Published in final edited form as:

Stat Med. 2012 June 15; 31(13): 1342–1360. doi:10.1002/sim.4448.

Supervised Bayesian latent class models for high-dimensional data

Stacia M. Desantis^{a,*†}, E. Andrés Houseman^b, Brent A. Coull^c, Catherine L. Nutt^d, and Rebecca A. Betensky^c

^aDepartment of Medicine, Medical University of South Carolina, USA

^bDepartment of Community Health, Brown University, USA

^cDepartment of Biostatistics, Harvard University, USA

^dDepartment of Pathology, Harvard Medical School, USA

Abstract

High-grade gliomas are the most common primary brain tumors in adults and are typically diagnosed using histopathology. However, these diagnostic categories are highly heterogeneous and do not always correlate well with survival. In an attempt to refine these diagnoses, we make several immunohistochemical measurements of YKL-40, a gene previously shown to be differentially expressed between diagnostic groups. We propose two latent class models for classification and variable selection in the presence of high-dimensional binary data, fit by using Bayesian Markov chain Monte Carlo techniques. Penalization and model selection are incorporated in this setting via prior distributions on the unknown parameters. The methods provide valid parameter estimates under conditions in which standard supervised latent class models do not, and outperform two-stage approaches to variable selection and parameter estimation in a variety of settings. We study the properties of these methods in simulations, and apply these methodologies to the glioma study for which identifiable three-class parameter estimates cannot be obtained without penalization. With penalization, the resulting latent classes correlate well with clinical tumor grade and offer additional information on survival prognosis that is not captured by clinical diagnosis alone. The inclusion of YKL-40 features also increases the precision of survival estimates. Fitting models with and without YKL-40 highlights a subgroup of patients who have glioblastoma (GBM) diagnosis but appear to have better prognosis than the typical GBM patient.

Keywords

penalization; cancer; variable selection; ridge; latent class; glioma; supervised

1. Introduction

The data that motivate the current research pertain to the diagnosis and prognosis of patients with high-grade gliomas, the most common primary brain tumor in adults (<http://www.cbtrus.org>). In modern neuro-oncology, tumor classification is the variable that most affects therapeutic decisions and prognostic estimation; however, there are limitations of the

current glioma classification schemes with respect to prognosis. Another challenge in the diagnosis of gliomas is that the diagnosis of tumors with nonclassic histology is controversial and subject to interobserver variability. To address these issues, one exploratory gene expression study demonstrated that a gene expression-based prediction model could better distinguish anaplastic oligodendrogliomas from glioblastomas in nonclassic cases [1]. In followup studies, immunohistochemical expression of YKL-40 was found to be significantly different between histological lineages of glioma [2,3]. The current study utilizes data on measures of several immunohistochemical features of the YKL-40 gene in four classes of gliomas: anaplastic oligodendroglioma (AO), anaplastic oligoastrocytoma (AOA), anaplastic astrocytoma (AA), and glioblastoma (GBM) (Nutt *et al.*, in preparation). We collected data on 206 patients at Massachusetts General Hospital in Boston, MA (Table I). The first four columns of the table show prevalence by tumor grade. Overall there are 28 AO, 33 AOA, 44 AA, and 101 GBM tumors. The goal of this study is to elucidate refined classes of glioma patients that correlate with patient survival in order to supplement clinical diagnosis.

Over the last decade, there has been a growing need for statistical methods for analyzing a large number of correlated binary or correlated outcomes recorded on a small to moderate number of individuals. A variety of clustering techniques have been developed for this purpose, often in the context of genomic data. In this paper, we develop supervised latent class analytic methods that provide a classification mechanism trained using observed variables from diagnosis together with survival, with the goal of augmenting standard pathological classification. With many model parameters and few subjects, numerical complications in existing latent class model formulations such as local maxima and non-convergence occur. The inclusion of unnecessary variables in a cluster analysis also may obscure the recovery of clusters [4, 5]. Application of constraints, penalization, or variable selection is needed to identify important variables and stabilize parameter estimation. To address the scientific questions posed by the glioma data, with its multitude of measurements but limited sample size, our goal is threefold. Firstly, we want the survival end point to inform latent class structure. Secondly, we want a large number of binary immunohistochemical measurements to inform the class structure. Thirdly, we want to mitigate effects of irrelevant variables. The first goal is met by considering supervision of the latent class model by survival [6], whereas the goals are met by translating and extending this approach to the Bayesian framework where penalization and variable selection can be incorporated using freely available software. This paper contributes to the current literature on clustering and variable selection by presenting latent class models supervised by a survival end point where prior specification invokes regularization. We also consider conditionally dependent models in this setting, which is a novel contribution to the literature. We validate our approaches via a simulation study that demonstrates that our model-based classification outperforms commonly applied two-stage approaches. When applied to the motivating data, we show that the incorporation of immunohistochemical information improves survival prediction over histopathologic tumor diagnosis alone.

In Section 2, we review the use of regularization and variable selection in classification and existing methods for accommodating survival data in the presence of other prognostic variables. In Section 3, we present the joint analysis of latent class and a time-to-event outcome to the moderate and high-dimensional setting by using both Bayesian ridge and spike and slab priors [7]. In Section 4, we discuss computational issues and model selection. In Section 5, we present results of the simulation study. In Section 6, we apply these methods to the glioma study, and in Section 7 we conclude.

2. Existing methods

2.1. Penalization

Unsupervised, constrained latent class models [8, 9] can be applied to high-dimensional data in a Bayesian setting; however, this restrictive method would require a priori knowledge in order to specify reasonable constraints. Penalization and variable selection are more appropriate in the absence of complete subject-matter knowledge. Penalization may or may not directly invoke variable selection; several authors have developed variable selection in a regularization framework via likelihood penalization [10–12]. Others have developed frequentist penalization methods specifically for latent class analysis [13–15]. In general, for

some parameter, β , the penalized log-likelihood may take the form $l_i - \alpha \sum_{c=1}^K \sum_{j=1}^J |\beta_{cj}|^r$, where l_i is the log of the latent class likelihood contribution of the i th subject, α is the penalty parameter, $c = 1, \dots, K$ indexes latent class, and $j = 1, \dots, J$ indexes the observed variable [13], for example, taking $r = 1$ results in lasso estimates of β [16] and taking $r = 2$ results in ridge estimates of β [17]. Maximum likelihood estimates of β for fixed α are typically obtained using the expectation maximization (EM) algorithm [18]. Houseman *et al.* implemented such an approach in a latent class analysis where penalization was necessary to regularize the Hessian matrix [13]. In their study of 93 oligodendroglioma brain tumor patients for whom binary loss of heterozygosity (LOH) was measured at 19 chromosomal markers, LOH profiles were found to be associated with survival in a post hoc (i.e., two-stage) analysis. However, in the presence of a survival end point, a joint modeling approach might be optimal in differentiating patients with respect to survival.

2.2. Supervised latent class models

Larsen [6] jointly modeled an event process in a latent class setting. Without the presence of censoring, one could use a regularization framework and incorporate the event time as a continuous or ordinal variable in the latent class measurement model. But in the case that not every event is observed, i.e., the outcome is right-censored, we cannot treat survival as another observed variable. In addition, we may not want to treat survival as simply another outcome in the latent class formulation, as then it would have minimal impact on supervising the clustering in the setting of high-dimensional predictors.

A cluster structure that not only utilizes information contained in important variables, such as YKL-40, but that also correlates with the survival outcome is of primary interest in order to facilitate survival prognosis for future patients. Several such methods for moderate and high-dimensional analysis exist in the literature. As mentioned, Larsen [6] jointly modeled an event process in a latent class setting; Moustaki and Steele [19] proposed a similar model for discrete time events, and Lin *et al.* [20] considered a latent class model for the joint analysis of an event process and an observed longitudinal variable. For supervised classification of high-dimensional data, one could extend the approaches of Houseman *et al.* [13] and Shedden and Zucker [15] by penalizing the joint model likelihood. However, parameter estimation would be computationally intensive requiring one to fit the model for a grid of possible penalty parameters. This approach also does not have the attractive feature of variable selection as a byproduct of the estimation technique.

2.3. Variable selection in model-based clustering

In recent years, there have been several model-based approaches to simultaneous variable selection in model-based clustering. Tadesse *et al.* [21] presented a Bayesian variable selection method for clustering based on variable selection priors by using a reversible-jump Markov chain Monte Carlo (MCMC) sampler that searches for models comprising different clusters and subsets of continuous variables. Kim *et al.* [22] considered variable selection in

clustering by using Dirichlet process mixture models. Raftery and Dean [23] presented latent class analysis variable selection for discrete data by using a headlong search algorithm and demonstrated that this approach improved the accuracy in choosing the number of latent classes. However, these computationally intensive methods do not immediately lend themselves to the incorporation of a censored survival end point.

Although not model-based, two-stage approaches are most commonly applied in the current data setting. Typically, a pre-processing step selects variables associated with survival, and those variables are subsequently used for clustering patients [24,25]. For example, in their analysis of gene expression profiles, Bair and Tibshirani [25] and Bair *et al.* [26] pre-processed the data based on a Cox score, selecting only the genes most highly associated with survival. In a second stage, they performed unsupervised clustering on those selected genes. This semi-supervised approach allows survival to inform the cluster structure while mitigating the influence of unimportant variables. Overall, the preliminary selection of only the important genes has been shown to improve the predictive accuracy over unsupervised and supervised techniques that use all of the genes. Because similar two-stage methodologies are very commonly applied in the context of cancer genomics [27, 28], we compare our methods in both simulation and practice with one such procedure developed by Bair *et al.* and Tibshirani. In our simulation study, we demonstrate that a drawback of the two-stage methodology is that important variables may be discarded in the pre-processing stage. As a result, it is potentially more desirable to have an integrated variable selection procedure as part of a unified probability model for binary data.

3. Models

3.1. Supervised latent class model

The observed data for subject i consist of $\{y_i, t_i, \delta_i\}$, where $y_i = \{y_{i1}, \dots, y_{iJ}\}$ are the J binary outcomes, t_i is a possibly right-censored survival time, and δ_i is the failure indicator. Censoring is assumed to be uninformative. It is assumed that the population consists of K unobserved latent diagnostic groups or classes, where C_i is the latent class variable, with $P(C_i = c) = \eta_c$. Conditional on unobserved class membership, assume that the J binary outcomes are mutually independent given latent class, i.e.,

$$P(Y_i = y_i | C_i = c) = \prod_{j=1}^J \pi_{cj}^{y_{ij}} (1 - \pi_{cj})^{1-y_{ij}}, y=0, 1.$$

It follows that the joint distribution of (Y_i, C_i) is

$$P(Y_i = y_i, C_i = c) = P(C_i = c) P(Y_i = y_i | C_i = c) = \eta_c \prod_{j=1}^J \pi_{cj}^{y_{ij}} (1 - \pi_{cj})^{1-y_{ij}}.$$

We now incorporate survival into a joint model with latent class. Allowing latent class to be a covariate in a proportional hazards model, let ν be a $K-1$ dimensional parameter vector that quantifies the effect of latent class on the hazard, where $\nu_1 = 0$ for identification. The baseline hazard $\lambda_0(t)$ is left unspecified, and the density of the event time, t_i is

$$P(T_i = t_i, \Delta_i = \delta_i | C_i = c) = \{\lambda_0(t_i) \exp(\nu_c)\}^{\delta_i} \exp\{-\Lambda_0(t_i) \exp(\nu_c)\},$$

Where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is the integrated baseline hazard. The observed likelihood contribution for the i th subject is thus

$$L_i(\theta; t_i, \delta_i, y_i) = \sum_{c=1}^K P(c) P(y_i|c) p(t_i, \delta_i|c) = \sum_{c=1}^K \eta_c \prod_{j=1}^J \pi_{cj}^{y_{ij}} (1 - \pi_{cj})^{1-y_{ij}} \times (\lambda_0(t_i) \exp(v_c))^{\delta_i} \exp(-\Lambda_0(t_i) \exp(v_c)), \quad (1)$$

where $\theta = (\boldsymbol{\pi}, \mathbf{v}, \lambda_0(t))$. It is straightforward to relax the conditional independence assumption implied by (1) by incorporating first-order interactions between categorical variables (see Section 3.4).

In our experience, in a frequentist setting, it is difficult to impose penalization on this model even for moderate-dimensional data as the penalty search is computationally intensive especially as multiple starting values need to be considered. As a remedy to this problem, we propose a Bayesian approach with variable selection to stabilize parameter estimation. The first procedure incorporates a ridge type of penalty, and the second results in nonlinear shrinkage (NLS) of covariate effects with the additional property of variable selection [16]. Both approaches are supervised in that they allow survival to inform the choice of latent classes. However, by incorporating shrinkage on the binary latent variables but not on the hazard ratios (HRs), the survival outcome is treated differently from the other observable outcomes and therefore upweighted when classifying subjects into latent classes. We show via a simulation study that both models can accommodate high-dimensional data, even where the majority of observed variables provide little information about class structure.

For the estimation of the baseline hazard rate, we assume a piecewise constant exponential hazard. This approach has been shown to be robust in that it is approximately nonparametric and is easy to model in a Bayesian setting. In particular, the event times are ordered, and the time axis is split into L intervals, $(0, s_1], (s_1, s_2], \dots, (s_{L-1}, s_L]$, where, for each interval, there is a constant baseline hazard, $\lambda_0(t) = \lambda_l$ for $t \in I_l = (s_{l-1}, s_l]$. The optimal number of intervals depends on the number of subjects and events. A general rule is to allow approximately 10 events in each interval [29]. To properly define the likelihood over the L intervals, let $d_{il} = 1$ if subject i has an event or is censored in interval l , and 0 otherwise. Recalling that $\delta_i = 1$ if subject i has failed, the density of the event time, T_i , is as follows:

$$P(t_i, \delta_i|c_i) = \prod_{l=1}^L \{\lambda_l \exp(v_c)\}^{d_{il}\delta_i} \exp \left\{ -d_{il} \left[\lambda_l (t_i - s_{l-1}) + \sum_{g=1}^{l-1} \lambda_g (s_g - s_{g-1}) \right] \exp(v_c) \right\}. \quad (2)$$

The likelihood contribution for the i th person is then

$$L_i(\theta; t_i, \delta_i, y_i) = \sum_{c=1}^K P(c) P(y_i|c) p(t_i, \delta_i|c) = \sum_{c=1}^K \eta_c \prod_{j=1}^J \pi_{cj}^{y_{ij}} (1 - \pi_{cj})^{1-y_{ij}} \times \prod_{l=1}^L \{\lambda_l \exp(v_c)\}^{d_{il}\delta_i} \exp \left\{ -d_{il} \left[\lambda_l (t_i - s_{l-1}) + \sum_{g=1}^{l-1} \lambda_g (s_g - s_{g-1}) \right] \exp(v_c) \right\}, \quad (3)$$

where $\pi_{1j} = \frac{\exp(\beta_{0j})}{1 + \exp(\beta_{0j})}$ and $\pi_{cj} = \frac{\exp(\beta_{0j} + \sum_{l=1}^{c-1} \beta_{lj})}{1 + \exp(\beta_{0j} + \sum_{l=1}^{c-1} \beta_{lj})}$ for $c = 2, \dots, K$ and $j = 1, \dots, J$ and where $\theta = (\boldsymbol{\beta}, \mathbf{v}, \boldsymbol{\lambda})$ and $\boldsymbol{\beta} = (\beta_{01}, \dots, \beta_{(K-1)J})$.

To obtain unpenalized estimates, we specify noninformative priors for the latent class coefficients from the survival part of the model, i.e., $\mathbf{v}_c \sim N(0, 1000)$. A common prior assumption for the baseline hazard is the independent gamma prior, i.e., $\lambda_l \sim \mathcal{G}(a_{0l}, d_{0l})$,

where c_{0j} and d_{0j} can be elicited through the prior mean and variance of λ_j [30]. In all analyses, we set $c_{0j} = d_{0j} = 0.01$, resulting in a noninformative prior for the baseline hazard. Because the number of subjects within a class can be modeled by a multinomial distribution with parameters $(\eta_1, \dots, \eta_K)$ and N , a conjugate prior density for the vector of class probabilities is a Dirichlet distribution with fixed hyperparameters $a_1, \dots, a_K$. We set the Dirichlet hyperparameters to 1, giving equal probability to each latent class assignment. We

specify $\beta_{0j} \sim N\left(0, \frac{1}{\tau^2}\right)$ and $\beta_{cj} \sim N\left(0, \frac{1}{\tau^2}\right)$ for the intercepts and slopes, respectively, where τ is fixed at 0.0001. This small precision puts virtually no restriction on the posterior distribution of the class-specific parameters, β_{cj} , allowing each to be estimated separately. In the absence of penalization, the model would treat survival on par with the other variables, likely diminishing its impact on the clustering.

3.2. Ridge shrinkage

The Bayesian ridge estimates for β can be derived as the posterior mode under a normal prior. The form of ridge regression shrinkage depends on the correlation of the variables. This penalized approach might be useful to consider when most of the variables are actually associated with latent class.

In our setting, we obtain ridge estimates by incorporating the mean and variance of β_{cj} as

hyperparameters in the model, that is, $\beta_{cj} \sim N\left(\mu_c, \frac{1}{\tau_c^2}\right)$, and we specify noninformative priors on the mean and variance, $\mu_c \sim N(0, 10)$ and $\tau_c^2 \sim \text{gamma}(0.01, 0.01)$. The priors for the intercepts are the same as in the unpenalized model. Use of the stochastic variance is analogous to the frequentist tuning parameter search (as discussed in Section 2.1); the result is that coefficients are smoothed toward a common mean where the amount of smoothing depends on the posterior estimate of the tuning parameter.

3.3. Nonlinear shrinkage

Nonlinear shrinkage is sensible when one wishes to deduce the discriminating variables from a large number of possibly irrelevant variables. Yuan and Lin [7] showed a connection between Bayesian variable selection and the Bayesian lasso; the latter can be derived as the posterior mode under a double exponential prior, whereas the former uses ‘spike and slab’ prior that encourages sparsity [31, 32]. George and McCulloch proposed a similar variable selection technique for the purpose of model selection in the linear regression setting [33].

To perform variable selection via NLS, we employed the spike and slab prior for the non-intercept elements of β . This mimics the frequentist behavior of the lasso by identifying a subset of variables that are useful in predicting survival by setting coefficients almost exactly to zero for variables that have little discriminatory ability. We assume

$\beta_{cj} | \gamma_{cj} \sim a_0 \gamma_{cj} + (1 - \gamma_{cj}) N\left(\mu_c, \frac{1}{\tau_c^2}\right)$, where $\mu_c \sim N(0, 10)$, $\tau_c^2 \sim \text{gamma}(0.01, 0.01)$, $\gamma_{cj} \sim \text{Bern}(p_j)$, and a_0 is a point mass at 0. The point mass facilitates variable selection through the conditional posterior distribution of γ_{cj} . The mixture prior elicits a variable selection procedure that sets some of the coefficients almost exactly equal to 0; thus, the β_{cj} will be selected on the basis of how well they can be distinguished from 0 rather than their absolute size. The prior on the variance hyperparameter controls the amount of smoothing. The Bernoulli probability can be altered depending on which of J variables are, a priori, expected to be important. This may be interpreted as the researcher’s prior probability that the specific Y_j s are associated with latent class. In all analyses, we assume that roughly 50% of the variables are associated with latent class and set $p_j = 0.5$.

3.4. Conditional dependence model

Violations of the conditional independence assumption (i.e., the assumption that conditional on latent class, the observed variables are independent) can be assessed by classifying patients into the latent class for which they have the highest posterior probability and assessing within-class correlation [18]. If violations occur, it is possible to relax the assumption by allowing mutual associations between variables within a latent class. Following the loglinear notation for a binary outcome, $Y_j, j = 1, \dots, J$ and latent class indicator, C , consider the expected cell frequency $u_{y_1 \dots y_J c}$ in the table formed by cross-classifying J variables and latent class [34–36]. The unsupervised conditional independence latent class model can be expressed in loglinear notation as

$$\log(u_{y_1 \dots y_J c}) = \Lambda + \sum_{j=1}^J \Lambda_j^Y I[y_j=1] + \Lambda_c^C + \sum_{j=1}^J \Lambda_{jc}^{YC} I[y_j=1, c] I[\gamma_{cj}=1],$$

where for identification $\Lambda_1^C = \Lambda_{j1}^{YC} = 0$, for all $j = 1, \dots, J$. Here, I is an indicator function, and as in Section 3.3, γ_{cj} denotes whether latent class is important for variable j . A general conditional dependence (CD) model that allows pairwise intra-class correlation between variables to vary across pairs of variables is

$$\begin{aligned} \log(u_{y_1 \dots y_J c}) = & \Lambda + \sum_{j=1}^J \Lambda_j^Y I[y_j=1] + \Lambda_c^C + \sum_{j=1}^J \Lambda_{jc}^{YC} I[y_j=1, c] I[\gamma_{cj}=1] \\ & + \sum_{i \neq j} \Lambda_{ij}^{YY} I[y_i=1, y_j=1] + \sum_{i \neq j} \Lambda_{ijc}^{YYC} I[y_i=1, y_j=1, c] \max[\gamma_i, \gamma_j] \end{aligned} \quad (4)$$

where $\Lambda_{ij1}^{YYC} = 0$ for all i, j , for identifiability. To reduce the dimensionality of the CD model, constraints can be imposed to force the within-class interaction terms Λ_{ijc}^{YYC} to be constant either across outcomes (i and j) within a latent class or across classes (c). For example, setting $\Lambda_{ijc}^{YYC} = 0$ for all c specifies two-way interactions that are constant across latent class, resulting in a more parsimonious model. We could also incorporate further constraints by specifying only one interaction parameter for all i, j variables within a class. We make both of these assumptions when fitting CD models to the glioma data in Section 6. The CD model is supervised through the incorporation of the survival likelihood contribution as discussed in Section 3.1

To fit the model in (4), we specify a likelihood and prior distributions for all model parameters. First, we assume that the number of subjects in latent class $C = c$ with $Y_1 = y_1, \dots, Y_J = y_J$ is distributed as a Poisson random variable with mean $u_{y_1 \dots y_J c}$. We set noninformative normal prior distributions for the collection of parameters, λ , and NLS priors for the collection of γ parameters. The latent class variable for each individual represents missing data. In the Bayesian framework, we treat these variables as model parameters and assume Dirichlet prior distributions for each subject's latent class C . The algorithm iterates between updating the collection of Λ parameters (given (Y, C)) and then updating latent class assignment, C , given (Λ, Y) .

4. Markov chain Monte Carlo estimation

Because the form of the joint posterior distribution is mathematically intractable, we cannot sample from this distribution directly. To obtain samples from the posterior distributions of the model parameters, we sample from the conditional posterior distributions by using MCMC, as implemented via the Gibbs sampler in WinBUGS version 1.4 [37, 38]. For

simulations, following a burn-in of 1000 iterations, we monitor the posterior distributions over a further 1000 iterations of the chain. We summarize the posterior distributions by using means, medians and standard deviations. We also obtain the posterior distribution of latent class membership for each subject. For the glioma data in Section 6, following a burn-in period of 15 000 iterations, we ran the Gibbs sampler for an additional 30 000 iterations, saving the results of every third iteration.

In some hierarchical formulations, assuming a gamma versus a uniform prior for the variance components can lead to different inferences [39]. Therefore, we conducted a sensitivity analysis to determine whether or not the choice of prior distributions of these parameters affects our resulting inferences. We considered a uniform density, $\tau_c^2 \sim \text{Uniform}(0, 10)$, and in fact, the choice of prior did not affect our results (results not shown).

4.1. Initial values

We obtain the values $(\tilde{\beta}_{01}, \dots, \tilde{\beta}_{(K-1)J})$, $\tilde{\gamma}$, $\tilde{\tau}_c$, $\tilde{\gamma}_{cj}$ and $\tilde{\lambda}_l$ to initialize the Gibbs sampler from first-order approximations. First, subjects are assigned to each of K latent classes based on quantiles of the mean number of responses on the J binary variables, providing initial values for the prevalence parameters, $\tilde{\eta}$. For responses that are heterogeneous in interpretation, nonparametric clustering procedures could also be used. Using this class assignment, we determine the initial values $(\tilde{\beta}_{01}, \dots, \tilde{\beta}_{(K-1)J})$ and the hyperparameter, $\tilde{\tau}_c$, using univariate logistic regression. Similarly, we obtain survival coefficients, $\tilde{\gamma}$, by fitting a proportional hazards model to the survival data. We obtain the class-specific initial values, $\tilde{\gamma}_{cj}$, by simulating a random draw from a Bernoulli distribution with probability, $p_j = 0.5$, and $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_L)$ from random draws from an inverse gamma distribution with shape and scale parameter equal to 0.01. We could have also used the Nelson–Aalen estimates of baseline hazard as starting values, although our simpler approach works well in practice.

4.2. Model selection and identifiability

The latent class survival model assumes that the number of classes, K , is fixed. In the current setting, a parsimonious model with just a few survival-driven latent classes that have clear interpretations is preferred to one that provides a slightly better fit but is more difficult to interpret. To calculate the effective number of parameters, p_K , and compare models based on different values of K , one could consider the deviance information criterion (DIC) [40].

The DIC is defined as $-4E_\theta [\log f(X|\theta)] + 2\log f(X|\tilde{\theta})$, where $\tilde{\theta}$ is commonly taken as the posterior mean. A plug-in estimate of the posterior mean of θ for the second term is not appropriate because of the discrete nature of classes, but it is possible to replace this with a weighted average across classes, where the deviance is evaluated at the posterior mean. The model with the lowest DIC is preferred, and DIC differences of 5 or more between models are considered substantial.

Poor behavior of the DIC has been well-documented in the context of discrete latent variables because there is no obvious plug-in estimate for θ . Nonsensical values of p_K resulted in some of our high-dimensional settings rendering DIC incalculable [41]. We also applied alternative DIC measures that were proposed specifically for comparison of mixture models [42], many of which still resulted in negative p_K . Further research into a model selection criterion perhaps is needed; however, for simulations, we are typically able to calculate and present DIC₅ [42]. For the glioma model comparison, DIC is not calculable, so we use deviance measures as an informal summary of model fit.

Finally, to assess identifiability of parameters, we compare plots of prior and posterior parameter distributions [43]. Suppose that $f(\theta; X)$ denotes the model, and θ is partitioned as $\theta = (\theta_1, \theta_2)$. If $f(\theta_2|\theta_1, X) = f(\theta_2|\theta_1)$, then θ_2 is not identifiable; that is, its prior distribution supplies the only information available about θ_2 , and so $f(\theta_2|X) \cong f(\theta_2)$ [44]. Plots indicating poor identifiability display strong overlap of prior and posterior distributions; otherwise, the posterior distribution should be sufficiently shifted from the prior distribution. We consider these plots in simulations and present them for the glioma data in Section 6.

5. Simulations

We perform a simulation study assuming several high-dimensional settings to analyze the behavior of the methodologies. The goal is to show that when J is a substantial fraction of N , regularization is necessary to stabilize parameter estimation in the joint survival and latent class setting. We vary the number of subjects, N , and the number of variables, J , to see how well the methods perform with respect to bias, standard errors and survival prediction and whether we improve upon unpenalized methods. Let $\tilde{J}$ denote the number of variables associated with latent class in the conditional independence model. We denote the set of coefficients for the $\tilde{J}$ variables that are associated with latent class by $\tilde{\mathcal{J}}$ and the remainder of the coefficients by $\mathcal{J} - \tilde{\mathcal{J}}$. We generate values according to an effect size of 1.5 on the logit scale. For all simulations, we consider a two-class model where the latent class indicator, U_i , is generated as a Bernoulli random variable with probability equal to 0.35. We generate survival times according to a skewed Weibull distribution with shape parameter equal to $\frac{1}{2}$ and scale parameter dependent on latent class membership and with the log HR (class 2 vs class 1), ν , equal to -0.5 . We simulate censoring times from a Uniform(3,6) distribution to mimic a patient accrual process with approximately 20% censoring and take the minimum of survival and censoring time as the observed time to an event. Because true latent class is known, we assess the sensitivity and specificity of each method in identifying the class to which each subject was classified by using receiver operator characteristic (ROC) analysis. That is, we compare probabilities of being assigned via highest posterior probability of latent class membership to the correct class (sensitivity) versus the probability of being assigned to the wrong class (false positive) via area under the ROC curve (AUC), providing a univariate measure of quality of the classification. We present DIC_5 of Celeux *et al.* [42], (where calculable) as well as the root mean square error (rMSE) averaged over

$$\text{simulations, } rMSE = \left\{ \sum_{c=1}^K (\hat{\eta}_c - \eta_c)^2 + \sum_{c=1}^K \sum_{j=1}^J (\hat{\pi}_{cj} - \pi_{cj})^2 + \sum_{c=1}^K (\hat{\nu}_c - \nu_c)^2 \right\}^{1/2}.$$

Tables II, III and IV present simulation results for several scenarios. Posterior means and standard errors of the $\tilde{J}$ variables, survival coefficients, posterior means of the class prevalence and the mean of all posterior means and posterior standard errors of the remaining $J - \tilde{J}$ variables are listed. We include a low-dimensional case where $N = 1000$, $J = 10$ and $\tilde{J} = 10$ in Table II in order to demonstrate a situation where all methods give similar estimates. These results show that when $N \gg J$ the unpenalized approach naturally performs well and results in the lowest DIC_5 .

As expected, in high-dimensional scenarios, the unpenalized model results in biased parameter estimates with large standard errors. However, the NLS and ridge methods provide markedly less biased estimates of both the zero and nonzero coefficients and also perform better with respect to standard errors, AUC and rMSE. In the two-class setting, the identifiability of parameters using the unpenalized approach is questionable even with moderate-sized J . In the left panel of Table III, where $N = 200$; $J = 50$; $\tilde{J} = 20$; $\beta_{1,14}$ and $\beta_{1,16}$ are severely biased (note that 112 parameters need to be estimated with an N of only 200). In the right panel, where J is increased to 80, all unpenalized estimates, including the log

HR, are markedly biased, and the mean of the $(J - \hat{J})$ elements that are not associated with latent class is -1.50 even though it should be zero. In contrast, in the same figure, the NLS and ridge estimators are only slightly biased toward the null. More importantly, these models are able to discern the zero from nonzero β elements as well as the correct log HR, ν . Weak identifiability in the unpenalized model is also evidenced by negative p_K obtained for most simulations, and for this reason, DIC_5 is generally incalculable for the unpenalized model fit to high-dimensional simulated data.

In Tables II, III and IV, NLS results in the lowest rMSE and appropriately sets coefficients of non-discerning variables, on average, to 0.01. The ridge solution also does well when only a small proportion of the variables distinguish latent class. This appears to be true for both censored and uncensored data. Thus, ‘borrowing’ information across variables is beneficial in this setting, even though, in general, the ridge does not set non-discerning variable coefficients to zero. Across simulations, the ridge standard error estimates are also markedly smaller than those for the NLS parameterization possibly because the ridge parameterization uses more variables to estimate the elements of β whereas the NLS uses only those actually associated with latent class. In the high-dimensional setting, the NLS and ridge approach both outperform the unpenalized approach with respect to standard errors and AUC. Smaller AUCs for the unpenalized scenarios presented in Tables III, II and IV suggest that posterior probabilities of latent class membership obtained from the unpenalized model frequently misclassify subjects. In comparison, the NLS and ridge methods accurately classify subjects in most scenarios considered, with AUCs greater than 0.90. It may seem surprising that ridge estimates are less biased than NLS estimates in fairly sparse situations; this may be due to the fact that many more parameters must be estimated to achieve nonlinear shrinkage, which can affect mixing of the Gibbs sampler.

Overall, the simulations demonstrate the necessity of penalization in clustering of moderate to high-dimensional data, in the presence of survival. Although the unpenalized model is appropriate for $N \gg J$, it results in biased parameter estimates, inflated variance and lack of identifiability of model parameters when J is even moderately large compared with N . In addition, AUCs demonstrate that latent class prediction is poor. In contrast, the penalized models fit the data well even when J is a substantial fraction of N . We note that with uncensored data (Table IV), we were able to fit models with much higher J relative to N .

Whether or not a variable is associated with latent class is estimated by the posterior mean of the corresponding γ element in the NLS model described in Section 3.3. In the two-stage approach of Bair *et al.*, the discerning variables are obtained by ranking univariate Cox scores (standardized coefficients from a proportional hazards model [25]). To compare how well NLS performs versus univariate selection based on Cox scores, we conduct an ROC analysis to assess the sensitivity and specificity of each approach in identifying variables that are associated with survival. Because the true indicator of association is known in simulation, we compare this with the posterior mean of γ_{cj} , the indicator of association between a variable and latent class. Table V compares AUCs for the NLS method and the pre-processing Cox score stage of the two-stage method [25]. With respect to AUC, the NLS method outperforms the naive univariate approach in discriminating the important variables in this simulation setting.

After retaining variables with the highest univariate Cox scores in the first stage, we apply a latent class analysis to these variables in the second stage. The right-hand columns of Table V show the Wald test statistic for the log HR resulting from a weighted proportional hazards model with latent class as estimated using NLS as a predictor and the weighted Wald test statistic for the two-stage model (with weights equal to the posterior probabilities of latent class membership). The NLS approach consistently outperforms the two-stage approach

with respect to the Wald statistic demonstrating the advantage of our joint supervised approach that considers all candidate variables simultaneously.

6. YKL-40 in gliomas

The data consist of 11 immunohistochemical (IHC) variables relating to YKL-40, two histopathological variables and three clinical variables (seven ordinal, six dichotomous and one continuous) measured on 206 glioma brain tumor patients (Table I). Because the data are a mix of true binary variables and ordinal variables with many unrepresented levels, it was most sensible to dichotomize all IHC variables into presence/absence. In general, there are a preponderance of zeros in the data, which also lends justification to a binary rather than an ordinal representation of the IHC variables in our dataset. We also dichotomize age at 70 years and Karnofsky performance status at 70, which are clinically relevant cut points in the disease process. The study measured time to death or exit from the study in days and censoring status; 73 patients were censored, and 133 died during follow-up. The focus of the analysis is on the estimation and interpretation of a small number, K , of latent classes after including all measured variables in the analysis, with supervision by the survival outcome. Our primary goal is to identify which, if any, immunohistochemical features of YKL-40 are useful for refining established diagnostic categories to render them more prognostic for survival in the presence of histopathological and clinical features and clinical diagnosis. Finally, we compare the conclusions from our analysis with those obtained using the two-stage approach.

6.1. Number and interpretation of latent classes

First, we fit conditional independence models to the data. Table VI lists the deviance ($-2 \times \log\text{-likelihood}$) for each of three Bayesian latent class models fit using all 16 variables plus diagnosis. All models result in a negative effective number of parameters. Because the implications of this for these models have not been established, we rely on the deviance measure to select the best model and further explore model fit via the graphical displays in Figure 1. Figure 1 displays plots of the posterior and prior distributions for select β_{2j} parameters for each model indicating weak identifiability of the unpenalized three-class model, thereby excluding this model from further consideration. However, it appears that the penalized solutions are well-identified, despite the negative degrees of freedom associated with the DIC. As the three-class NLS model exhibits the lowest deviance of all fits considered, this is the model we report. When the prior distribution over the models is uniform, it can be shown that selecting the lowest deviance is equivalent to selecting the highest posterior model probability. However, although the models compared have equal priors within a single choice of K , this is not necessarily true across different choices of K .

As a result of violations of conditional independence within feature groups (i.e., IHC, histopathological and clinical), we incorporate first-order interactions between variables within a latent class as discussed in Section 3.4. A priori, variables within a feature group are more highly correlated than variables between groups. Consequently, we constrain two-way interactions to be constant across feature group, but these interactions are allowed to vary across class. Further, we fit a common interaction model that assumes one interaction term across all i, j variables within a latent class. Similar to the other coefficients, we set noninformative normal priors for interaction terms.

Because interaction terms across classes and feature groups for the full model exhibit overlapping 95% credible intervals and because model deviance is much smaller for the three-class NLS with a common interaction term (Deviance = 4527.63) versus the three-class NLS with feature-specific and class-specific interaction terms (Deviance = 7926.55), the parsimonious model assuming a common interaction among observed variables is

avored. The deviance for the three-class CD NLS model is lower than that for any of the conditional independence models, suggesting that relaxing this assumption improves model fit.

Because of the relaxation of the conditional independence assumption across the features, the tabulated class-specific probabilities depend on the presence or expression of other variables within a group. Figure 2 displays the probability of each variable conditional on latent class where this probability, because of the inclusion of an interaction term, is conditional on the number of other variables expressed. This probability changes as one moves up the y -axis, i.e., the number of other binary variables expressed. Black rectangles indicate a probability of 1, whereas white rectangles indicate a probability of 0; different probabilities resulting from the relaxation of conditional independence are evident by the shades of gray along the y -axis within a class.

The posterior mean hazard rates [95% credible interval] associated with membership in class 2 versus 1 and class 3 versus 2 are 4.22 [2.62, 6.86] and 17.19 [10.05, 31.16], respectively. The classes in the figure are presented in order from best to worst prognosis, and the median survival times for patients assigned to each class were 1279, 469 and 115 days, respectively.

6.2. Utility of YKL-40

One goal is to observe whether expression of the 11 YKL-40 immunohistochemical variables differs across survival-driven latent classes. Figure 2 shows that many of the YKL-40 expression patterns do not appear to be differentially expressed across latent classes, so it is not surprising that the model incorporating variable selection performs the best. Important features include normal blood vessel staining, extracellular matrix (ECM) staining, MVP staining and perivascular staining. The most important of these is normal blood vessel staining as members in class 1 are much more likely to exhibit normal blood vessel staining than members in the poorer prognosis classes. The model-based estimates reflected in this graph imply that this YKL-40 staining pattern is very sensitive for good prognosis as a positive stain is highly indicative of a better prognosis. In addition, this conclusion does not change as more features are expressed (i.e., as one moves up the y -axis). On the contrary, patients in class 1 are overall less likely than members of classes 2 and 3 to exhibit ECM staining, so we conclude that this staining pattern is sensitive for poorer prognosis. However, the class distinction based on this particular feature is lost as more features are expressed. Finally, members in classes 1 and 3 are both more likely to exhibit MVP staining than members in class 2 but less likely to exhibit perivascular staining, indicating that these particular staining patterns are not particularly sensitive in light of the other variables. Overall, we conclude that even in the presence of diagnosis and survival, latent class is additionally informative, and indeed some of the YKL-40 variables are responsible for the additional information garnered.

6.3. Predictive ability

We use the C -statistic as a measure of prediction accuracy of the three-class NLS model [45]. We fit two proportional hazards models, one with latent class as a predictor (observed $C = 0.77$, cross-validated $C = 0.76$) and one with diagnosis as a predictor (observed $C = 0.68$, cross-validated $C = 0.66$). The relatively larger statistic demonstrates the improved ability of our model-based classifications over diagnosis to predict survival.

Figure 3 displays heat maps of the tumor variables. Both heat maps show the 17 variables along the x -axis, which are hierarchically clustered across patients by using the Ward method (clusterings are visualized in the top dendrogram). Purple corresponds to expression or presence of the variable, yellow corresponds to absence, and white denotes that the

measure is missing for that patient. Figure 3a includes an annotation track for the diagnosis of GBM for each patient, and Figure 3b includes an annotation track for the posterior probabilities of class membership, both along the y -axis. The y -axis in Figure 3a blue indicates a diagnosis of AA/AO/AOA and red indicates a diagnosis of GBM. On the y -axis in Figure 3b contains shades of blue, green and red denoting the posterior probabilities of membership in latent classes 1, 2 and 3, respectively (as in Figure 2, the classes are ordered from longest to shortest survival times). The darker the shading, the closer the posterior probability is to 1. The two annotation tracks can be compared to demonstrate the additional information that estimated latent class provides over diagnosis alone.

The first thing to note is that the posterior probabilities approached 0 or 1 so we did not observe much ambiguity in latent class assignment. Comparing the two annotation tracks and as discussed in the previous subsection, five of 83 latent class 1 (blue) members have a non-GBM diagnosis, class 2 (green) members have varying diagnoses, and 40 of 44 class 3 (red) members have a GBM diagnosis. So whereas latent classes 1 and 3 typically follow diagnoses of non-GBM versus GBM, respectively, latent class 2 further discriminates a subgroup of 23 GBM patients who exhibit a better prognosis than the typical GBM patient (the median survival time for class 2 is 469 days, whereas the median survival time for GBM patients overall is only 288 days). Interestingly, falling into this latent class correlates well with having undergone resection for the tumor; thus, this variable helps distinguish the better prognosis GBMs from the other GBMs. Overall, the ability to further differentiate two distinct prognostic subgroups of GBM patients (e.g., falling into class 2 versus class 3) demonstrates the appeal of our supervised latent class model over grouping by diagnosis alone.

The purpose of a biostatistical model is to describe the current data as well as make predictions about future patients. For a new brain tumor patient, i , we could potentially observe their entire data vector, y_i ; however, a priori we would know neither their latent class membership indicator, C_i , nor their survival time and censoring status (t_i , δ_i). Given that patient i 's survival time is assumed to be conditionally independent of y_i given latent class membership, it is straightforward to calculate the unsupervised posterior probability of latent class membership given y_i when (t_i , δ_i) are unknown, as the expression $P(T_i = t_i, \Delta_i = \delta_i | C_i = c)$ factors out of the likelihood. The ability to make straightforward prediction in the absence of survival was the rationale for formulating survival and variable expression in a conditionally independent manner while still allowing the model to flexibly incorporate many conditionally dependent binary variables.

6.4. Comparisons with other models

We compare our results with those obtained by applying the two-stage technique outlined in [25]. In stage 1, we ranked univariate Cox scores and keep only those variables greater than the optimal threshold as determined by 10-fold cross-validation. The optimal Cox score cutoff is 1.53 resulting in the retention of nine variables—all five of the histological and clinical variables and four of the YKL-40 variables, including both of the astro scoring variables, normal blood vessel staining and ECM staining intensity. In stage 2 we perform a supervised latent class analysis on these nine variables and assign patients to latent classes based on their highest posterior probabilities. We subsequently fit a proportional hazards model with latent class included as a covariate in order to assess utility of latent class assignment in predicting survival from this two-stage approach. The primary difference between this comparative approach and our approach is that this approach pre-selects variables whereas our approach selects variables as part of the estimation procedure. The resulting HRs [95% confidence interval] comparing the worst with best and worst with middle prognosis latent class from the two-stage approach are 4.36 [2.63, 7.25] and 2.04 [0.42, 1.52], respectively. These HRs are substantially smaller than those reported from the

supervised CD model in Section 6.1, and the second one did not reach significance. This may be due to the fact that spurious univariate associations affect which variables move forward to the second (clustering) stage of the model. Overall, whereas two-stage approaches retain variables that are highly associated with survival on univariate analysis, our approach results in groupings that are more survival driven, providing identifiable parameter estimates obtained from a joint probability model in the presence of multi-collinearity.

Finally, to assess predictive ability of approach compared with a commonly used variable selection model, we fit the L1 (lasso) penalized Cox regression model [46] are all variables where considered on their original scale. The variables with nonzero coefficients in order of importance include: diagnosis, resection, microvascular proliferation, astro scoring intensity, necrosis and ECM staining intensity. Whereas our Bayesian NLS approach to classification also indicates that diagnosis, resection and ECM staining are sensitive for latent class, the variable normal blood vessel staining is among the most sensitive for latent class, with its expression indicative of a good prognosis and upheld by the fact that the posterior mean of the variable selection parameters for this variable is 1. The lasso method does not select this variable. In addition, MVP and perivascular staining were determined to be somewhat sensitive for latent class in the NLS model but are not selected using the lasso.

The observed and cross-validated C -statistics resulting from the lasso penalized Cox model are 0.73 and 0.71, respectively. As mentioned in Section 6.3, the post hoc observed and cross-validated C -statistic resulting from the Cox regression of latent class on survival are 0.77 and 0.76, respectively, indicating a greater predictive ability of our approach than the lasso penalized Cox model.

7. Discussion

There is considerable intratumoral heterogeneity and interobserver variability in the current histopathological diagnosis of gliomas. For this reason, a study was conducted to investigate whether diagnoses could be refined through consideration of YKL-40 immunohistochemical staining, with supervision by survival, in addition to histopathological and clinical variables. Although appropriate for a small number of variables, an unpenalized supervised latent class model produces weakly identifiable estimates for even a moderate J . However, we were able to use penalization and variable selection within the supervised latent class analysis framework to fit well-identified, conditionally dependent two-class and three-class models, allowing us to consider the diagnostic and prognostic utilities of YKL-40. The new methods offer a unifying probability model that accommodates high-dimensional binary data in the presence of survival data. We discerned three distinct survival groups that correlate with diagnosis, possibly providing new diagnoses that are more informative about prognosis. YKL-40 increases the precision of survival estimates allowing for more effective classification of high-grade gliomas. Importantly, with the inclusion of YKL-40, we were able to ‘reclassify’ a small group of patients. These discordant patients (those whose prognosis differs with and without the inclusion of YKL-40) are of great interest in this and in future studies.

In simulations we obtain less biased estimates than the unpenalized model, especially in cases where many items play no role in discerning latent class. In the high-dimensional setting, penalized parameter estimates for the class-specific parameters, β , as elicited through prior specification, are easily calculated using the Gibbs sampler. The NLS standard error estimates are generally larger than those for the ridge parameterization. This is not surprising because the former uses only those variables actually associated with latent class to estimate the elements of β . For YKL-40, although both methods yielded similar results, it

is not surprising that the NLS is preferred because the majority of YKL-40 immunohistochemical variables did not discern latent class. An ROC analysis also showed that the NLS model outperforms the two-stage supervised approach in the settings we considered, and applying a two-stage approach to the glioma data did not result in significant HRs. Thus, pre-screening using Cox scores with varying thresholds does not necessarily select the appropriate variables that discriminate between classes.

We also applied our methodology to the LOH data from [13]. These data consist of 19 binary markers of LOH measured on 93 oligodendroglioma patients. We compared our results with those obtained in their unsupervised, penalized analysis where DIC preferred a three-class model, uncovering low, low/moderate and high LOH profiles. In their second-stage survival analysis, the high and low LOH classes differed significantly with respect to patient survival, but the low/moderate LOH class did not differ from the low LOH class. Our supervised analysis was consistent with these results, as a two-class model corresponding to low and high LOH profiles was preferred; that is, the influence of survival on the latent class structure made the prevalence of the low/medium LOH tend toward zero. The two estimated latent classes corresponded to high and low LOH probabilities, respectively, across all 19 markers and were highly associated with survival, as the hazard for the low LOH class was 3.33 times that for the high LOH class. This result validates the ability of survival to drive the latent class structure and corroborates previous findings.

There are some interesting topics that future research might address. First is the numerical instability encountered when the class-specific probabilities approach either 0 or 1. This is problematic mainly when fitting models with more than two latent classes and where the classes are not well-defined. Although our penalized methods seek to avoid this issue by putting a hyperprior on the variance parameter, in some cases it is still necessary to avoid probabilities on the edge of the parameter space by bounding the logit of the class-specific probabilities. Secondly, the penalized methods easily converge for situations in which J is approximately equal to N ; however, no methods reliably converged for J larger than N . Thus, in these situations, some pre-processing of the data based on associations with survival might be necessary in order to fit high-dimensional latent class models. Thirdly, a problem that has been addressed in other research contexts is that normal priors with large variances for the regression coefficients can result in unnatural priors for the underlying probabilities [47]. In the context of Bayesian logistic regression, one can address this problem by specifying a prior for the outcome probabilities rather than for the model coefficients themselves and by using a change-of-variables technique to induce a prior distribution for the coefficients. One of the benefits of this formulation is that it is simpler to elicit information about success probabilities as opposed to the logit of the probabilities. It may be worthwhile to implement this approach in the latent class setting as well. Finally, investigation of model selection in this context is needed. We considered several DIC criteria, and their poor behavior was evident in simulations and the data example, forcing us to rely on model selection by using the deviance. A way to avoid relying on a selection criterion would be to utilize the reversible-jump technique to estimate the number of latent classes [21]; however, this is not straightforward to apply in the current context. One could also consider Dirichlet process priors [48]. It would be interesting to explore this further as it has been successfully applied in unsupervised continuous data settings.

References

1. Nutt CL, Mani DR, Betensky RA, Ramayo P, Cairncross JG, Ladd C, Pohl U, Hartmann C, McLaughlin ME, Batchelor TT, Black PM, Von Deimling A, Pomeroy SL, Golub TR, Louis DN. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Research*. 2003; 63:1602–1607. [PubMed: 12670911]

2. Nutt CL, Betensky RA, Brower MA, Batchelor TT, Louis DN, Stemmer-Rachamomov AO. YKL-40 is a differential diagnostic marker for histological subtypes of high-grade gliomas. *Clinical Cancer Research*. 2005; 11(6):2258–2264. [PubMed: 15788675]
3. Rousseau A, Nutt CL, Betensky RA, Iafrate AJ, Han M, Ligon KL, Rowitch DH, Louis DN. expression of oligodendroglial and astrocytic lineage markers in diffuse gliomas: use of YKL-40, ApoE, ASCL1, and NKX2-2. *Journal of Neuropathology and Experimental Neurology*. 2006; 65(12):1149–1156. [PubMed: 17146289]
4. Fowlkes EB, Gnanadesikan R, Kettering JR. Variable selection in clustering. *Journal of Classification*. 1998; 5:205–228.
5. Brusco MJ, Cradit JD. A variable selection heuristic for K -means clustering. *Psychometrika*. 2001; 66:249–270.
6. Larsen K. Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*. 2004; 60:85–92. [PubMed: 15032777]
7. Yuan M, Lin Y. Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*. 2005; 100:1215–1225.
8. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*. 1974; 61:215–231.
9. Hoijtink H. Constrained latent class analysis using the Gibbs sampler and posterior predictive p -values: applications to educational testing. *Statistica Sinica*. 1998; 8:691–711.
10. Pan W, Shen X. Penalized model-based clustering with application to variable selection. *Journal of Machine Learning Research*. 2006; 8:1145–1164.
11. Wang S, Zhu J. Variable selection for model-based high dimensional clustering and its application to microarray data. *Biometrics*. 2007; 64:440–448. [PubMed: 17970821]
12. Guo J, Levina E, Michailidis G, Zhu J. Pairwise variable selection for high-dimensional model-based clustering. *Biometrics*. 2010; 66(3):793–804. [PubMed: 19912170]
13. Houseman EA, Coull BA, Betensky RA. Feature-specific constrained latent class analysis for genomic data. *Biometrics*. 2006; 62:1062–1070. [PubMed: 17156280]
14. DeSantis SM, Houseman EA, Coull BA, Stemmer-Rachamimov A, Bentensky RA. A penalized latent class model for ordinal data. *Biostatistics*. 2007; 9(2):249–262. [PubMed: 17626225]
15. Shedden K, Zucker RA. Regularized finite mixture models for probability trajectories. *Psychometrika*. 2008; 73(4):625–646. [PubMed: 19956348]
16. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*. 1996; 58:267–288.
17. Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric Regression*. Cambridge University Press; New York: 2003.
18. Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ. Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*. 1997; 92(440):1375–1386.
19. Moustaki I, Steele F. Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in Bangladesh. *Statistical Modelling*. 2005; 5:327–342.
20. Lin HQ, Turnbull BW, McCulloch CE, Slate EH. Latent class models for joint analysis of longitudinal biomarker and event process data. *Journal of the American Statistical Association*. 2002; 457:53–65.
21. Tadesse MG, Sha N, Vannucci M. Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Society*. 2005; 100(470):602–617.
22. Kim S, Tadesse MG, Vannucci M. Variable selection in clustering via Dirichlet process mixture models. *Biometrics*. 2006; 93(4):877–893.
23. Raftery A, Dean N. Variable selection for model-based clustering. *Journal of the American Statistical Association*. 2006; 101:168–178.
24. McLachlan RW, Bean D, Peel D. A mixture model-based approach to the clustering of microarray expression data. *Bioinformatics*. 2002; 18:413–422. [PubMed: 11934740]
25. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLOS Biology*. 2004; 2(4):511–522.

26. Bair E, Hastie T, DeBashis P, Tibshirani R. Prediction by supervised principle components. *Journal of the American Statistical Association*. 2006; 101(473):119–137.
27. Tan Q, Thomassen M, Kruse TA. Feature selection for predicting tumor metastases in microarray experiments using paired design. *Cancer Informatics*. 2007; 3:213–218. [PubMed: 19455244]
28. Baker SG, Kramer BS. Identifying genes that contribute most to good classification in microarrays. *BMC Bioinformatics*. 2006; 7:407. [PubMed: 16959042]
29. Lindsey JC, Ryan LM. Tutorial in biostatistics methods for interval-censored data. *Statistics in Medicine*. 1998; 17:219–238. [PubMed: 9483730]
30. Ibrahim, JG.; Chen, MH.; Sinha, D. *Bayesian Survival Analysis*. Springer; New York: 2001.
31. Park T, Casella G. The Bayesian lasso. *Journal of the American Statistical Association*. 2008; 103(482):681–686.
32. Ishwaran H, Rao JS. Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics*. 2005; 33:730–773.
33. George EI, McCulloch RE. Approaches for Bayesian variable selection. *Statistica Sinica*. 1997; 7:339–373.
34. Agresti A, Lang JB. Quasi-symmetric latent class models, with application to rater agreement. *Biometrics*. 1993; 49:131–139. [PubMed: 8513099]
35. Yang I, Becker MP. Latent variable modeling of diagnostic accuracy. *Biometrics*. 1997; 53:948–958. [PubMed: 9290225]
36. Hagenaars JA. Latent structure models with direct effects between indicators: local dependence models. *Sociology Methodology Research*. 1988; 16:379–405.
37. Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. *The Statistician*. 1994; 43:169–178.
38. Spiegelhalter, DJ.; Thomas, A.; Best, NG.; Gilks, WR.; Lunn, D. *Bayesian Inference Using Gibbs Sampling*. MRC Biostatistics Unit; Cambridge, England: 2003. 1994 www.mrc-bsu.cam.ac.uk/bugs/
39. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006; 1(3):515–534.
40. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*. 2002; 64(4):583–639.
41. DeIorio M, Robert CP. Discussion of Spiegelhalter et al. *Journal of the Royal Statistical Society, Series B*. 2002; 64:629–630.
42. Celeux G, Forbes F, Robert CP, Titterton DM. Deviance information criteria for missing data models. *Bayesian Analysis*. 2006; 1(4):1–23.
43. Garrett ES, Zeger SL. Latent class model diagnosis. *Biometrics*. 2000; 56:1055–1067. [PubMed: 11129461]
44. Gelfand, AE.; Sahu, SK. Identifiability, propriety and parameterization with regard to simulation-based fitting of generalized linear mixed models. University of Connecticut, Storrs, Dept. of Statistics; 1996. Technical Report 96-36
45. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine*. 2004; 23:2109–2123. [PubMed: 15211606]
46. Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in Medicine*. 1997; 16(4):385–395. [PubMed: 9044528]
47. Bedrick EJ, Christensen R, Johnsonm W. Bayesian methods for binomial regression. *The American Statistician*. 1997; 51:211–218.
48. Pennell ML, Dunson DB. Fitting semiparametric random effects models to large data sets. *Biostatistics*. 2007; 8(4):821–834. [PubMed: 17429104]

Figure 1.

Prior (solid) versus posterior (dotted) distributions of logit probabilities for three-class model parameters fit to the YKL-40 glioma data. The plots compare weakly identifiable β_{1j} parameters obtained from fitting the three-class unpenalized model to the glioma data, with the corresponding well-identified three-class NLS and ridge parameters.

Figure 2. Class-specific probabilities resulting from fitting the three-class NLS conditional dependence model. The x-axis displays the variables of interest, whereas the y-axis displays the total number of binary variables present/expressed. Black rectangles indicate a probability of 1, whereas white rectangles indicate a probability of 0. Latent class prevalence is 0.40, 0.39 and 0.21, respectively.

Figure 3.

The first figure is a heat map of variables with an indicator of GBM (red) diagnosis on the annotation track. The second figure is a heat map of variables with the posterior probabilities of latent class on the annotation track. Latent classes are ordered from best (class 1, blue) to worst (class 3, red) prognosis. Purple is expressed (1), yellow is not expressed (0), and white is missing.

Table I

Prevalences of histological features by tumor type and estimates obtained from fitting the ridge model to the YKL-40 glioma data.

Variable	Type	Prevalence				
		AO	AOA	AA	GBM	Overall
Number of cells stained (Astro scoring)	IHC	0.14	0.36	0.50	0.77	0.66
Staining intensity (Astro scoring)	IHC	0.11	0.24	0.47	0.75	0.60
Number of cells stained (Oligo scoring I)	IHC	0.00	0.09	0.00	0.03	0.09
Staining intensity (Oligo scoring I)	IHC	0.14	0.21	0.02	0.02	0.21
Number of cells stained (Oligo scoring II)	IHC	0.14	0.54	0.02	0.03	0.39
Staining intensity (Oligo scoring II)	IHC	0.18	0.36	0.02	0.02	0.30
Normal blood vessel staining	IHC	0.25	0.48	0.48	0.81	0.39
ECM staining intensity	IHC	0.21	0.30	0.23	0.43	0.33
Peri-necrotic staining	IHC	0.00	0.06	0.00	0.16	0.31
Microvascular proliferation staining	IHC	0.36	0.15	0.02	0.42	0.51
Perivascular staining	IHC	0.07	0.09	0.00	0.31	0.46
Microvascular proliferation present	Histopathological	0.57	0.30	0.02	0.86	0.55
Necrosis present	Histopathological	0.14	0.24	0.02	0.69	0.40
Resection	Clinical	0.93	0.79	0.30	0.73	0.72
KPS at diagnosis	Clinical	0.54	0.52	0.45	0.39	0.79
Age (70 vs > 70 years)	Clinical	0.04	0.03	0.02	0.25	0.14

Ordinal variables were collapsed into 0 vs 1 or more, and continuous variables were dichotomized at clinically relevant cut points. The prevalence of each immunohistochemical feature by tumor diagnosis, and overall prevalence. Overall, 28 patients were diagnosed as having anaplastic oligodendroglioma, 33 as having astrocytic oligoastrocytoma, 44 as having astrocytic astrocytoma, and 101 as having glioblastoma. Among those with glioblastoma, older people have a worse prognosis. Those variables in boldface were the ones with nonzero γ_{Cj} .

AO, anaplastic oligodendroglioma; AOA, astrocytic oligoastrocytoma; AA, astrocytic astrocytoma; GBM, glioblastoma; IHC, immunohistochemical; ECM, extracellular matrix; KPS, Karnofsky performance status.

Table II

Simulation study of two-class unpenalized, NLS and ridge model estimates (standard error) for right-censored survival data.

Parameter	$N = 200, J = 50, \tilde{J} = 10$			$N = 1000, J = 10, \tilde{J} = 10$		
	Unpenalized	NLS	Ridge	Unpenalized	NLS	Ridge
$\beta_{1,1}$	18.94 (47.82)	1.14(0.66)	1.01 (0.33)	1.45 (0.20)	1.47 (0.17)	1.48 (0.11)
$\beta_{1,2}$	19.36 (59.12)	1.04 (0.59)	1.04 (0.39)	1.48 (0.22)	1.44 (0.18)	1.49 (0.08)
$\beta_{1,3}$	17.98 (63.82)	0.93 (0.53)	0.96 (0.38)	1.51 (0.22)	1.50 (0.16)	1.50 (0.10)
$\beta_{1,4}$	12.12(52.66)	0.99 (0.66)	0.94 (0.29)	1.50 (0.22)	1.49 (0.22)	1.50 (0.11)
$\beta_{1,5}$	21.19 (56.34)	0.99 (0.59)	0.98 (0.39)	1.57 (0.21)	1.54 (0.24)	1.52 (0.12)
$\beta_{1,6}$	15.37 (60.22)	1.08 (0.68)	0.95 (0.32)	1.54 (0.20)	1.52(0.18)	1.51 (0.10)
$\beta_{1,7}$	9.84 (73.98)	1.13 (0.63)	0.97 (0.35)	1.54 (0.17)	1.51 (0.13)	1.51 (0.08)
$\beta_{1,8}$	11.84 (53.05)	1.21 (0.68)	0.99 (0.35)	1.49 (0.16)	1.45 (0.19)	1.49 (0.10)
$\beta_{1,9}$	17.98 (55.33)	1.19 (0.63)	1.04 (0.39)	1.51 (0.22)	1.55 (0.22)	1.50 (0.10)
$\beta_{1,10}$	11.93 (55.79)	1.08 (0.65)	1.05 (0.39)	1.45 (0.27)	1.42 (0.17)	1.48 (0.09)
ν	-0.53 (13.19)	-0.37 (1.91)	-0.75 (0.73)	-0.49 (0.15)	-0.49 (0.11)	-0.49 (0.11)
η_2	0.13	0.35	0.40	0.35	0.36	0.36
$\text{Min}(\beta_1 \in J - \tilde{J})$	-60.73	-0.09	0.02	-	-	-
$\text{Max}(\beta_1 \in J - \tilde{J})$	75.76	0.10	0.23	-	-	-
$\text{Mean}(\beta_1 \in J - \tilde{J})$	-5.95	0.01	0.15	-	-	-
$\text{Mean}(\text{SE}(\beta_1 \in J - \tilde{J}))$	69.17	0.18	0.28	-	-	-
P_K	-	132.11	209.25	316.48	563.58	564.27
DIC_5	-	14 228.54	14 339.27	17 592.99	17 839.19	17 820.08
AUC	0.511	0.826	0.901	0.952	0.952	0.952
rMSE	60.23	1.49	2.01	0.53	0.53	0.52

Posterior means are shown (SE) for the parameters of interest, where $\beta_1 j$ denotes the non-intercept inverse logits of the class-specific probabilities, ν denotes the coefficient from the proportional hazards model, and η_2 denotes prevalence of latent class 2. The posterior maximum, minimum, mean and mean standard error of non-discerning parameters are also calculated, where $\beta_1 \in J - \tilde{J}$ denotes the set of non-discerning β coefficients. The effective number of parameters, P_K , DIC_5 , AUC and rMSE are calculated for each model as described in the text.

NLS, nonlinear shrinkage; DIC_5 , deviance information criterion; AUC, area under the curve; rMSE, root mean square error.

Table III

Simulation study of two-class unpenalized, NLS and ridge model estimates (standard error) for right-censored survival data.

Parameter	$N = 200, J = 50, \tilde{J} = 20$			$N = 200, J = 80, \tilde{J} = 20$		
	Unpenalized	NLS	Ridge	Unpenalized	NLS	Ridge
$\beta_{1,1}$	1.57 (0.36)	1.47 (0.33)	1.35 (0.25)	6.72 (18.81)	1.30 (0.44)	1.19 (0.31)
$\beta_{1,2}$	1.61 (0.46)	1.42 (0.43)	1.38 (0.33)	10.54 (32.46)	1.39 (0.33)	1.27 (0.27)
$\beta_{1,3}$	1.56 (0.45)	1.23 (0.40)	1.36 (0.33)	7.20 (18.20)	1.36 (0.36)	1.23 (0.34)
$\beta_{1,4}$	1.46 (0.32)	1.46 (0.42)	1.29 (0.24)	1.85 (2.67)	1.48 (0.35)	1.25 (0.25)
$\beta_{1,5}$	1.58 (0.40)	1.33 (0.23)	1.36 (0.26)	1.80 (6.36)	1.35 (0.31)	1.26 (0.26)
$\beta_{1,6}$	1.55 (0.42)	1.44 (0.48)	1.35 (0.30)	5.80 (16.13)	1.40 (0.28)	1.25 (0.29)
$\beta_{1,7}$	1.49 (0.42)	1.37 (0.36)	1.32 (0.33)	4.46 (12.13)	1.36 (0.39)	1.32 (0.27)
$\beta_{1,8}$	1.52 (0.37)	1.45 (0.47)	1.33 (0.28)	2.94 (5.22)	1.41 (0.27)	1.35 (0.26)
$\beta_{1,9}$	1.69 (0.49)	1.45 (0.35)	1.42 (0.36)	3.97 (8.87)	1.44 (0.37)	1.20 (0.27)
$\beta_{1,10}$	1.70 (0.42)	1.43 (0.36)	1.46 (0.31)	2.45 (3.09)	1.33 (0.45)	1.28 (0.29)
$\beta_{1,11}$	1.43 (0.33)	1.54 (0.42)	1.28 (0.25)	6.46 (15.15)	1.35 (0.42)	1.33 (0.30)
$\beta_{1,12}$	1.53 (0.35)	1.41 (0.40)	1.35 (0.25)	3.17 (8.18)	1.35 (0.33)	1.30 (0.25)
$\beta_{1,13}$	1.90 (1.52)	1.41 (0.34)	1.54 (0.38)	5.38 (17.60)	1.38 (0.36)	1.20 (0.33)
$\beta_{1,14}$	4.83 (14.77)	1.44 (0.37)	1.39 (0.33)	5.18 (16.61)	1.49 (0.49)	1.31 (0.33)
$\beta_{1,15}$	1.73 (0.54)	1.36 (0.44)	1.42 (0.35)	2.80 (6.03)	1.46 (0.36)	1.41 (0.32)
$\beta_{1,16}$	4.48 (15.44)	1.51 (0.39)	1.28 (0.37)	7.02 (15.89)	1.46 (0.42)	1.30 (0.25)
$\beta_{1,17}$	1.55 (0.33)	1.32 (0.40)	1.35 (0.23)	6.56 (17.98)	1.24 (0.29)	1.34 (0.27)
$\beta_{1,18}$	1.74 (0.48)	1.48 (0.35)	1.46 (0.36)	2.98 (15.36)	1.37 (0.37)	1.30 (0.29)
$\beta_{1,19}$	1.55 (0.44)	1.34 (0.33)	1.35 (0.33)	7.11 (17.88)	1.41 (0.39)	1.32 (0.21)
$\beta_{1,20}$	1.55 (0.40)	1.49 (0.34)	1.35 (0.26)	6.02 (17.01)	1.27 (0.27)	1.29 (0.25)
ν	-0.50 (0.28)	-0.54 (0.20)	-0.50 (0.20)	-1.49 (5.23)	-0.48 (0.21)	-0.50 (0.19)
η_2	0.31	0.34	0.34	0.29	0.36	0.36
$\text{Min}(\beta_1 \in J - \tilde{J})$	-2.01	-0.12	-0.03	-5.37	-0.09	-0.06
$\text{Max}(\beta_1 \in J - \tilde{J})$	0.86	0.071	0.27	3.80	0.06	0.16
$\text{Mean}(\beta_1 \in J - \tilde{J})$	-0.10	-0.00	0.12	-1.5	-0.01	0.07
$\text{Mean}(\text{SE}(\beta_1 \in J - \tilde{J}))$	0.38	0.17	0.30	11.37	0.18	0.28
P_K	—	67.41	61.26	—	53.63	60.52
DIC_5	—	14 080.91	14 062.48	—	21 986.74	21 998.93
AUC	0.981	0.987	0.985	0.929	0.986	0.984
rMSE	4.58	0.77	1.17	642.28	0.93	1.35

NLS, nonlinear shrinkage; DIC₅, deviance information criterion; AUC, area under the curve; rMSE, root mean square error.

Table IV

Simulation study of two-class unpenalized, NLS and ridge estimates (standard error) for uncensored survival data.

Parameter	$N = 100, J = 50, \tilde{J} = 20$			$N = 100, J = 80, \tilde{J} = 20$		
	Unpenalized	NLS	Ridge	Unpenalized	NLS	Ridge
$\beta_{1,1}$	19.46 (68.87)	1.38 (0.44)	1.47 (0.39)	14.28 (25.44)	1.11 (0.52)	1.12 (0.32)
$\beta_{1,2}$	31.56 (51.28)	1.33 (0.67)	1.42 (0.39)	18.76 (23.13)	1.15 (0.63)	1.12 (0.36)
$\beta_{1,3}$	35.61 (59.95)	1.17 (0.55)	1.23 (0.38)	15.42 (23.58)	1.13 (0.57)	1.14 (0.32)
$\beta_{1,4}$	31.28 (49.71)	1.28 (0.66)	1.46 (0.38)	11.11 (24.22)	1.32 (0.66)	1.06 (0.35)
$\beta_{1,5}$	36.52 (59.17)	1.14 (0.49)	1.33 (0.35)	13.01 (23.63)	0.98 (0.56)	1.09 (0.43)
$\beta_{1,6}$	30.68 (58.62)	1.27 (0.50)	1.44 (0.43)	11.90 (25.76)	1.05 (0.58)	1.13 (0.46)
$\beta_{1,7}$	33.63 (57.10)	1.36 (0.67)	1.37 (0.33)	19.15 (20.53)	1.06 (0.63)	1.10 (0.32)
$\beta_{1,8}$	27.19 (65.87)	1.38 (0.55)	1.45 (0.41)	10.74 (25.74)	0.95 (0.57)	1.11 (0.35)
$\beta_{1,9}$	30.85 (64.62)	1.13 (0.68)	1.45 (0.32)	15.12 (24.11)	1.12 (0.66)	1.14 (0.27)
$\beta_{1,10}$	38.98 (52.11)	1.33 (0.66)	1.43 (0.41)	11.79 (24.32)	1.21 (0.53)	1.23 (0.36)
$\beta_{1,11}$	53.98 (68.18)	1.28 (0.57)	1.54 (0.48)	14.76 (24.57)	1.17 (0.65)	1.04 (0.36)
$\beta_{1,12}$	35.11 (55.17)	1.17 (0.61)	1.41 (0.37)	10.97 (26.11)	1.09 (0.49)	0.98 (0.33)
$\beta_{1,13}$	28.89 (65.77)	1.20 (0.68)	1.41 (0.33)	11.18 (24.64)	1.20 (0.64)	1.06 (0.31)
$\beta_{1,14}$	33.46 (56.42)	1.33 (0.62)	1.44 (0.37)	10.89 (25.01)	1.13 (0.54)	1.02 (0.33)
$\beta_{1,15}$	34.13 (62.61)	1.48 (0.64)	1.36 (0.30)	10.88 (26.00)	1.06 (0.55)	1.02 (0.22)
$\beta_{1,16}$	43.96 (49.40)	1.15 (0.52)	1.51 (0.30)	13.97 (24.33)	1.07 (0.62)	1.06 (0.36)
$\beta_{1,17}$	35.03 (52.58)	1.26 (0.68)	1.32 (0.36)	15.32 (22.95)	1.13 (0.65)	1.12 (0.35)
$\beta_{1,18}$	32.74 (59.93)	1.33 (0.65)	1.48 (0.34)	16.96 (23.82)	1.23 (0.42)	1.14 (0.36)
$\beta_{1,19}$	15.87 (49.21)	1.20 (0.48)	1.34 (0.24)	13.52 (24.61)	1.22 (0.72)	1.09 (0.44)
$\beta_{1,20}$	38.86 (62.98)	1.37 (0.48)	1.49 (0.46)	15.74 (21.15)	1.17 (0.54)	1.08 (0.36)
ν	-4.00 (11.65)	-0.43 (0.29)	-0.53 (0.25)	-3.06 (5.61)	-0.26 (1.28)	-0.50 (0.37)
η_2	0.12	0.34	0.35	0.05	0.36	0.36
$\text{Mean}(\beta_1 \in J - \tilde{J})$	-10.58	-0.01	0.16	-5.51	-0.01	0.14
PK	–	38.21	48.10	–	45.83	64.43
DIC ₅	–	7129.71	7139.39	–	11963.84	11896.82
AUC	0.579	0.958	0.972	0.651	0.980	0.982

NLS, nonlinear shrinkage; DIC₅, deviance information criterion; AUC, area under the curve.

Table V

Area under curve and weighted logrank test statistic comparing the NLS with the two-stage method.

Simulation	AUC		Wald Test	
	NLS	Bair–first stage	NLS	Two–stage model
$N = 200, J = 80, \tilde{J} = 20$	0.989	0.629	7.06	0.67
$N = 200, J = 50, \tilde{J} = 20$	0.974	0.638	8.69	3.83
$N = 200, J = 50, \tilde{J} = 10$	0.933	0.596	7.15	0.68
$N = 100, J = 80, \tilde{J} = 20$ (no cens)	0.933	0.581	5.05	0.21
$N = 100, J = 50, \tilde{J} = 20$ (no cens)	0.937	0.586	3.34	2.45

The Wald test statistic is calculated for the survival models including latent class as a predictor, where latent class is weighted by posterior probabilities of class membership.

N , number of subjects; J , number of variables; $\tilde{J}$, number of variables associated with latent class.

Table VI

Deviance for K-class CD models fit to the glioma data.

	Larsen	NLS	Ridge
Two-class	4846.24	4719.00	4720.56
Three-class	98 008.46	4527.63	5180.34

the smallest value of the model Deviance determines the best model fit.