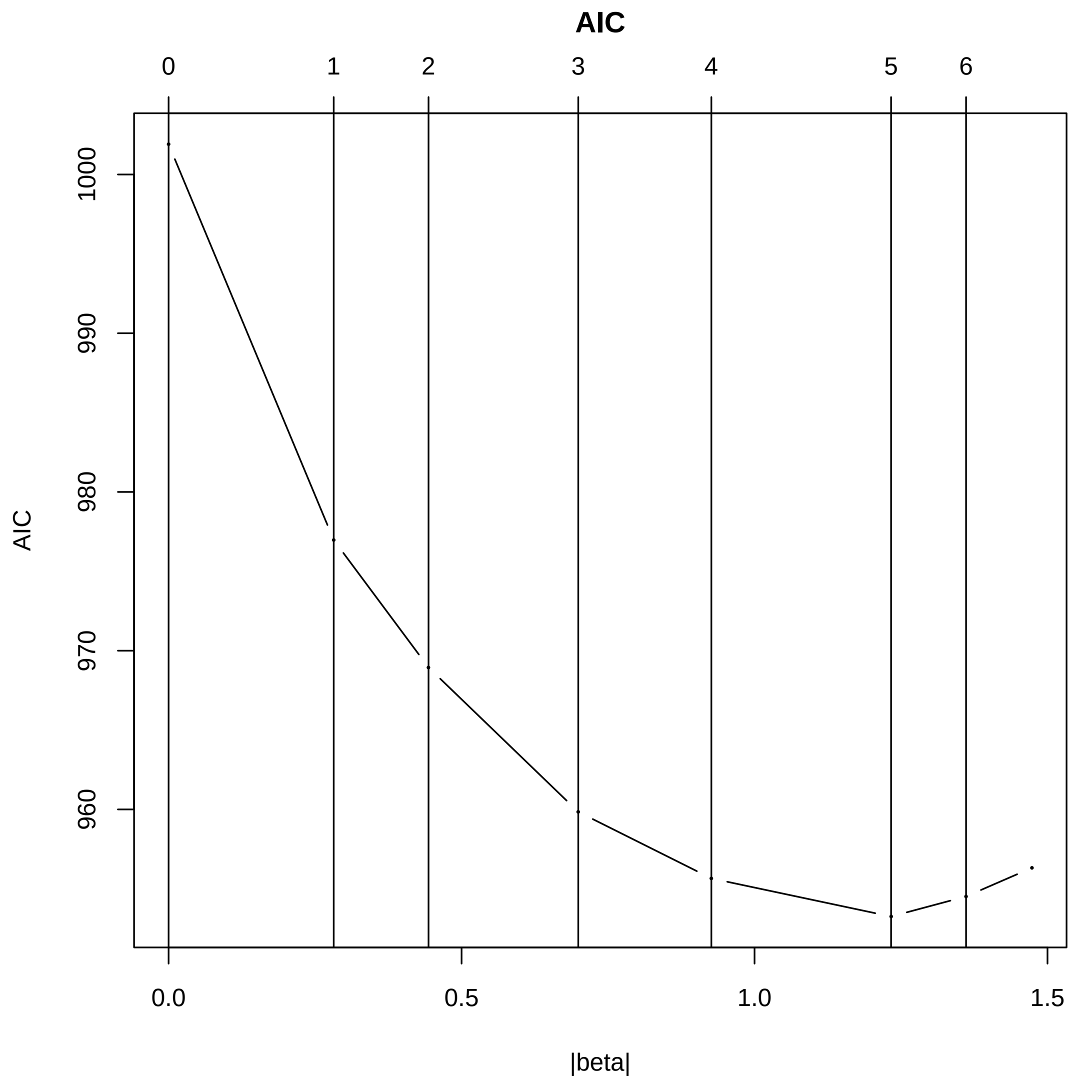


Figure 4: Based on the minimum value of AIC five variables included in the model. The variables PS,C1,C3,C2 and treatment are the most significant variables on the survival of lung cancer.

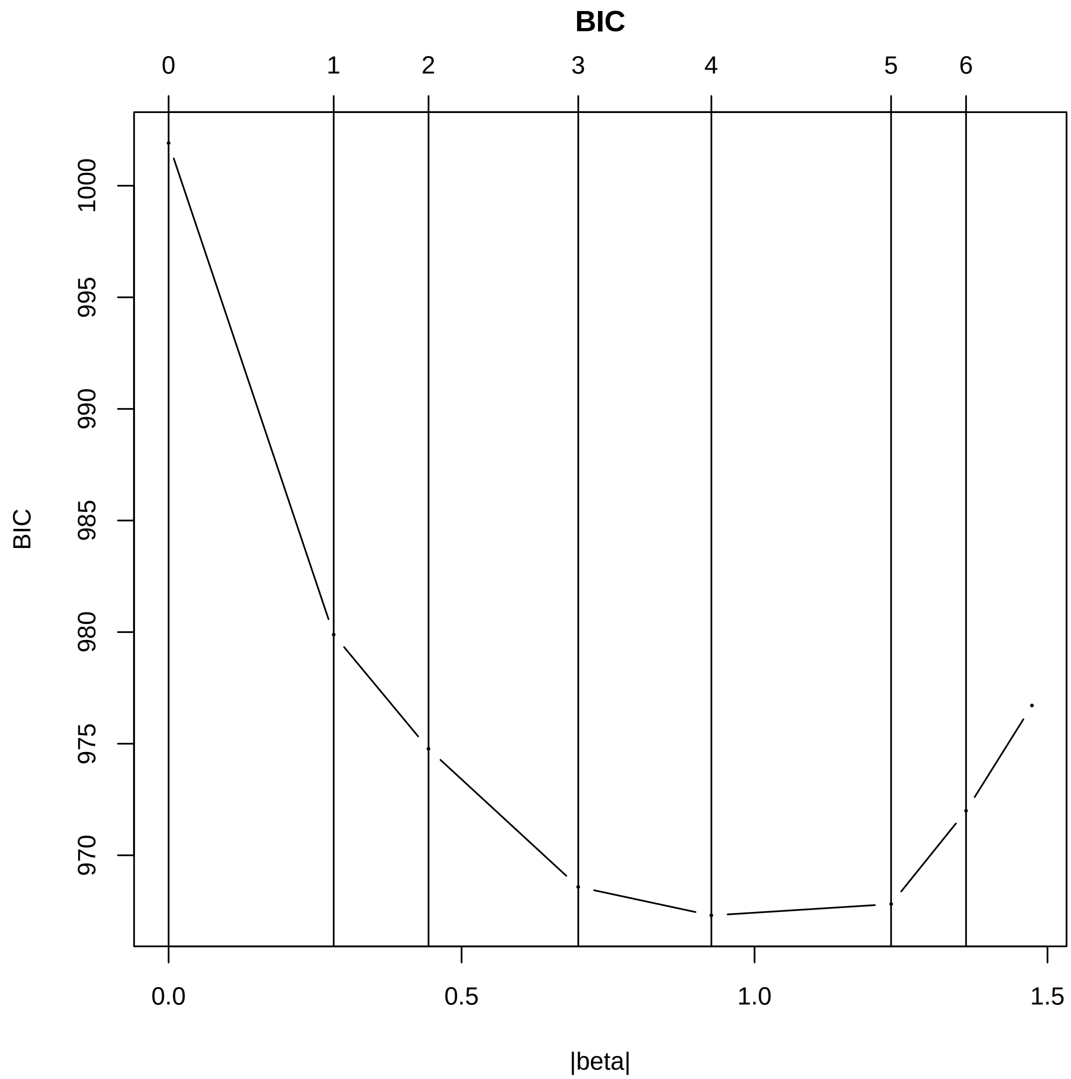
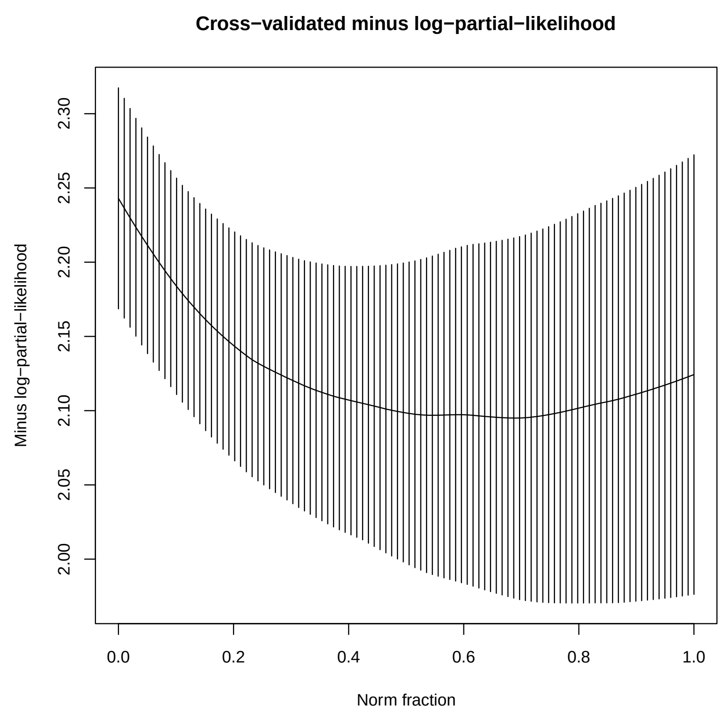


Figure 5: Based on the minimum value of BIC four variables included in the model



Cross−validated minus log−partial−likelihood
Minus log−partial−likelihood
2.00
2.05
2.10
2.15
2.20
2.25
2.30
0.0
0.2
0.4
0.6
0.8
1.0
Norm fraction

7.2 Analysis on Breast Cancer Data

Below the data set consists of 198 patients for both time to event. All patients received the same type of treatment, and recurrence is considered as the time from diagnosis of a disease to event. Those didn't exprience a recurrence before the end of the study are considered as Censored.

7.2.1 Breast Cancer Data on event A

Here is the second part of the project the data set consists of gene and treatment of 198 samples of breast cancer patients diagnosed at Denmark Hospital. Define survival as the event of interest in the Cox Proportional hazards model. The data set has 52 events and 146 Censored observation. The data classified like a variable genes, treatment and the survival time of breast cancer. Then I decided one of the statistical methods family that is regression analysis is appropriated for such kind of data set. Since in regression analysis I can see the relationship between the effect(response) in my case study survival time on cancer and the cause(covariates) which contains genes and clinical trials that is treatment. However by how much unit increase or decrease the amount of covariates when the response variable increase by a unit and by using the regression coefficient which of the most significant factor influence the survival time on cancer. But a type of regression analysis that is Ordinary Least Square does not work since has given high dimensional data. So I used glmpath algorithm to solve the given data. Therefore, based on this algorithm I got many graphs and interpreted as follows, Figure.6 shows a regularization path for a distribution negative likelihood with an L1-regulizer. The horizontal axis is the bound on the L_1 norm of the coefficient vector. Here after all the final step 149 variables are include in the model.
Figure.7 shows that since I used $\lambda \geq 0$, and from the plot I can see that when λ approaches 0 the Coxpath is approximately to Least Square estimation method while λ goes to infinity none of the variables including in the model or the only variable in the model is intercept there.
Figure.8 from the step graph results of Coxpath in full data, each colour of line represents the shrinked regression coefficient of one features and also in the graph the vertical lines indicates that each active variables are modified, however the number lies along on the lines indicates numbers of degrees of freedom in the upper and steps in down respectively. In addition to these, using the Coxpath generate the all data and I got 687 steps among these based on statistical criterion like AIC, the minimum one is the good one, i.e is up to 587 steps with 121 variables are most significant variables on the

survival of breast cancer.
The other two figures, figure.9 and figure.10 can see that, evaluate the model at step 587 which is suggested by partial likelihood version of AIC, then the resulting model includes 121 features. And Select an appropriate value of λ that yields the smallest BIC or AIC. But in my case AIC is the best one to choose the coefficients. Obviously BIC and AIC are doing the same purpose, but when the sample size is very small compare to the covariates AIC is much better than BIC since I have 198 samples with 17910 covariates, that is the reason why I chosen AIC.

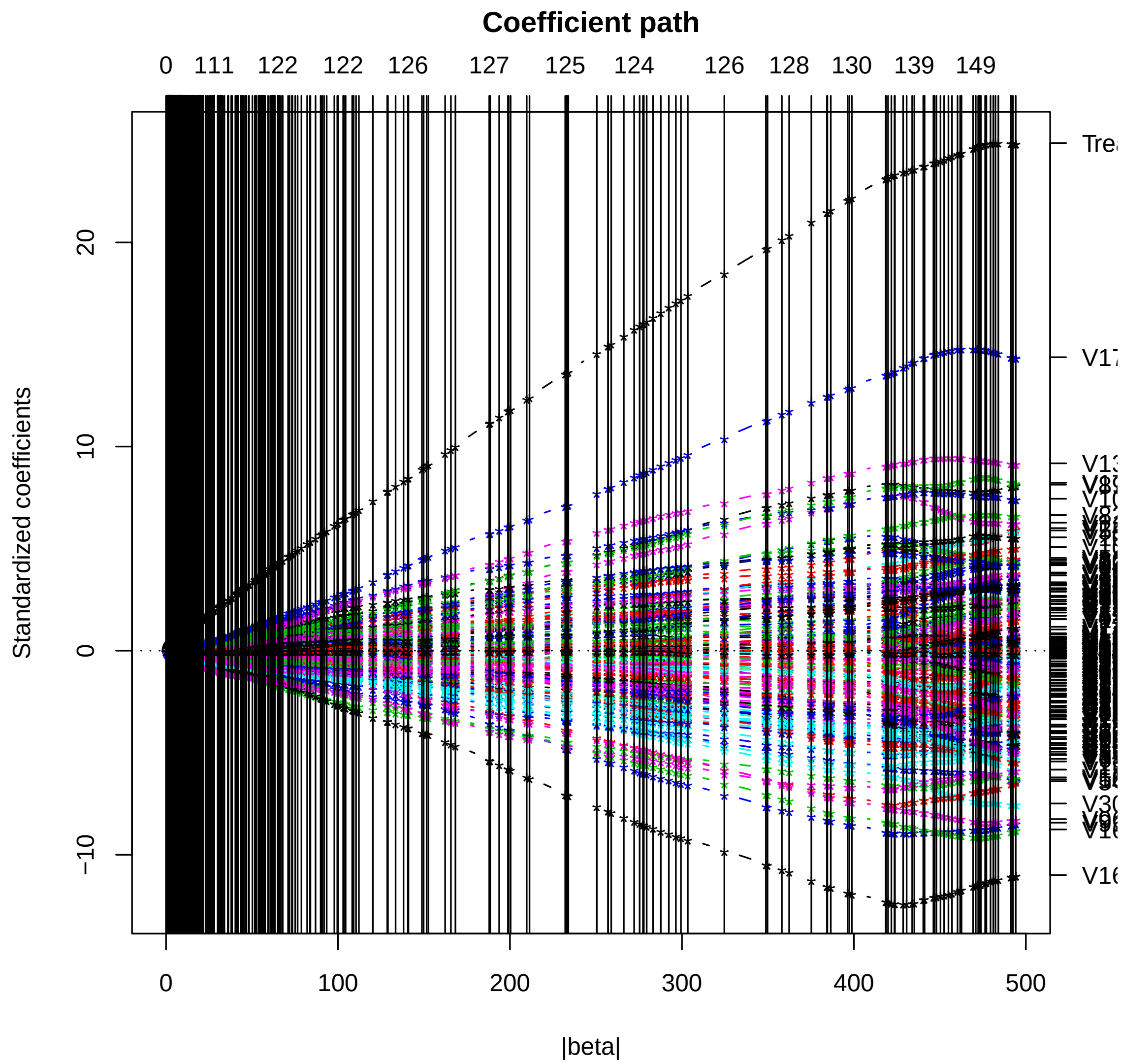


Figure 6: A regularization path for a distribution negative likelihood with an L1-regulizer path for Cox-proportional hazards.

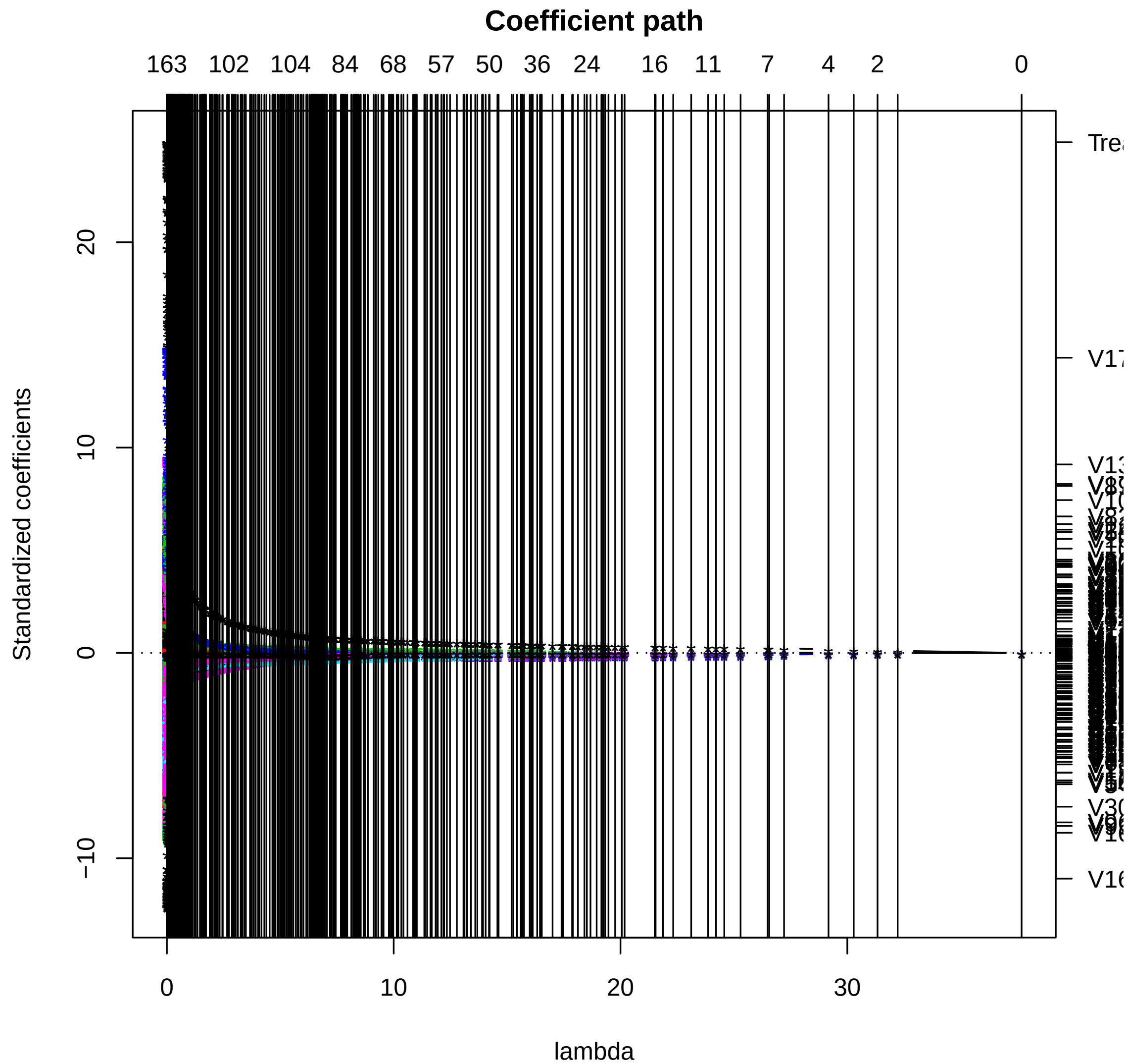


Figure 7: In this graph , since I used $\lambda \geq 0$,and when λ approaches 0 and the Coxpath approximately to Least Square estimation method while λ goes to infinity none of the variables including in the model may be the intercept will be there

Afterall the factor treatment is highly significant variables in the response on the time of event. Therefore I have done a model using only treatment for both time to event A and time to event B.

```
>fittreata<-coxph(Surv(ytimea,ystatusa)~factor(Treat),data=dataa)
> fittreata
Call:
coxph(formula = Surv(ytimea, ystatusa) ~ factor(Treat), data = dataa)


               coef exp(coef) se(coef)    z       p
factor(Treat)1 1.67      5.31    0.353 4.73 2.2e-06

Likelihood ratio test=29.2  on 1 df, p=6.66e-08  n= 198
>fittreatb<-coxph(Surv(ytimeb,ystatusb)~factor(Treat),data=datab)
> fittreatb
Call:
coxph(formula = Surv(ytimeb, ystatusb) ~ factor(Treat), data = datab)


                coef exp(coef) se(coef)    z    p
factor(Treat)1 0.293      1.34    0.186 1.58 0.11

Likelihood ratio test=2.49  on 1 df, p=0.114  n= 198
```

Therefore, for time to event A: The coefficient for treatment 1.67 is the hazard ratio for a patient given not treated compared with a patient give treated. Whereas for time to event B: The coefficient for treatment 0.293 is the hazard ratio for a patient given not treated compared with a patient given treated.

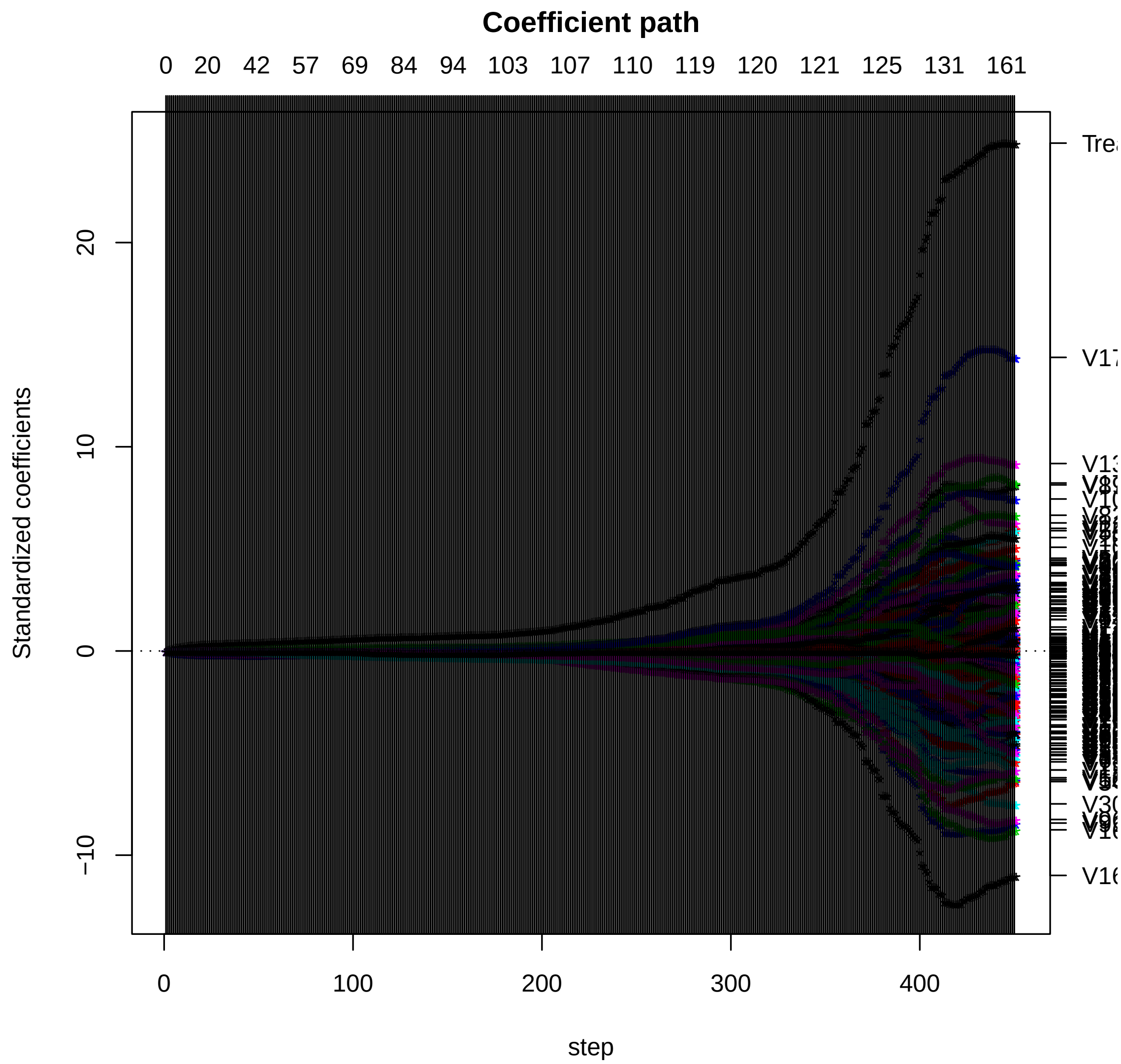


Figure 8: In the step graph results of CoxPath in full data, each colour of line represents the shrinked regression coeefficient of one features.

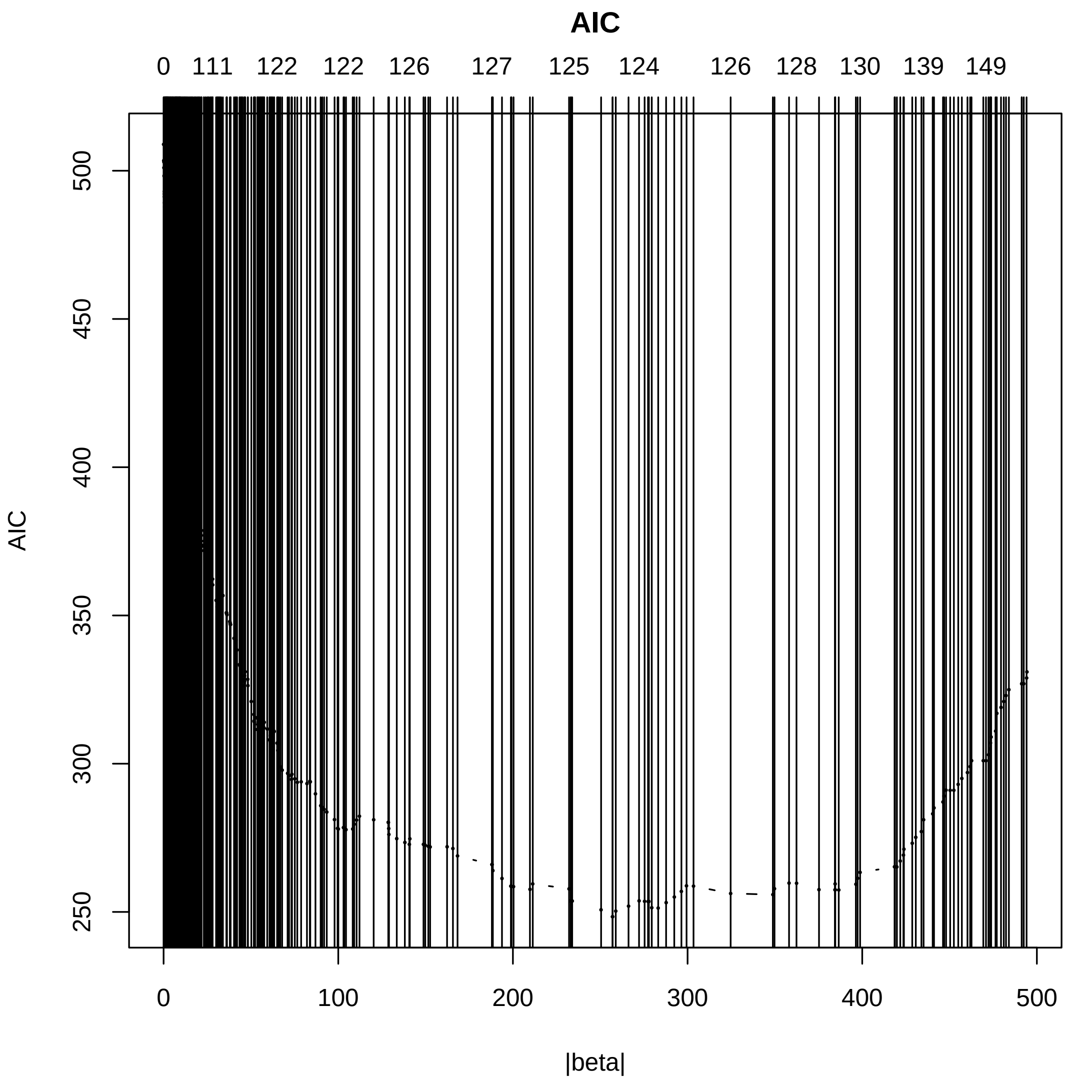


Figure 9: Evaluate the model at step 587 which is sugested by partial likelihood version of AIC, then the resulting model includes 121 features.

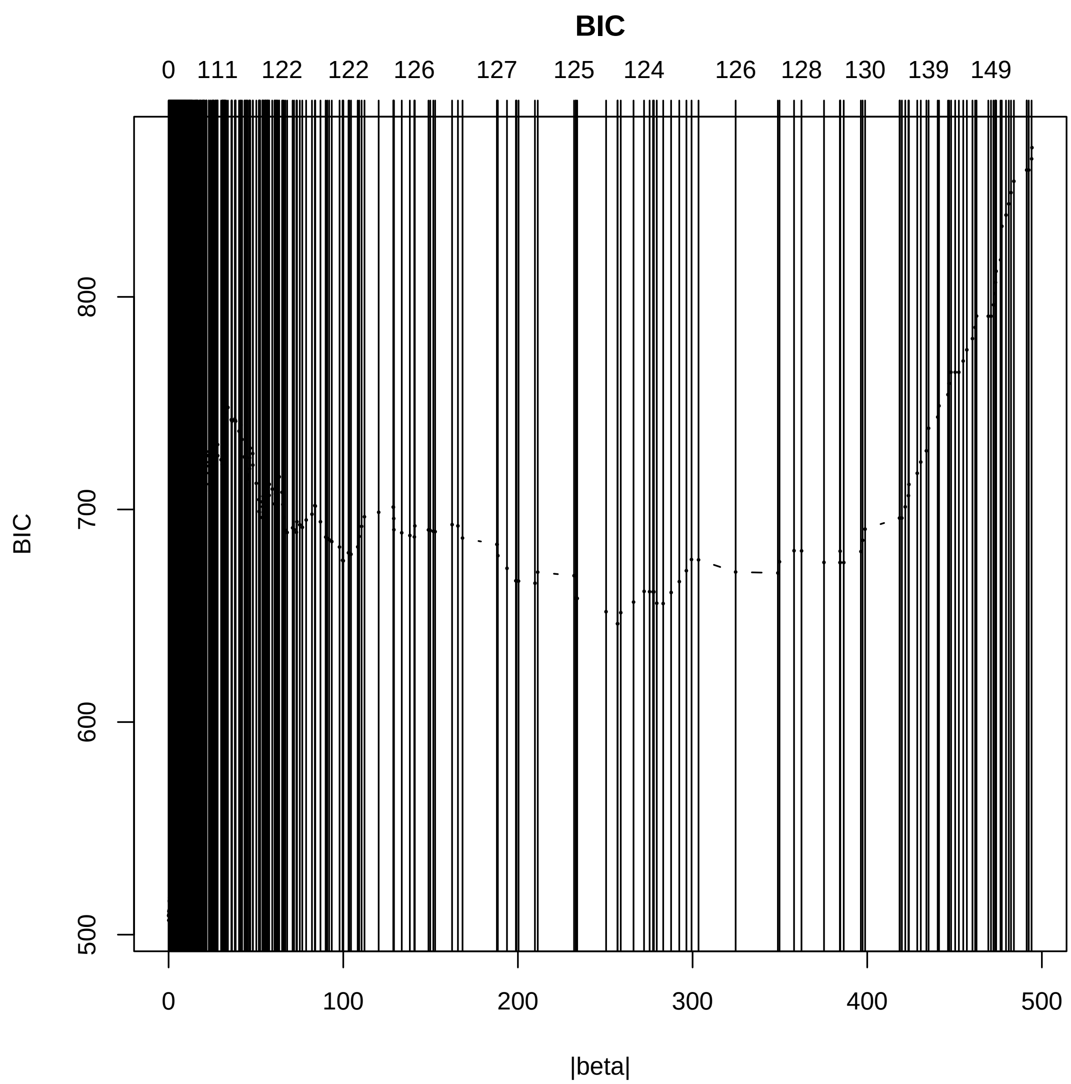


Figure 10: Select an appropriate value of λ that yields the smallest BIC or AIC.

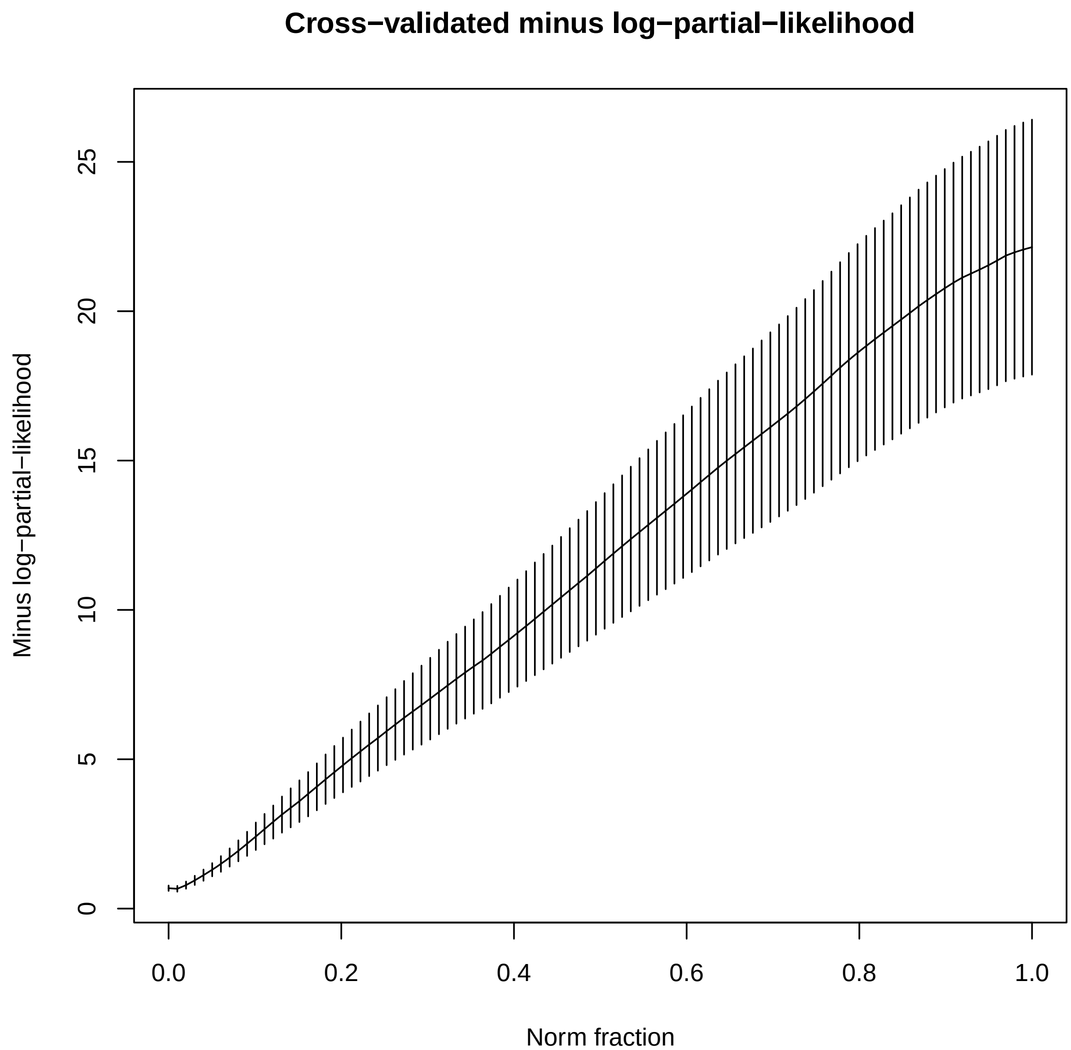


Figure 11: Here I can see 10-fold CV error curve using lasso of breast cancer data(17910 variables with 198 observations).

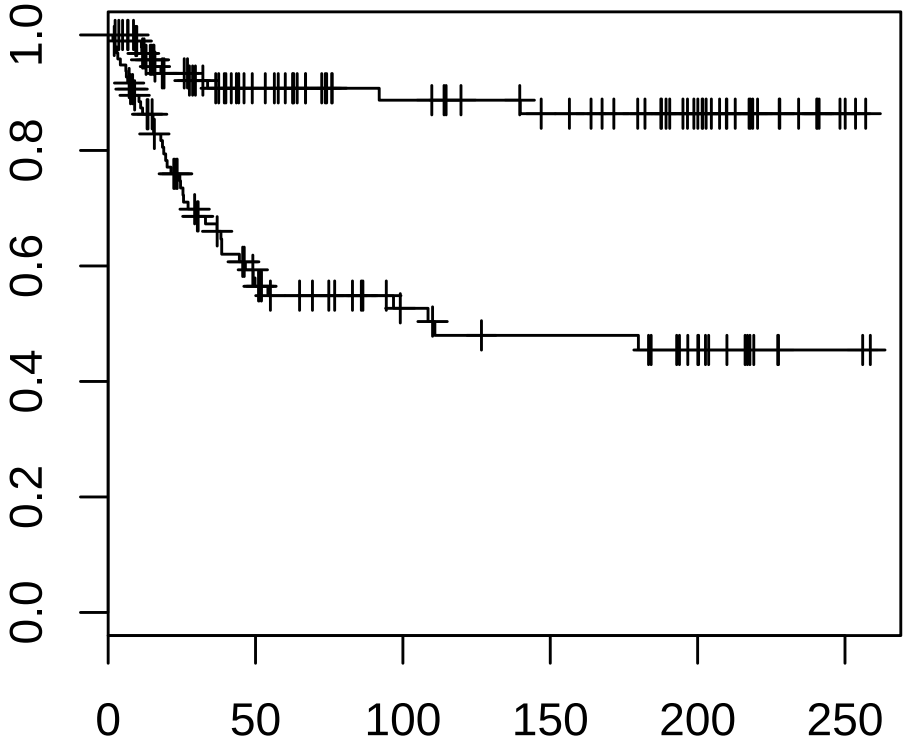


Figure 12:

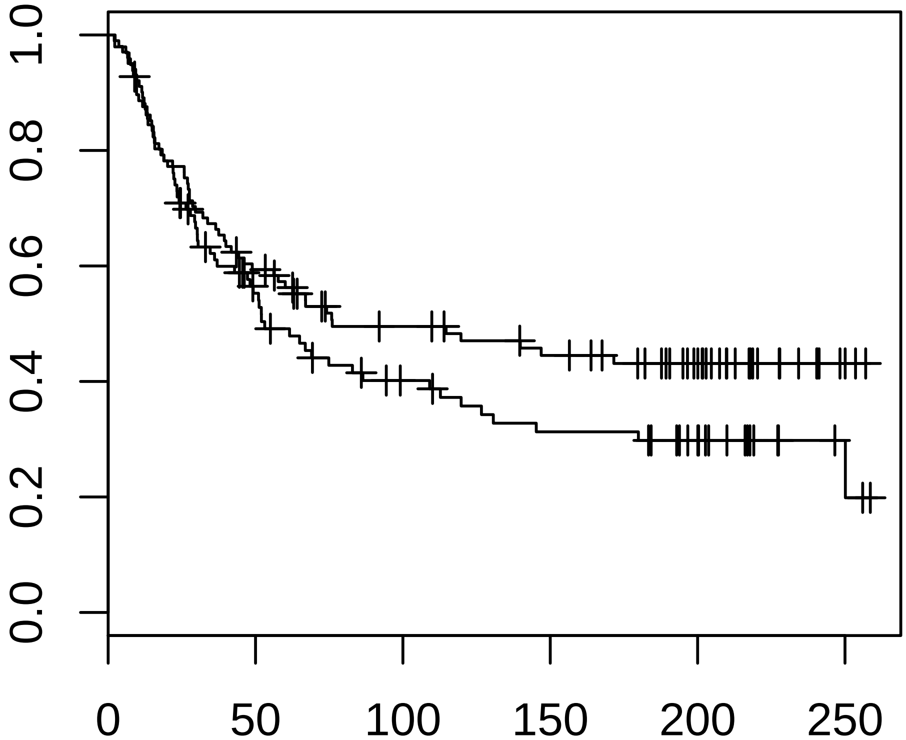
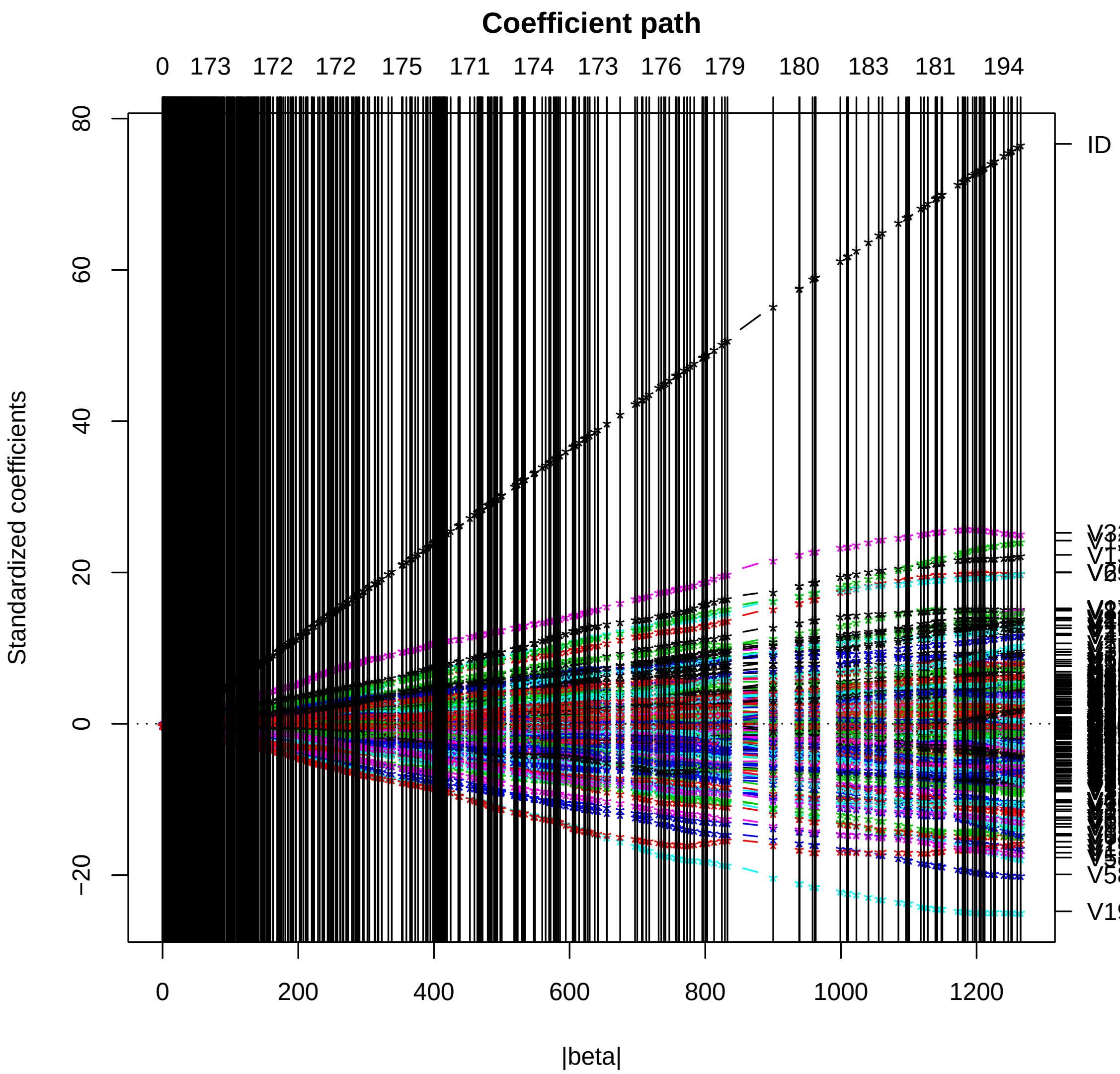


Figure 13:

And from the survival plot, I can see that in both the estimated curves are not crossed so that I have confidently speaking the hazard are proportional, the assumptions of Cox Proportional hazard are proved.

7.2.2 Breast Cancer Data on event B

The data set has 17910 genes and treatment per sample and 198 samples which consisted 176 events and 22 Censored observation. But the time to event for B takes so much time to execute the algorithm even could not see all the result or selections of the variable steps in the visual window.



Coefficient path
0 173 172 172 175 171 174 173 176 179 180 183 181 194
80
60
40
20
0
−20
Standardized coefficients
0 200 400 600 800 1000 1200
|beta|
ID

Figure 14:

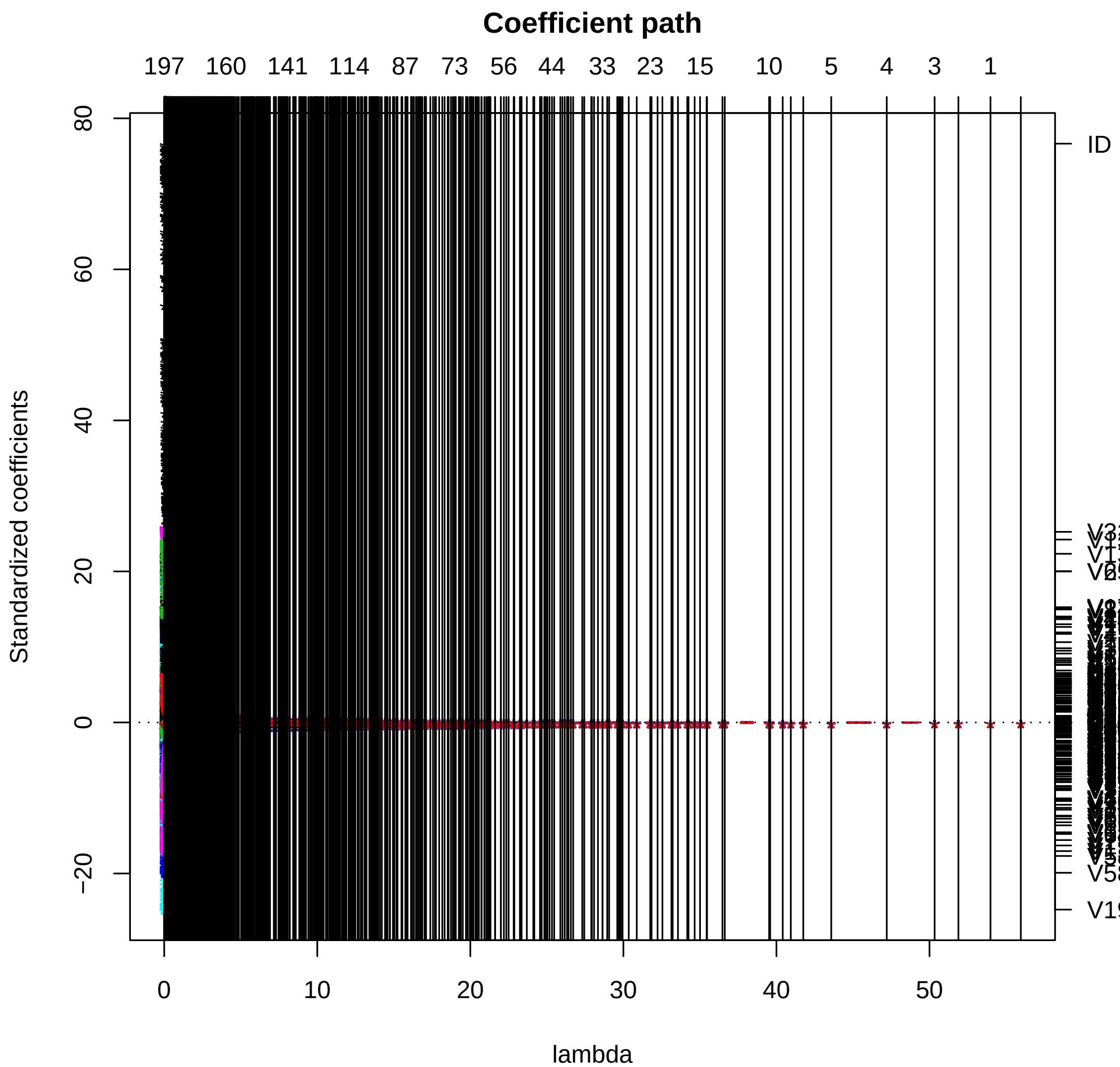

Coefficient path
197 160 141 114 87 73 56 44 33 23 15 10 5 4 3 1
Standardized coefficients
80
60
40
20
0
−20
0 10 20 30 40 50
lambda
ID

Figure 15:

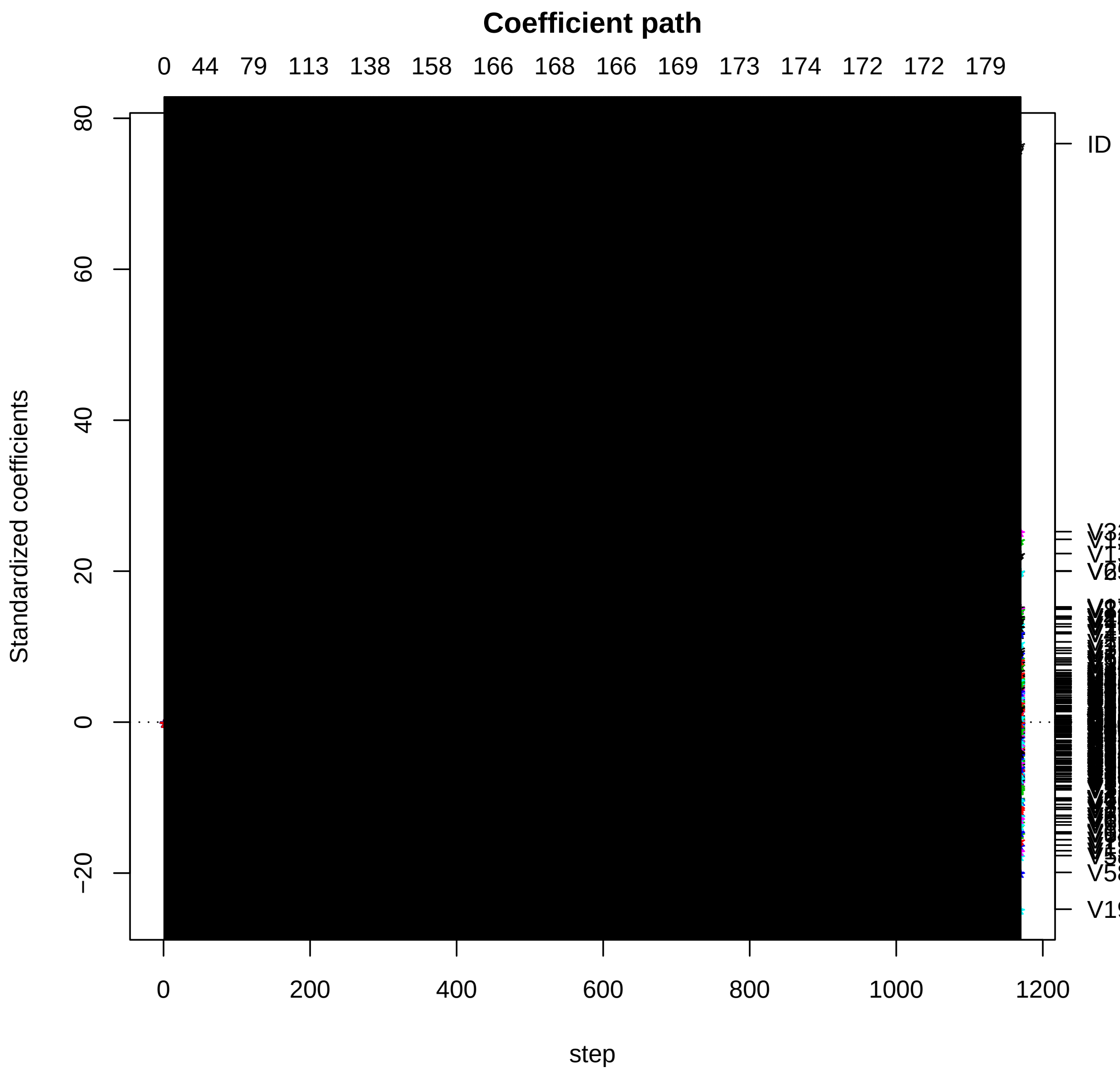

Coefficient path
0 44 79 113 138 158 166 168 166 169 173 174 172 172 179
Standardized coefficients
80
60
40
20
0
−20
0
200
400
600
800
1000
1200
step
ID

Figure 16:

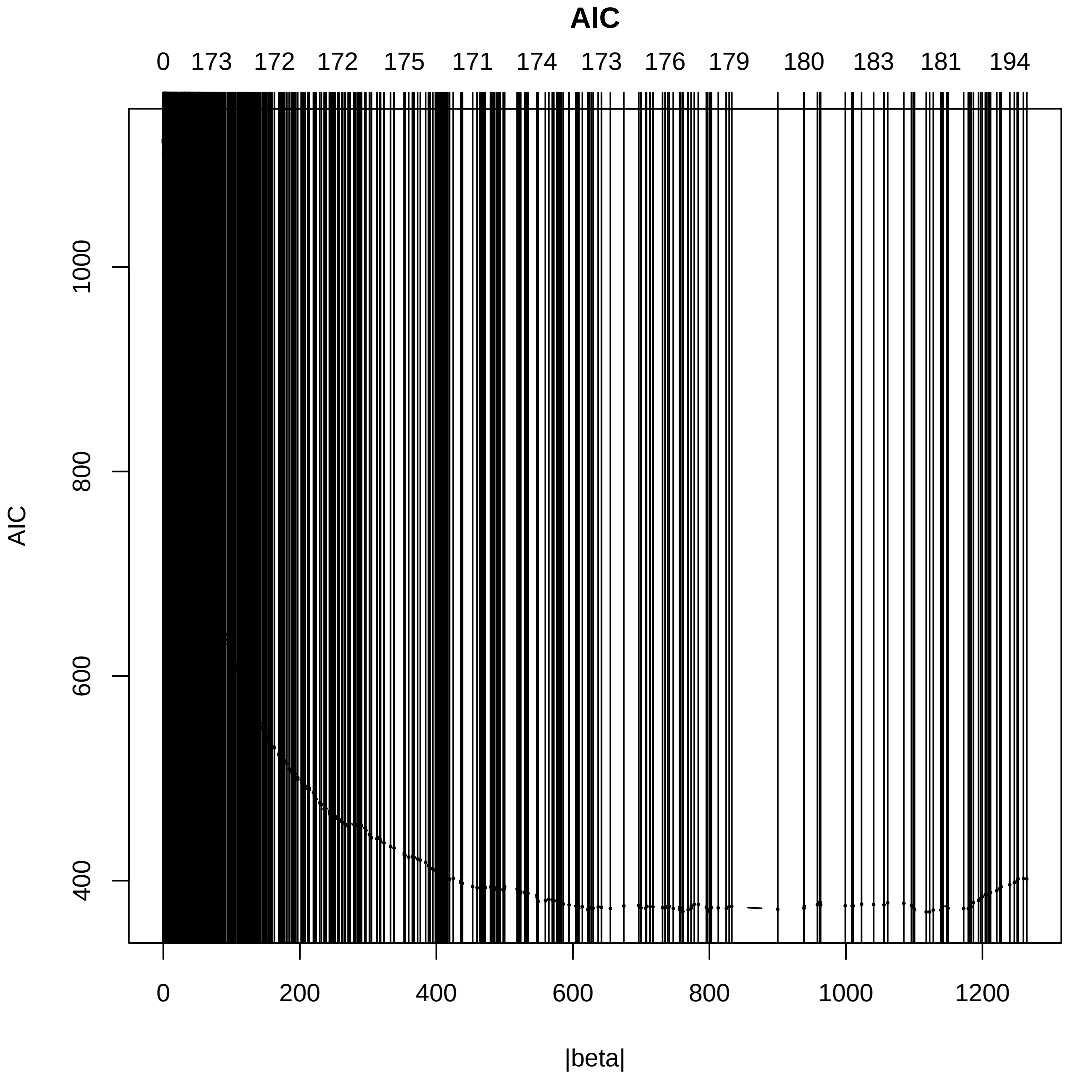
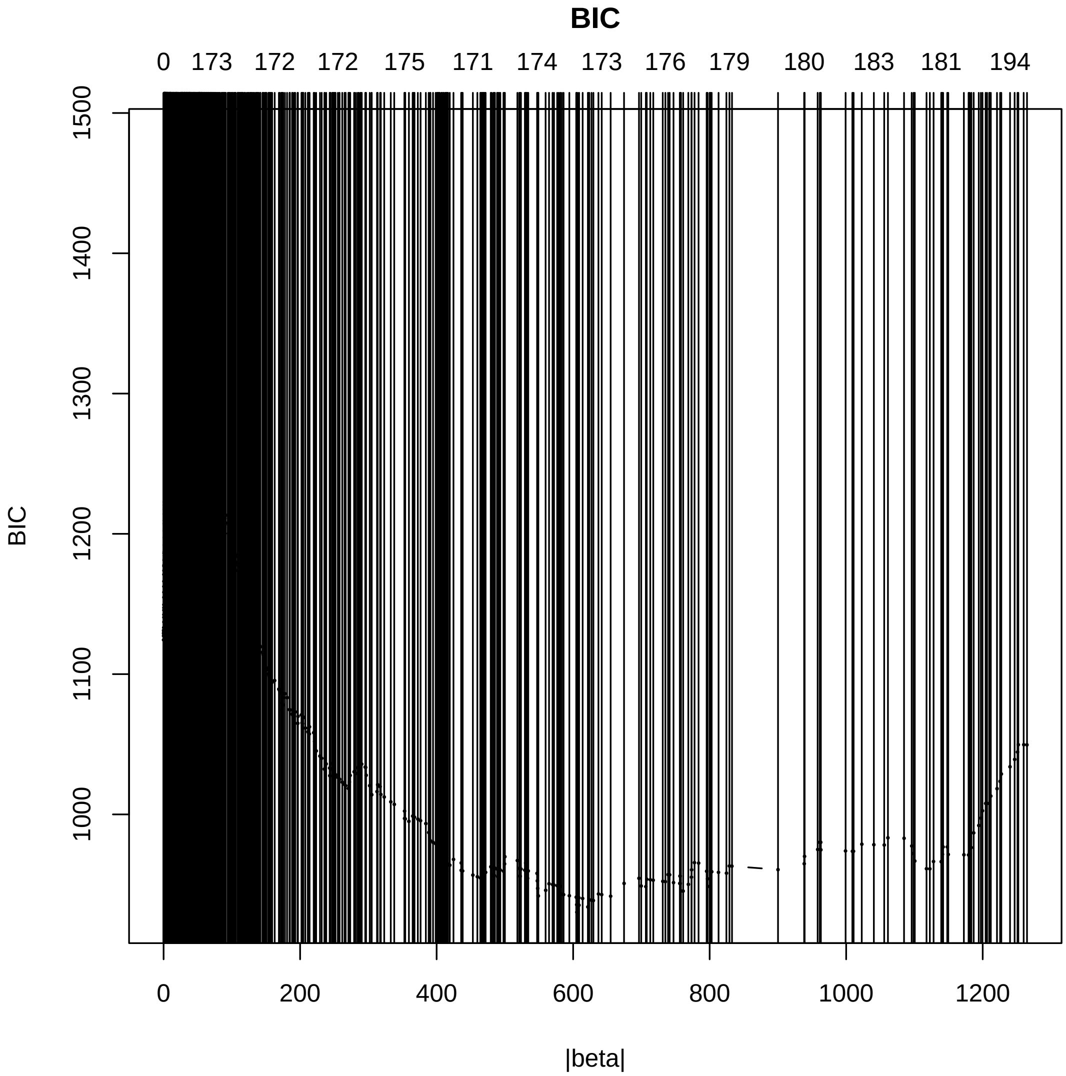

AIC
0 173 172 172 175 171 174 173 176 179 180 183 181 194
1000
800
600
400
AIC
0 200 400 600 800 1000 1200
|beta|

Figure 17:



BIC
0 173 172 172 175 171 174 173 176 179 180 183 181 194
1500
1400
1300
1200
1100
1000
BIC
0 200 400 600 800 1000 1200
|beta|

Figure 18:

Acknowledgement

Thanks the Almighty God. It only remains to thank the people who have helped me along the way and some of them I want to mention by name. First and foremost, I would like to thank my supervisor Professor Bo Lindqvist for advice and criticism has been extremely helpful with the cocept and writing during the course of this work. He has always had time for discussions and provided me with invaluable guidance. I would also like to thank who has been my co-supervisor Professor Arnaldo Frugessi for useful comments especially in the beginning of the processs. I thank a Phd student Hayat Mohammed for providing and editing the data of the Breast Cancer. I thank also Sara Martino for useful discussion on simulation techniques.
Special thanks for Professor Henning Omre, you have supported me at every stage, from my first arrival here Trondheim Norway to the final completion of my thesis. He has always been there to help make things possible; I say Thank You. And also I am grateful to Dr.Eng.Ayele Taye, Mathematical and Physical Sciences at Hawassa University.
I am also please to thank the institution of both Hawassa University and Norwegian University of Science and Technology for granted me MASTMO program.
The deepest thanks go to my family and my friends. I say, Thank you to all for staying close, despite the distance.
Thank you all
Mahder

References

[1] Hui Zou, Trevor Hastie, and Robert Tibshirani(2007) On the Degrees of Freedom of the Lasso ,*Annals of Statistics*. Vol 35, Number 5, 2173-2192

[2] Jianqing Fan and Runze Li(2002) Variable Selection for Cox's Proportional Hazards Model and Frailty model *Chinese University of Hong Kong and Pennsylvania State University*, The Annals of statistics, Vol.30, No.1, 74-99

[3] J. Friedman, T. Hastie and R. Tibshirani.(2009) Regularization Paths For Generalized Linear Models via Coordinate Descent. Department of Statistics, Stanford University

[4] John Fox(2002) Cox Proportional-Hazards Regression for Survival Data *Appendix to An R and S-PLUS Companion to Applied Regression*

[5] Lucila Ohno-Machado(1996),A comparison of Cox proportional hazards and artificial neural network models for medical prognosis.*Decision Systems Group,Brigham and Women's Hospital. Harhavrd Medical School, 75 Francis St, Boston,MA 02115, U.S.A.*

[6] Mee Young Park and Trevor Hastie.(2007) L_1-regularization path algorithm for generalized linear models.J.R. Statist. Soc. B(2007)**69**, Part 4,pp.659-677

[7] Shao, J. (1997) An asymptotic theory for linear model selection. *Statistica Sinica*, **7**, 221-242

[8] Sugiure, N, (1978). Further Analysis of the data by Akaike's Information Criterion and finite corrections *Comm.Statist* **A7** *13-26*

[9] Tibshirani, R.(1996). Regression shrinkage and selection via the lasso, J. Royal. Statist Soc B, Vol. 58, No.1,pages 267-288

[10] Tibshirani, R.(1997).The Lasso method for variable selection in cox model, Statistics in Medicine, Vol.16,pp 385-395

[11] Yuhong Yang(2003) Regression with multiple candidate models: selecting or mixing? *Statistica Sinica,* **9**, 475-499

[12] Zhenqiu Liu, et.al.(2007),Sparse Logistic Regression with L_p Penalty for Biomarker Identification ,*Journal Statistical Applications in Genetics and Molecular Biology* Volume 6, Issue 1.

Appendix

```
> getwd()
[1] "/home/sylow/a/tuji"
> setwd("/home/sylow/a/tuji")
> library(glmpath)
Loading required package: survival
Loading required package: splines
> library(survival)
> library(base)
> slung3<-read.table("standardlung.txt",header=T)
> sxlung3<-as.matrix(slung3[,1:7])
> sytime3<-as.vector<-(slung3[,8])
> systatus3<-as.vector<-(slung3[,9])
> sfitlung3<-coxph(Surv(ytime3,ystatus3)~Treat+C1+C2+C3+PS+Month+Prior,data=slung3)
> sfitlung3
Call:
coxph(formula = Surv(ytime3, ystatus3) ~ Treat + C1 + C2 + C3 +
    PS + Month + Prior, data = slung3)


         coef exp(coef) se(coef)      z       p
Treat  0.1216     1.129   0.1019  1.193 2.3e-01
C1    -0.1590     0.853   0.1254 -1.268 2.0e-01
C2     0.2227     1.249   0.1290  1.726 8.4e-02
C3     0.3275     1.387   0.1231  2.660 7.8e-03
PS    -0.6140     0.541   0.1056 -5.814 6.1e-09
Month  0.0207     1.021   0.0965  0.215 8.3e-01
Prior  0.0190     1.019   0.1068  0.177 8.6e-01

Likelihood ratio test=60.2  on 7 df, p=1.35e-10  n= 136
> slung3list<-list(x=sxlung3,time=sytime3,status=systatus3)
> class(slung3list)
[1] "list"
> attach(slung3list)

The following object(s) are masked _by_ .GlobalEnv :

 x

> scvlung3<-cv.coxpath(slung3list)
```

```
CV Fold 1
CV Fold 2
CV Fold 3
CV Fold 4
CV Fold 5
> dev.copy2eps(file="scvlung3.eps")
X11cairo
       2
> slung3coxpath<-coxpath(slung3list)
> slung3coxpath
Call:
coxpath(data = slung3list)
Step 1 :  PS
Step 2 :  C1
Step 4 :  C3
Step 5 :  C2
Step 7 :  Treat
Step 10 :  Month
Step 11 :  Prior
> plot(slung3coxpath)
> dev.copy2eps(file="slung3coxpath.eps")
X11cairo
       2
> plot(slung3coxpath,xvar="lambda")
> dev.copy2eps(file="slambda3.eps")
X11cairo
       2
> plot(slung3coxpath,xvar="step")
> dev.copy2eps(file="sstep3.eps")
X11cairo
       2
> plot(slung3coxpath,xvar="step",xlimit=20)
> dev.copy2eps(file="slimit3.eps")
X11cairo
       2
> plot(slung3coxpath,type="aic")
> dev.copy2eps(file="saic.eps")
X11cairo
       2
> plot(slung3coxpath,type="bic")
> dev.copy2eps(file="sbic.eps")
```

```
X11cairo
       2

>getwd()
>setwd("/home/sylow/a/tuji")
>library(glmpath)
>library(survival)
>library(base)
>co<-as.matrix(read.table("fullcov.txt"))
>tco<-t(co)
>header=tco[1,]
>tco<-matrix(as.numeric(tco[-1,]),198,17912)
>colnames(tco)=header
>evA<-read.table("eventAA.txt",header=T)
>ytime=as.vector(evA[,2])
>ystatus=as.vector(evA[,3])
>covlist<-list(x=tco,time=ytime,status=ystatus)
>covlistA<-list(x=tco,time=ytime,status=ystatus)
>coxpathA<-coxpath(covlistA,add.newvars=1,bshoot.threshold=0.1,relax.lambda=1e-7,st
> coxpathA
Call:
coxpath(data = covlistA, add.newvars = 1, bshoot.threshold = 0.1,
    relax.lambda = 1e-07, standardize = TRUE)
Step 1 :  Treat
Step 2 :  V13450
Step 4 :  V713
Step 5 :  V7803
Step 6 :  V935
Step 8 :  V15013
Step 9 :  V3892
Step 10 :  V17090
Step 12 :  V8045
Step 14 :  V10831
Step 16 :  V8785
Step 17 :  V4291
Step 19 :  V8586
Step 21 :  V3096
Step 23 :  V5825
Step 25 :  V3485
Step 27 :  V3178
Step 28 :  V228
```

```
Step 30 :  V6053
Step 31 :  V11822
Step 33 :  V17602
Step 34 : - V3178
Step 35 :  V11287
Step 36 :  V386
Step 38 :  V10839
Step 39 :  V9832
Step 41 :  V16110
Step 43 :  V4807
Step 45 :  V16423
Step 47 :  V7340
Step 49 :  V1127
Step 50 :  V7790
Step 52 : - V9832
Step 53 :  V14845
Step 55 :  V2204
Step 57 :  V16362
Step 58 :  V9302
Step 59 :  V17398
Step 61 :  V12241
Step 62 :  V13027
Step 63 :  V12455
Step 65 :  V13388
Step 67 : - V10831
Step 68 :  V17747
Step 70 :  V9832
Step 72 :  V15478
Step 74 :  V5152
Step 75 :  V9119
Step 77 :  V9796
Step 78 :  V5998
Step 79 :  V4459
Step 81 :  V8297
Step 83 :  V7678
Step 85 :  V8941
Step 86 :  V5711
Step 88 :  V1929
Step 89 :  V17357
Step 90 : - V228
Step 92 :  V10508
```

```
Step 94 :  V5540
Step 96 :  V16231
Step 97 :  V14805
Step 99 :  V13847
Step 100 : - V17090
Step 101 :  V5593
Step 103 :  V17158
Step 104 :  V5496
Step 106 :  V4112
Step 107 : - V3485
Step 109 :  V15307
Step 110 : - V16362
Step 112 :  V9289
Step 113 : - V10839
Step 114 :  V15292
Step 115 : - V16110
Step 116 :  V12197
Step 117 : - V713
Step 119 :  V9671
Step 121 :  V15177
Step 123 :  V7863
Step 125 :  V4529
Step 127 :  V14168
Step 128 :  V16880
Step 130 :  V12255
Step 131 : - V12455
Step 133 :  V12272
Step 135 :  V606
Step 136 :  V5276
Step 138 :  V420
Step 140 : - V13847
Step 142 : - V12197
Step 144 :  V3259
Step 146 :  V1807
Step 147 :  V228
Step 149 :  V1394
Step 150 : - V4459
Step 151 : - V11822
Step 153 :  V17878
Step 154 :  V10779
Step 155 :  V4349
```

Step 157 : - V7340
Step 158 : V7206
Step 159 : - V935
Step 160 : V10841
Step 161 : V12197
Step 163 : V17620
Step 165 : V16362
Step 166 : - V228
Step 167 : V6130
Step 169 : V2797
Step 170 : V10182
Step 172 : V8841
Step 173 : - V14845
Step 174 : V726
Step 176 : V14815
Step 177 : V17851
Step 178 : - V8045
Step 180 : V6198
Step 181 : V17052
Step 183 : V17045
Step 185 : V3019
Step 187 : V458
Step 189 : V2400
Step 190 : V2025
Step 192 : V1606
Step 194 : - V15307
Step 195 : V16482
Step 197 : V322
Step 199 : V3740
Step 200 : V6135
Step 202 : V8672
Step 203 : - V16423
Step 204 : V7156
Step 205 : V935
Step 207 : V5410
Step 208 : V10944
Step 209 : - V16482
Step 211 : V1338
Step 212 : - V3259
Step 213 : - V420
Step 214 : - V4112

```
Step 216 :  V8584
Step 217 : - V8586
Step 218 : - V13388
Step 220 :  V4472
Step 221 : - V9302
Step 222 :  V6814
Step 224 :  V3117
Step 226 :  V4477
Step 228 :  V17266
Step 230 :  V16961
Step 231 :  V17354
Step 233 :  V16234
Step 235 :  V8873
Step 237 :  V5350
Step 239 :  V17289
Step 240 : - V13450
Step 242 : - V15177
Step 243 :  V12625
Step 244 :  V16034
Step 245 :  V10734
Step 247 :  V12063
Step 249 :  V12099
Step 251 : - V12625
Step 252 :  V3486
Step 253 :  V1065
Step 255 :  V994
Step 256 :  V6054
Step 257 :  V10035
Step 259 : - V5711
Step 260 : - V10734
Step 261 :  V12625
Step 262 :  V7427
Step 264 : - V7156
Step 266 :  V5892
Step 267 : - V3019
Step 269 :  V10414
Step 271 : - V16034
Step 272 : - V17354
Step 274 :  V2952
Step 275 : - V5276
Step 277 :  V5699
```

```
Step 279 :  V748
Step 281 : - V14805
Step 282 :  V6410
Step 284 : - V606
Step 285 : - V4529
Step 287 :  V15038
Step 289 :  V17354
Step 290 : - V6135
Step 292 :  V10368
Step 293 : - V17354
Step 294 :  V10734
Step 296 :  V16403
Step 297 :  V10170
Step 298 :  V12860
Step 300 : - V11287
Step 301 : - V458
Step 302 : - V12241
Step 304 :  V7331
Step 306 :  V13795
Step 308 : - V8873
Step 310 : - V12860
Step 311 :  V7180
Step 313 :  V13979
Step 315 :  V8777
Step 316 : - V12625
Step 318 :  V7188
Step 320 :  V5927
Step 321 : - V5350
Step 323 :  V1120
Step 324 : - V3117
Step 326 :  V2247
Step 327 : - V7331
Step 328 : - V8841
Step 329 : - V994
Step 330 : - V17045
Step 331 : - V7206
Step 332 : - V5699
Step 334 : - V17878
Step 336 :  V12537
Step 338 : - V16403
Step 340 : - V10508
```

```
Step 341 : - V2952
Step 343 :  V3101
Step 344 : - V7790
Step 345 : - V3740
Step 346 :  V13388
Step 347 :  V4397
Step 348 : - V17357
Step 349 :  V17148
Step 350 :  V11287
Step 352 :  V15989
Step 354 :  V8576
Step 356 :  V2920
Step 357 :  V14189
Step 359 :  V8144
Step 360 :  V2196 - V8672
Step 361 :  V4270
Step 362 : - V4291
Step 363 : - V14189
Step 365 :  V3902
Step 367 :  V366
Step 368 :  V15273
Step 369 :  V12860
Step 370 : - V4477
Step 372 :  V14189
Step 374 :  V12857
Step 375 : - V935
Step 376 :  V16034
Step 378 :  V14733
Step 379 : - V8144
Step 380 :  V12074
Step 381 :  V15990
Step 382 : - V10734
Step 384 :  V14349
Step 386 :  V8052
Step 388 :  V15890
Step 390 :  V16078
Step 392 :  V10924
Step 394 :  V1116
Step 396 :  V11767
Step 397 : - V4270
Step 398 : - V12063
```

```
Step 400 :  V11802
Step 402 :  V228
Step 404 :  V5305
Step 405 : - V228
Step 406 :  V16403
Step 407 : - V9832
Step 409 :  V228
Step 411 :  V10468
Step 412 : - V3101
Step 414 :  V8013
Step 415 : - V6410
Step 418 :  V9523
Step 420 :  V5696
Step 421 : - V9796
Step 422 : - V11767
Step 423 : - V12857
Step 425 :  V7739
Step 426 : - V5540
Step 427 : - V2247
Step 429 :  V16415
Step 430 : - V13979
Step 432 :  V5407
Step 433 :  V10444
Step 434 :  V2467
Step 436 :  V2343
Step 437 : - V16415
Step 438 : - V366
Step 439 : - V10779
Step 440 : - V11802
Step 441 : - V16362
Step 442 :  V16380
Step 443 : - V12272
Step 445 :  V2462
Step 447 :  V8672
Step 448 : - V1116
Step 450 : - V7739
Step 452 :  V7804
Step 454 :  V12237
Step 456 :  V7739
Step 458 :  V3101
Step 459 : - V10444
```

```
Step 460 : - V12255
Step 462 :  V13948
Step 463 : - V7427
Step 464 : - V12197
Step 466 :  V1587
Step 467 :  V16415
Step 468 : - V7739
Step 470 :  V9010
Step 472 :  V11798
Step 475 :  V3091
Step 476 : - V13027
Step 477 : - V17266
Step 479 :  V5544
Step 480 :  V9534
Step 481 :  V12570
Step 482 : - V3101
Step 484 :  V9488
Step 486 :  V366
Step 487 : - V9523
Step 488 : - V9534
Step 489 : - V3892
Step 490 : - V12537
Step 491 : - V228
Step 492 :  V16162
Step 495 :  V7739
Step 496 : - V10182
Step 500 :  V10734
Step 502 :  V10513
Step 503 : - V11798
Step 505 :  V13090
Step 507 :  V5108
Step 510 :  V17499
Step 512 :  V7340
Step 513 : - V8052
Step 514 : - V16415
Step 515 : - V10368
Step 517 :  V15630
Step 518 : - V386
Step 520 :  V11885
Step 521 : - V2920
Step 522 : - V16078
```

```
Step 524 :  V4594
Step 525 :  V384
Step 526 : - V5108
Step 528 :  V10368
Step 530 :  V16038
Step 532 :  V15851
Step 534 :  V12255
Step 536 :  V8820
Step 539 :  V5073
Step 542 :  V16163
Step 543 : - V16034
Step 545 : - V16163
Step 546 : - V15851
Step 549 :  V15215
Step 550 : - V16403
Step 552 :  V10966
Step 554 :  V15810
Step 555 : - V11287
Step 557 :  V16163
Step 558 : - V15478
Step 560 :  V15851
Step 563 :  V7646
Step 564 : - V11885
Step 565 : - V3902
Step 568 :  V10826
Step 569 : - V17499
Step 571 : - V15810
Step 572 : - V366
Step 573 : - V2025
Step 575 :  V3679
Step 576 : - V7863
Step 577 :  V17481
Step 579 :  V7556
Step 580 : - V2467
Step 581 : - V17158 - V5073
Step 582 :  V386
Step 583 : - V14168
Step 584 : - V17481
Step 585 : - V12570
Step 587 :  V12367
Step 589 :  V17045
```

```
Step 594 :  V2467
Step 596 :  V9447
Step 597 : - V12074
Step 599 :  V177
Step 600 : - V10826
Step 601 : - V1807
Step 603 :  V3178
Step 605 :  V17878
Step 607 :  V7123
Step 609 :  V10826
Step 610 :  V16403
Step 611 : - V3679
Step 612 : - V7739
Step 615 :  V6689
Step 616 :  V15478
Step 618 :  V17499
Step 620 : - V12099
Step 621 : - V10966
Step 622 :  V10926  V12074
Step 623 : - V15630
Step 624 : - V10926
Step 625 :  V2025
Step 626 :  V366
Step 629 :  V13027
Step 630 :  V14223
Step 632 :  V11141
Step 633 : - V17499
Step 634 :  V5044
Step 636 :  V6681
Step 637 :  V11885
Step 638 :  V17481
Step 639 :  V7739
Step 640 :  V4756
Step 644 :  V8034  V17499
Step 645 :  V14168
Step 646 :  V12849
Step 647 :  V1477
Step 650 :  V16470
Step 651 :  V6410
Step 652 :  V15810
Step 653 : - V2462
```

```
Step 654 :  V2410
Step 656 :  V9796
Step 658 :  V10003
Step 659 :  V10197
Step 661 :  V15081
Step 663 :  V12154
Step 665 : - V8820
Step 666 :  V1807
Step 668 :  V17170
Step 669 :  V4952
Step 670 :  V12241
Step 672 :  V420
Step 673 :  V6166  V12099  V16482
Step 674 :  V9178
Step 675 :  V13536
Step 676 :  V12751
Step 678 :  V6027
Step 680 :  V9114
Step 682 :  V12570
Step 683 :  V4291 - V17499
Step 685 :  V7427 - V17052
Step 686 :  V2462
> plot(coxpathA)
> dev.copy2eps(file="coxpathA.eps")
X11cairo
       2
> plot(coxpathA,xvar="lambda")
> dev.copy2eps(file="coxpathAlambda.eps")
X11cairo
       2
> plot(coxpathA,xvar="step")
> dev.copy2eps(file="coxpathAstep.eps")
X11cairo
       2
> plot(coxpathA,type="aic")
> dev.copy2eps(file="coxpathAaic.eps")
X11cairo
       2
> plot(coxpathA,type="bic")
> dev.copy2eps(file="coxpathAbic.eps")
X11cairo
```

```
       2
> cvcoxpathA<-cv.coxpath(covlistA,method="efron",nfold=10,fraction=seq(from=0,to=1,
CV Fold 1
CV Fold 2
CV Fold 3
CV Fold 4
CV Fold 5
CV Fold 6
CV Fold 7
CV Fold 8
CV Fold 9
CV Fold 10
> dev.copy2eps(file="cvcoxpathA.eps")
X11cairo
       2
> summary(coxpathA)
Call:
coxpath(data = covlistA, add.newvars = 1, bshoot.threshold = 0.1,
    relax.lambda = 1e-07, standardize = TRUE)
          Df   Log.p.lik      AIC      BIC
Step 1     0 -254.448193 508.8964 508.8964
Step 2     1 -250.717223 503.4344 506.7227
Step 4     2 -249.482701 502.9654 509.5419
Step 5     3 -247.497988 500.9960 510.8608
Step 6     4 -245.126715 498.2534 511.4065
Step 8     5 -241.072607 492.1452 508.5865
Step 9     6 -239.634917 491.2698 510.9994
Step 10    7 -239.469362 492.9387 515.9566
Step 12    8 -236.608358 489.2167 515.5229
Step 14    9 -234.892344 487.7847 517.3791
Step 16   10 -234.044843 488.0897 520.9724
Step 17   11 -233.203209 488.4064 524.5774
Step 19   12 -231.293917 486.5878 526.0470
Step 21   13 -229.294344 484.5887 527.3362
Step 23   14 -228.120645 484.2413 530.2770
Step 25   15 -227.242381 484.4848 533.8088
Step 27   16 -227.132769 486.2655 538.8778
Step 28   17 -223.696850 481.3937 537.2942
Step 30   18 -223.375964 482.7519 541.9407
Step 31   19 -222.597297 483.1946 545.6717
Step 33   20 -221.777187 483.5544 549.3197
```

```
Step 34   20 -221.254876 482.5098 548.2751
Step 35   20 -220.941048 481.8821 547.6474
Step 36   21 -220.696345 483.3927 552.4463
Step 38   22 -219.981537 483.9631 556.3049
Step 39   23 -219.016181 484.0324 559.6625
Step 41   24 -218.454593 484.9092 563.8276
Step 43   25 -218.070102 486.1402 568.3469
Step 45   26 -216.941877 485.8838 571.3787
Step 47   27 -216.011485 486.0230 574.8062
Step 49   28 -215.909628 487.8193 579.8907
Step 50   29 -215.872711 489.7454 585.1052
Step 52   29 -214.042608 486.0852 581.4450
Step 53   29 -213.955862 485.9117 581.2715
Step 55   30 -213.695681 487.3914 586.0394
Step 57   31 -211.861522 485.7230 587.6593
Step 58   32 -209.724947 483.4499 588.6744
Step 59   33 -209.693715 485.3874 593.9002
Step 61   34 -209.541152 487.0823 598.8834
Step 62   35 -209.355271 488.7105 603.7999
Step 63   36 -208.764105 489.5282 607.9058
Step 65   37 -207.832977 489.6660 611.3318
Step 67   37 -207.609179 489.2184 610.8842
Step 68   37 -207.584584 489.1692 610.8350
Step 70   38 -207.172413 490.3448 615.2990
Step 72   39 -205.864787 489.7296 617.9720
Step 74   40 -205.481943 490.9639 622.4946
Step 75   41 -205.481390 492.9628 627.7817
Step 77   42 -205.129824 494.2596 632.3669
Step 78   43 -205.124714 496.2494 637.6449
Step 79   44 -204.137229 496.2745 640.9582
Step 81   45 -203.198230 496.3965 644.3685
Step 83   46 -202.766675 497.5334 648.7936
Step 85   47 -199.382098 492.7642 647.3127
Step 86   48 -199.076092 494.1522 651.9890
Step 88   49 -197.014587 492.0292 653.1543
Step 89   50 -196.927946 493.8559 658.2692
Step 90   50 -196.837247 493.6745 658.0878
Step 92   50 -195.942243 491.8845 656.2978
Step 94   51 -195.253250 492.5065 660.2081
Step 96   52 -195.132022 494.2640 665.2539
Step 97   53 -194.977427 495.9549 670.2330
```

```
Step 99   54 -193.657321 495.3146 672.8811
Step 100  54 -193.080050 494.1601 671.7265
Step 101  54 -191.909176 491.8184 669.3848
Step 103  55 -190.884647 491.7693 672.6240
Step 104  56 -190.705015 493.4100 677.5530
Step 106  57 -190.102789 494.2056 681.6368
Step 107  57 -189.793430 493.5869 681.0181
Step 109  57 -187.905271 489.8105 677.2418
Step 110  57 -185.952698 485.9054 673.3366
Step 112  57 -185.077866 484.1557 671.5870
Step 113  57 -184.305466 482.6109 670.0422
Step 114  57 -184.124645 482.2493 669.6805
Step 115  57 -183.491773 480.9835 668.4148
Step 116  57 -182.445413 478.8908 666.3220
Step 117  57 -181.945906 477.8918 665.3230
Step 119  57 -181.818062 477.6361 665.0673
Step 121  58 -180.710209 477.4204 668.1399
Step 123  59 -180.333887 478.6678 672.6755
Step 125  60 -179.322330 478.6447 675.9407
Step 127  61 -178.724293 479.4486 680.0329
Step 128  62 -178.465519 480.9310 684.8036
Step 130  63 -176.084509 478.1690 685.3298
Step 131  63 -175.855320 477.7106 684.8715
Step 133  63 -175.589029 477.1781 684.3389
Step 135  64 -175.279076 478.5582 689.0072
Step 136  65 -174.899189 479.7984 693.5357
Step 138  66 -173.049846 478.0997 695.1253
Step 140  66 -171.736988 475.4740 692.4996
Step 142  65 -171.052070 472.1041 685.8415
Step 144  65 -170.082164 470.1643 683.9017
Step 146  66 -169.713093 471.4262 688.4518
Step 147  67 -169.578737 473.1575 693.4714
Step 149  68 -168.440026 472.8801 696.4822
Step 150  68 -168.123632 472.2473 695.8494
Step 151  67 -167.819385 469.6388 689.9527
Step 153  67 -167.341353 468.6827 688.9966
Step 154  68 -167.012880 470.0258 693.6279
Step 155  69 -166.967401 471.9348 698.8252
Step 157  69 -164.999495 467.9990 694.8894
Step 158  69 -164.352537 466.7051 693.5955
Step 159  69 -164.231304 466.4626 693.3530
```

```
Step 160  69 -163.088572 464.1771 691.0676
Step 161  70 -162.278893 464.5578 694.7365
Step 163  71 -161.979939 465.9599 699.4268
Step 165  72 -161.356547 466.7131 703.4683
Step 166  72 -159.249506 462.4990 699.2542
Step 167  72 -158.230613 460.4612 697.2165
Step 169  73 -157.659854 461.3197 701.3632
Step 170  74 -156.604393 461.2088 704.5405
Step 172  75 -156.359693 462.7194 709.3394
Step 173  75 -155.765251 461.5305 708.1505
Step 174  75 -155.579684 461.1594 707.7794
Step 176  76 -155.156817 462.3136 712.2219
Step 177  77 -155.045072 464.0901 717.2867
Step 178  77 -154.639405 463.2788 716.4754
Step 180  77 -154.420134 462.8403 716.0368
Step 181  78 -154.329418 464.6588 721.1437
Step 183  79 -154.063470 466.1269 725.9000
Step 185  80 -153.809526 467.6191 730.6804
Step 187  81 -153.626770 469.2535 735.6032
Step 189  82 -152.852649 469.7053 739.3432
Step 190  83 -152.815038 471.6301 744.5562
Step 192  84 -151.077313 470.1546 746.3691
Step 194  84 -150.451855 468.9037 745.1181
Step 195  84 -150.201885 468.4038 744.6182
Step 197  85 -150.019879 470.0398 749.5425
Step 199  86 -149.489257 470.9785 753.7695
Step 200  87 -149.118422 472.2368 758.3161
Step 202  88 -149.010705 474.0214 763.3889
Step 203  88 -148.580162 473.1603 762.5278
Step 204  88 -148.534273 473.0685 762.4360
Step 205  89 -148.522276 475.0446 767.7003
Step 207  90 -148.367357 476.7347 772.6787
Step 208  91 -148.332207 478.6644 777.8967
Step 209  91 -148.296984 478.5940 777.8263
Step 211  91 -145.842782 473.6856 772.9179
Step 212  91 -145.636333 473.2727 772.5050
Step 213  90 -145.403325 470.8067 766.7507
Step 214  89 -145.298434 468.5969 761.2526
Step 216  89 -145.200320 468.4006 761.0564
Step 217  89 -144.990399 467.9808 760.6366
Step 218  88 -144.139119 464.2782 753.6457
```

Step	220	88	-143.729623	463.4592	752.8267
Step	221	88	-143.149254	462.2985	751.6660
Step	222	88	-143.146080	462.2922	751.6597
Step	224	89	-141.738416	461.4768	754.1326
Step	226	90	-141.617032	463.2341	759.1781
Step	228	91	-140.812665	463.6253	762.8576
Step	230	92	-140.576452	465.1529	767.6735
Step	231	93	-140.473827	466.9477	772.7565
Step	233	94	-139.879472	467.7589	776.8560
Step	235	95	-139.626706	469.2534	781.6388
Step	237	96	-139.449747	470.8995	786.5731
Step	239	97	-139.340949	472.6819	791.6438
Step	240	97	-138.869328	471.7387	790.7006
Step	242	96	-138.531174	469.0623	784.7360
Step	243	96	-137.874236	467.7485	783.4221
Step	244	97	-137.261075	468.5222	787.4841
Step	245	98	-137.171397	470.3428	792.5930
Step	247	99	-137.021572	472.0431	797.5816
Step	249	100	-136.365141	472.7303	801.5570
Step	251	100	-136.005111	472.0102	800.8369
Step	252	100	-135.567430	471.1349	799.9616
Step	253	101	-135.456485	472.9130	805.0279
Step	255	102	-135.192346	474.3847	809.7879
Step	256	103	-135.134755	476.2695	814.9610
Step	257	104	-135.095060	478.1901	820.1699
Step	259	104	-134.620716	477.2414	819.2212
Step	260	103	-134.513212	475.0264	813.7179
Step	261	103	-133.935398	473.8708	812.5623
Step	262	104	-133.858088	475.7162	817.6959
Step	264	104	-133.662060	475.3241	817.3039
Step	266	104	-132.918046	473.8361	815.8159
Step	267	104	-132.672302	473.3446	815.3244
Step	269	104	-131.685492	471.3710	813.3508
Step	271	104	-131.483858	470.9677	812.9475
Step	272	103	-130.922529	467.8451	806.5366
Step	274	103	-129.412199	464.8244	803.5159
Step	275	103	-128.443908	462.8878	801.5793
Step	277	103	-127.843361	461.6867	800.3782
Step	279	104	-126.692134	461.3843	803.3640
Step	281	104	-126.254305	460.5086	802.4884
Step	282	104	-126.179621	460.3592	802.3390

```
Step 284 104 -125.218328 458.4367 800.4164
Step 285 103 -123.889391 453.7788 792.4703
Step 287 103 -122.948316 451.8966 790.5881
Step 289 104 -122.369155 452.7383 794.7181
Step 290 104 -121.726571 451.4531 793.4329
Step 292 104 -121.235082 450.4702 792.4499
Step 293 104 -120.364316 448.7286 790.7084
Step 294 104 -120.082551 448.1651 790.1449
Step 296 105 -119.109262 448.2185 793.4866
Step 297 106 -119.074299 450.1486 798.7049
Step 298 107 -118.159503 450.3190 802.1636
Step 300 107 -117.992405 449.9848 801.8294
Step 301 106 -117.561257 447.1225 795.6788
Step 302 105 -117.017922 444.0358 789.3039
Step 304 105 -116.020586 442.0412 787.3092
Step 306 106 -115.249847 442.4997 791.0560
Step 308 106 -114.085722 440.1714 788.7277
Step 310 105 -113.247902 436.4958 781.7638
Step 311 105 -113.247159 436.4943 781.7624
Step 313 106 -112.688468 437.3769 785.9332
Step 315 107 -112.370286 438.7406 790.5851
Step 316 107 -110.784295 435.5686 787.4132
Step 318 107 -108.874754 431.7495 783.5941
Step 320 108 -107.665811 431.3316 786.4645
Step 321 108 -105.972641 427.9453 783.0781
Step 323 108 -104.687733 425.3755 780.5083
Step 324 108 -103.882475 423.7650 778.8978
Step 326 108 -102.437892 420.8758 776.0086
Step 327 108 -102.232722 420.4654 775.5983
Step 328 107 -102.082847 418.1657 770.0103
Step 329 106 -100.848631 413.6973 762.2536
Step 330 105  -99.689712 409.3794 754.6475
Step 331 104  -99.352387 406.7048 748.6845
Step 332 103  -98.863175 403.7264 742.4179
Step 334 102  -96.257329 396.5147 731.9179
Step 336 102  -95.259765 394.5195 729.9228
Step 338 102  -93.974855 391.9497 727.3529
Step 340 101  -93.560046 389.1201 721.2351
Step 341 100  -92.814228 385.6285 714.4552
Step 343 100  -90.991459 381.9829 710.8096
Step 344 100  -89.547235 379.0945 707.9212
```

Step	345	99	-89.338308	376.6766	702.2151
Step	346	99	-88.244809	374.4896	700.0281
Step	347	100	-88.170001	376.3400	705.1667
Step	348	100	-87.026848	374.0537	702.8804
Step	349	100	-87.018884	374.0378	702.8645
Step	350	101	-86.794821	375.5896	707.7046
Step	352	102	-84.693945	373.3879	708.7911
Step	354	103	-83.624405	373.2488	711.9403
Step	356	104	-83.554463	375.1089	717.0887
Step	357	105	-83.505835	377.0117	722.2797
Step	359	106	-83.345504	378.6910	727.2473
Step	360	106	-79.939178	371.8784	720.4347
Step	361	107	-79.872805	373.7456	725.5902
Step	362	107	-79.644862	373.2897	725.1343
Step	363	106	-77.552906	367.1058	715.6621
Step	365	106	-77.161027	366.3221	714.8784
Step	367	107	-75.337116	364.6742	716.5188
Step	368	108	-74.727357	365.4547	720.5876
Step	369	109	-74.652929	367.3059	725.7270
Step	370	109	-73.752295	365.5046	723.9257
Step	372	109	-73.215491	364.4310	722.8521
Step	374	110	-71.732011	363.4640	725.1734
Step	375	110	-71.220456	362.4409	724.1503
Step	376	110	-71.015697	362.0314	723.7408
Step	378	111	-70.185608	362.3712	727.3689
Step	379	111	-69.173044	360.3461	725.3437
Step	380	111	-69.171821	360.3436	725.3413
Step	381	112	-69.126979	362.2540	730.5399
Step	382	112	-65.553667	355.1073	723.3932
Step	384	112	-64.440430	352.8809	721.1668
Step	386	113	-64.189759	354.3795	725.9537
Step	388	114	-64.066388	356.1328	730.9952
Step	390	115	-63.766290	357.5326	735.6833
Step	392	116	-63.584864	359.1697	740.6087
Step	394	117	-62.899179	359.7984	744.5256
Step	396	118	-62.618687	361.2374	749.2529
Step	397	118	-62.032207	360.0644	748.0799
Step	398	117	-61.802889	357.6058	742.3330
Step	400	117	-61.581437	357.1629	741.8901
Step	402	118	-60.716422	357.4328	745.4484
Step	404	119	-59.347907	356.6958	747.9996

Step	405	119	-56.475124	350.9502	742.2540
Step	406	119	-56.451713	350.9034	742.2072
Step	407	119	-56.314961	350.6299	741.9337
Step	409	119	-56.120264	350.2405	741.5443
Step	411	120	-53.983243	347.9665	742.5585
Step	412	120	-53.723603	347.4472	742.0392
Step	414	120	-53.474375	346.9488	741.5408
Step	415	120	-53.469293	346.9386	741.5306
Step	418	120	-51.134219	342.2684	736.8605
Step	420	121	-50.297106	342.5942	740.4745
Step	421	121	-49.893370	341.7867	739.6671
Step	422	120	-49.187635	338.3753	732.9673
Step	423	119	-47.737679	333.4754	724.7791
Step	425	119	-47.508359	333.0167	724.3205
Step	426	119	-46.832081	331.6642	722.9679
Step	427	118	-46.803934	329.6079	717.6234
Step	429	118	-46.213725	328.4274	716.4430
Step	430	118	-45.988961	327.9779	715.9934
Step	432	118	-45.165435	326.3309	714.3464
Step	433	119	-44.932194	327.8644	719.1682
Step	434	120	-44.843353	329.6867	724.2788
Step	436	121	-44.505840	331.0117	728.8920
Step	437	121	-43.229610	328.4592	726.3395
Step	438	120	-43.157976	326.3160	720.9080
Step	439	119	-41.495231	320.9905	712.2942
Step	440	118	-40.315526	316.6311	704.6466
Step	441	117	-40.202583	314.4052	699.1324
Step	442	117	-40.180319	314.3606	699.0879
Step	443	117	-40.171854	314.3437	699.0710
Step	445	117	-40.003701	314.0074	698.7346
Step	447	118	-39.829340	315.6587	703.6742
Step	448	118	-39.753722	315.5074	703.5230
Step	450	117	-38.776818	311.5536	696.2809
Step	452	117	-38.753246	311.5065	696.2337
Step	454	118	-38.641288	313.2826	701.2981
Step	456	119	-38.460227	314.9205	706.2242
Step	458	120	-38.203765	316.4075	710.9996
Step	459	120	-37.761980	315.5240	710.1160
Step	460	119	-37.660664	313.3213	704.6251
Step	462	119	-37.468298	312.9366	704.2404
Step	463	119	-37.292063	312.5841	703.8879

Step	464	118	-37.076484	310.1530	698.1685
Step	466	118	-36.735201	309.4704	697.4859
Step	467	119	-36.708085	311.4162	702.7199
Step	468	119	-36.582404	311.1648	702.4686
Step	470	119	-36.379087	310.7582	702.0620
Step	472	120	-36.036711	312.0734	706.6655
Step	475	121	-35.988670	313.9773	711.8577
Step	476	121	-34.857507	311.7150	709.5953
Step	477	120	-34.047838	308.0957	702.6877
Step	479	120	-33.796280	307.5926	702.1846
Step	480	121	-33.729959	309.4599	707.3402
Step	481	122	-33.301461	310.6029	711.7715
Step	482	122	-33.206963	310.4139	711.5825
Step	484	122	-32.941929	309.8839	711.0524
Step	486	123	-32.503618	311.0072	715.4641
Step	487	123	-32.419327	310.8387	715.2955
Step	488	122	-31.473845	306.9477	708.1163
Step	489	121	-31.224751	304.4495	702.3298
Step	490	120	-30.893891	301.7878	696.3798
Step	491	119	-30.452976	298.9060	690.2097
Step	492	119	-29.926592	297.8532	689.1570
Step	495	120	-28.369509	296.7390	691.3311
Step	496	120	-27.972166	295.9443	690.5364
Step	500	120	-27.338701	294.6774	689.2694
Step	502	121	-27.197754	296.3955	694.2758
Step	503	121	-26.468273	294.9365	692.8169
Step	505	121	-25.863861	293.7277	691.6080
Step	507	122	-24.947798	293.8956	695.0642
Step	510	123	-23.650550	293.3011	697.7579
Step	512	124	-23.032329	294.0647	701.8098
Step	513	124	-22.950783	293.9016	701.6467
Step	514	123	-21.909014	289.8180	694.2749
Step	515	122	-20.928287	285.8566	687.0252
Step	517	122	-20.634086	285.2682	686.4367
Step	518	122	-20.308813	284.6176	685.7862
Step	520	122	-19.844326	283.6887	684.8572
Step	521	122	-18.568587	281.1372	682.3058
Step	522	121	-18.104213	278.2084	676.0887
Step	524	121	-17.987138	277.9743	675.8546
Step	525	122	-17.230202	278.4604	679.6290
Step	526	122	-17.111880	278.2238	679.3923

```
Step 528 122  -16.877416 277.7548 678.9234
Step 530 123  -15.983780 277.9676 682.4244
Step 532 124  -15.793383 279.5868 687.3319
Step 534 125  -15.490919 280.9818 692.0152
Step 536 126  -15.134551 282.2691 696.5907
Step 539 127  -13.540869 281.0817 698.6917
Step 542 128  -12.111287 280.2226 701.1208
Step 543 128  -12.097915 280.1958 701.0940
Step 545 127  -12.068249 278.1365 695.7464
Step 546 126  -12.051797 276.1036 690.4252
Step 549 126  -11.369972 274.7399 689.0616
Step 550 126  -10.717556 273.4351 687.7568
Step 552 126  -10.377840 272.7557 687.0773
Step 554 127  -10.339011 274.6780 692.2879
Step 555 127   -9.381208 272.7624 690.3723
Step 557 127   -9.266781 272.5336 690.1435
Step 558 127   -9.072017 272.1440 689.7539
Step 560 127   -8.957120 271.9142 689.5242
Step 563 128   -7.982128 271.9643 692.8624
Step 564 128   -7.676999 271.3540 692.2522
Step 565 127   -7.448814 268.8976 686.5075
Step 568 127   -5.993256 265.9865 683.5964
Step 569 127   -5.987319 265.9746 683.5846
Step 571 126   -5.959733 263.9195 678.2411
Step 572 125   -5.641552 261.2831 672.3165
Step 573 124   -5.353708 258.7074 666.4525
Step 575 124   -5.334800 258.6696 666.4147
Step 576 124   -5.271192 258.5424 666.2875
Step 577 124   -4.804225 257.6085 665.3536
Step 579 125   -4.728970 259.4579 670.4913
Step 580 125   -3.908397 257.8168 668.8502
Step 581 123   -3.874937 253.7499 658.2067
Step 582 123   -3.874509 253.7490 658.2059
Step 583 123   -3.847895 253.6958 658.1526
Step 584 122   -3.356722 250.7134 651.8820
Step 585 121   -3.190829 248.3817 646.2620
Step 587 121   -3.190318 248.3806 646.2609
Step 589 122   -3.147190 250.2944 651.4630
Step 594 123   -2.983240 251.9665 656.4233
Step 596 124   -2.860204 253.7204 661.4655
Step 597 124   -2.798807 253.5976 661.3427
```

Step	599	124	-2.762512	253.5250	661.2701
Step	600	124	-2.753771	253.5075	661.2527
Step	601	123	-2.723656	251.4473	655.9042
Step	603	123	-2.659747	251.3195	655.7763
Step	605	124	-2.585477	253.1710	660.9161
Step	607	125	-2.513409	255.0268	666.0602
Step	609	126	-2.453756	256.9075	671.2292
Step	610	127	-2.413017	258.8260	676.4359
Step	611	127	-2.360081	258.7202	676.3301
Step	612	126	-2.121389	256.2428	670.5644
Step	615	126	-1.920599	255.8412	670.1628
Step	616	127	-1.914015	257.8280	675.4379
Step	618	128	-1.859283	259.7186	680.6167
Step	620	128	-1.833479	259.6670	680.5651
Step	621	127	-1.763713	257.5274	675.1373
Step	622	127	-1.721117	257.4422	675.0521
Step	623	128	-1.720624	259.4412	680.3394
Step	624	127	-1.720203	257.4404	675.0503
Step	625	127	-1.711628	257.4233	675.0332
Step	626	128	-1.672331	259.3447	680.2428
Step	629	129	-1.668619	261.3372	685.5237
Step	630	130	-1.663360	263.3267	690.8014
Step	632	131	-1.600326	265.2007	695.9636
Step	633	131	-1.599639	265.1993	695.9623
Step	634	131	-1.596567	265.1931	695.9561
Step	636	132	-1.591352	267.1827	701.2340
Step	637	133	-1.586623	269.1732	706.5128
Step	638	134	-1.585860	271.1717	711.7995
Step	639	135	-1.573914	273.1478	717.0639
Step	640	136	-1.569030	275.1381	722.3424
Step	644	137	-1.561558	277.1231	727.6157
Step	645	139	-1.558765	281.1175	738.1866
Step	646	140	-1.547555	283.0951	743.4525
Step	647	141	-1.545983	285.0920	748.7376
Step	650	142	-1.535735	287.0715	754.0054
Step	651	143	-1.534468	289.0689	759.2911
Step	652	144	-1.532381	291.0648	764.5752
Step	653	144	-1.528031	291.0561	764.5665
Step	654	144	-1.524192	291.0484	764.5588
Step	656	145	-1.519817	293.0396	769.8384
Step	658	146	-1.516300	295.0326	775.1196

```
Step 659 147   -1.511062 297.0221 780.3974
Step 661 148   -1.508650 299.0173 785.6808
Step 663 149   -1.507430 301.0149 790.9666
Step 665 149   -1.497398 300.9948 790.9466
Step 666 149   -1.495221 300.9904 790.9422
Step 668 150   -1.493261 302.9865 796.2266
Step 669 151   -1.492042 304.9841 801.5124
Step 670 152   -1.491719 306.9834 806.8000
Step 672 153   -1.491160 308.9823 812.0872
Step 673 154   -1.487919 310.9758 817.3690
Step 674 157   -1.486969 316.9739 833.2319
Step 675 158   -1.483970 318.9679 838.5141
Step 676 159   -1.482030 320.9641 843.7985
Step 678 160   -1.480506 322.9610 849.0837
Step 680 161   -1.478584 324.9572 854.3682
Step 682 162   -1.470547 326.9411 859.6404
Step 683 162   -1.470448 326.9409 859.6402
Step 685 162   -1.469300 326.9386 859.6379
Step 686 163   -1.467662 328.9353 864.9229
Step 687 164   -1.467403 330.9348 870.2106
> > getwd()
[1] "m:/"
> setwd("m:/")
> library(survival)
Loading required package: splines
> library(base)
> dataa<-read.table("adata.txt",header=T)
> datab<-read.table("bdata.txt",header=T)
> ytimea<-as.vector(dataa[,3])
> ytimeb<-as.vector(datab[,3])
> ystatusb<-as.vector(datab[,4])
> ystatusa<-as.vector(dataa[,4])
> fittreata<-coxph(Surv(ytimea,ystatusa)~factor(Treat),data=dataa)
> fittreata
Call:
coxph(formula = Surv(ytimea, ystatusa) ~ factor(Treat), data = dataa)


               coef exp(coef) se(coef)    z       p
factor(Treat)1 1.67      5.31    0.353 4.73 2.2e-06
```

```
Likelihood ratio test=29.2  on 1 df, p=6.66e-08  n= 198
> fittreatb<-coxph(Surv(ytimeb,ystatusb)~factor(Treat),data=datab)
> fittreatb
Call:
coxph(formula = Surv(ytimeb, ystatusb) ~ factor(Treat), data = datab)


                 coef exp(coef) se(coef)    z    p
factor(Treat)1 0.293      1.34    0.186 1.58 0.11

Likelihood ratio test=2.49  on 1 df, p=0.114  n= 198
> survtreatb<-survfit(Surv(ytimeb,ystatusb)~factor(Treat),data=datab)
> plot(survtreatb)
> dev.copy2eps(file="survb.eps")
windows
2
> survtreata<-survfit(Surv(ytimea,ystatusa)~factor(Treat),data=dataa)
> plot(survtreata)
> dev.copy2eps(file="surva.eps")
windows
2
> summary(fittreatb)
Call:
coxph(formula = Surv(ytimeb, ystatusb) ~ factor(Treat), data = datab)

  n= 198

                 coef exp(coef) se(coef)     z Pr(>|z|)
factor(Treat)1 0.2928    1.3402   0.1857 1.577    0.115

               exp(coef) exp(-coef) lower .95 upper .95
factor(Treat)1     1.340     0.7461    0.9313     1.929

Rsquare= 0.013   (max possible= 0.997 )
Likelihood ratio test= 2.49  on 1 df,   p=0.1144
Wald test            = 2.49  on 1 df,   p=0.1148
Score (logrank) test = 2.5  on 1 df,   p=0.1136

> summary(fittreata)
Call:
coxph(formula = Surv(ytimea, ystatusa) ~ factor(Treat), data = dataa)
```

```
  n= 198

                 coef exp(coef) se(coef)     z Pr(>|z|)
factor(Treat)1 1.6690    5.3066   0.3527 4.732 2.23e-06 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

               exp(coef) exp(-coef) lower .95 upper .95
factor(Treat)1     5.307     0.1884     2.658     10.59

Rsquare= 0.137   (max possible= 0.923 )
Likelihood ratio test= 29.16  on 1 df,   p=6.66e-08
Wald test            = 22.39  on 1 df,   p=2.228e-06
Score (logrank) test = 27.98  on 1 df,   p=1.226e-07
```