A nonparameteric Bayesian classification technique to predict disease status using genomic data

Yuxin Zhang

Abstract

Predicting the state of disease is one of the most important problems in biostatistics research areas. Regression and classification are used to be the most common methods solving these problems.[1] Our research target, Naive Bayes classification is also a well-known classification technique. However, the disadvantages of a traditional NB classifier will be exposed when the number of features is huge and the assocation between feature and class variable has not be determined because the two main assumptions of a NB classifier are the distributions of features has to be known and no correlation between features. This paper explores a non-parametric Bayes classification framework for predicting and imputing the class variable by training distributions from the given data. We apply this new method to a inflammatory bowel disease data and compare our method with multinomial regression models, another traditional method for classification problems. Our results show that the transformed NB classification method can perform accurate and robust predictions, even higher than traditional methods.

1. Introduction

Bayes theorem[2] is widely used in dealing with data problems. The two main applications of Bayes theorem are classification and data imputation. In a classification problem with categorical features, Bayes theorem can be flexibily applied to find the posterior distribution given features' value based on the prior distribution of features and the conditional distribution of each feature at each response variable's class.

There are three main challenges in applying traditional Naive Bayes classification method on a genomic data.[3] One of the challenges is the gene expression data is used to be high-dimensional. Computing the posterior probability requires multiplying the number of (conditional) density equal to the number of features which is very likely to result in a value extremely closes to 0, far beyond the ability of ordinary calculation tools. Secondly, the distribution of gene features is continous in most cases and the value of each data point are distinct which makes it impossible to train a conditional probability density model from data. However, if the correlation between some features exists, the joint density can not be correctly estimated without conditonal distribution unless the independence between features are guaranteed. For the above two problem, dimension reduction is a proposed resolution. The final challenge is the unpredictability of a gene expression data distribution. For high-dimensional Gaussian distribution, classification can be done by Linear Discriminant Analysis[4] method. However, the gene distribution may distributed without any pattern or does not depend on any parameter. This inspires us to find a way simulating the distribution of a variable based on the samples only.

Our new method first uses the dimension reduction method: selecting a combination of features meet the independence assumption. One way is computing the sample correlations between features and then searching for a sequence of uncorrelated features. An alternative way is performing a basis transformation. After the selection step, kernel density estimation is applied to simulate the univariate density function. We also propose a new approach to choose the optimized bandwidth through the comparison between results from KDE and bootstrap method.

Here is a brief overview of the sections to follow. Section 2 introduces and describes the whole framework of the proposed method. Section 3 evaluates the method on a real genomic data. Section 4 discusses some concerns and limitations about the method.

2. Methods

In this part, a nonparametric Baysian classifier for continous features is described. The order of sections follow the actual operations.

2.1 Data Pre-processing

First of all, the gene expression data can be normalized[5] according to personal demands and perferences. Observations with missing values are not recommended to be imputed when sample size is large unless there are few observations because NB performs better when more training samples are given.

2.2 Feature Selection Methods

One challenge of nonparametric Bayes classification on continous data is that training a conditional density is unfeasible. The unique way to estimate the joint density is multiplying a series of estimated univariate probability density training from the given information. Therefore, a nonparametric Bayes classifier requires that the presense of a particular feature in a class is uncorrelated to the presence of any other feature. (more explanation and proof in next section) For genomic data, the existence of correlation between genes may greatly reduce the classification accuracy. This section discusses some feasible methods to eliminate the correlation between features or select a subset of features which are mutually independent.

2.2.1 Correlation-based feature selection

Pearson's Correlation is a measure of how strong the linear association is between two variables. Pearson' R ranges from -1 to 1 and its absoluate value indicates the strength of the association. The absolute value of coefficient is expected to be larger for a pair of highly correlated features. Pearson's sample correlation coefficient estimates the population correlation.[6] Pearson's sample correlation of a paired data (x_i,y_i), i = 1,. . . ,n can be computed by the formula below:

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

For a genomic data with more than two features, a correlation matrix can be generated and each entry represents a pairwise correlation. A set of mutually independent features can be selected based on the values in the correlation matrix. A workable selection algorithm is shown in the flow diagram (Figure 1):

2.2.2 Principal Component

An alternative method is performing a change of basis. Principal Component Analysis[7] is a technique to project a multidimensional data to a lower dimension space with minimizing information loss. The transformed new basis are proved to be uncorrelated. For a n << k case, PCA is a useful tool to seek a small number of uncorrelated features but still explaning most of the data variability. The first principal component is the one explaining the most variance and it is defined as (X is the feature matrix):

$$Maximize \;\; w_1 X^T X w_1^T \qquad subject\ to: \;\; w_1^T w_1 = 1$$

And the second principal component can be found by the same process:

$$Maximize \;\; w_2 X^T X w_2^T \qquad subject\ to: \;\; w_2^T w_2 = 1 \; and \; w_1^T w_2 = 0$$

It turns out that the principal components are generated from the eigenvectors of $X^T X$ (covaraince matrix). The first principal component is the generated from the eigenvector with the largest eigenvalue.

2.3 Nonparametric Bayes Classifier

We now derive a nonparametric Bayes classifer used for a multi-class classification problem. The classifier's decision is based on the posterior probability of each class given the feature values. The predicted class should be the one with highest posterior probability. Now suppose there is a complete gene expression data with k independent continous features after the feature selection process and n independent samples is

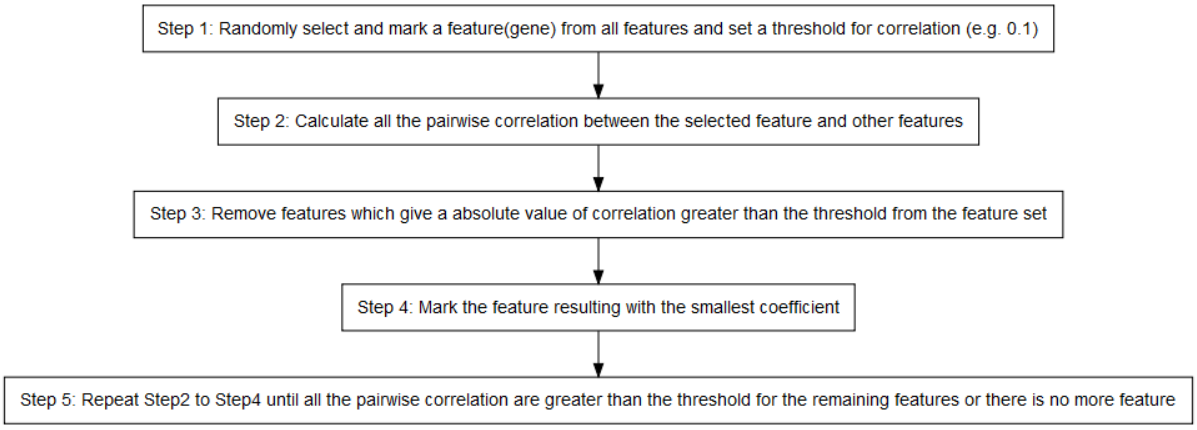


Figure 1: An Algorithm for Correlation-based Selection

represented by the following matrix:

$$X = \begin{pmatrix} x_{1,1} & \dots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,k} \end{pmatrix}, \quad Y = (y_1, \dots y_n)^T$$

The class variable is a class label with m classes.(e.g. disease status) According to Bayes' Theorem, the probability that given a new observation $x_{n+1} = (x_{n+1,1}, x_{n+1,2}, \dots x_{n+1,k})$ belongs to class j is equivalent to:

$$\begin{aligned} P(Y = j | X = x_{n+1}) &= \frac{f_{X_1, X_2, \dots X_k | Y=j}(x_{n+1}) \pi_j}{\sum_{c=1}^{m} f_{X_1, X_2, \dots X_k | Y=c}(x_{n+1}) \pi_c} \\ &= \frac{\prod_{i=1}^{k} f_{X_i | Y=j}(X_i = x_{n+1,i}) \pi_j}{\sum_{c=1}^{m} \prod_{i=1}^{k} f_{X_i | Y=c}(X_i = x_{n+1,i}) \pi_c} \end{aligned} \tag{2}$$

(proof of (2) in APPENDIX) where $\hat{\pi}_j = \sum_{i=1}^{n} 1(y_i = j)/n$ can be used to estimate π_j[8]. To estimate $f_{X_i|Y=j}$, by the fact that probability density function is the derivative of cumulative density function and empirical distribution estimates the true underlying distribution[9],

$$\hat{f}_{X_i | Y=j}(X_i = x_{n+1,i}) = \frac{1}{2n\hat{\pi}_j h} \sum_{r=1}^{n} 1(x_{n+1,i} - h \leq x_{r,i} \leq x_{n+1,i} + h) 1(y_r = j) \tag{3}$$

for some small h. (proof of (3) in APPENDIX) Therefore, our predicted class of a new observation x_{n+1} is the one with maximum probability:

$$\hat{class}(x_{n+1}) = \underset{j \in \{1, \dots, m\}}{\operatorname{argmax}} \hat{\pi}_j \prod_{i=1}^{k} \left(\frac{1}{2n\hat{\pi}_j h} \sum_{r=1}^{n} 1(x_{n+1,i} - h \leq x_{r,i} \leq x_{n+1,i} + h) 1(y_r = j) \right) \tag{4}$$

2.4 Kernel and Bandwidth Selection

smoothness and kernel density estimation
Recall from (3), the equation can be expressed as:

$$\frac{1}{2n\hat{\pi}_j h}\sum_{r=1}^{n} 1(x_{n+1,i}-h\le x_{r,i}\le x_{n+1,i}+h)1(y_r=j)=\frac{1}{n\hat{\pi}_j h}\sum_{r=1}^{n} K(\frac{x_{n+1,i}-x_{r,i}}{h})1(y_r=j) \tag{5}$$

where $K(\frac{x_{n+1,i}-x_{r,i}}{h})\sim Unif(-1,1)$ if we assumes all data points are having equal weight.(Proof of (5) in APPENDIX) To better depict a distribution with smooth curve(s), closer points should have larger weights than further points, uniform kernel is replaced by a normal kernel.[10] (i.e. $K(\frac{x_{n+1,i}-x_{r,i}}{h})\sim N(0,1)$)

bandwidth selection
Two proposed bandwidths[11] for a normal kernel are:

Scott's bandwidth[12]: $h_1 = 1.06n^{-0.2}\hat{\sigma}$
Silversman's bandwidth[13]: $h_2 = 0.9min(\hat{\sigma}, IQR/1.35)n^{-0.2}$

Since genomic data has a large number of features with different distributions. Therefore, there can be a more optimized selection method than choosing the same bandwidth for all features during the KDE step. We introduce an algorithm that selects the bandwidth with the better performance (a 'smart' bandwidth) from the proposed ones for each feature. Performance is measured by a loss function. Split the variable domain into several intervals. The expected setting of loss function is the sum of the absolute values of difference between the probability obtained by KDE with selected bandwidth and the actual probability in each interval. Since the actual distribution is unknown, we can estimate the partial probability by boostrap method.[14] The specific 'smart' bandwidth selection procedures are shown as follows:

Step 1: Partition the variable domain into N disjoint intervals (e.g. $\cup_{i=1}^{N} I_i=(-\infty,\infty)\ and\ I_j\cap I_i=\varnothing,\ \forall\, i,j$), each interval covers the region where the data points is distributed. Split the training data X_k (suppose this is the data for kth feature) into two parts with equal or similar size: $X_{k,1}, X_{k,2}$.
Step 2: Start with the first interval, use the normal kernel to estimate the probability that $x_{n+1,k}$ falls within the interval based on the data $X_{k,1}$ when bandwidth is equal to h_1 and h_2 respectively. The results are denoted as $\hat{p}_{1,h_1}$ and $\hat{p}_{1,h_2}$.
Step 3: Draw the number of samples equal to $X_{k,2}$ from $X_{k,2}$ with replacement. Calculate the proportion of samples falls within the interval. Repeat this step B times (e.g. 1000) and average the results as $\hat{p}_{1,boot}$.
Step 4: Repeat Step 2 and Step 3 until all the intervals are processed. The smart bandwidth h_k^* for the kth feature, is the one minimizing the loss equation:

$$h_k^* = \underset{h\in\{h_1,h_2\}}{\operatorname{argmin}} \sum_{i=1}^{N} \mid \hat{p}_{i,boot} \ - \ \hat{p}_{i,h} \mid \tag{6}$$

Finally, our final nonparametric Bayes classifier can be reexpressed as

$$\hat{class}(x_{n+1}) = \underset{j\in\{1,\ldots,m\}}{\operatorname{argmax}} \ \hat{\pi}_j \prod_{i=1}^{k}(\frac{1}{n\hat{\pi}_j h_i^*}\sum_{r=1}^{n} K(\frac{x_{n+1,i}-x_{r,i}}{h_i^*})1(y_r=j)) \tag{7}$$

3. Application and Results

To demonstrate the performance of nonparametric Bayes classfication technique on real genomic data, we evaluated the algorithm on an inflammatory bowel disease dataset. The dataset contains 126 individuals with their disease status and features. (i.e. gene expressions) The evalution method is training an nonparametric Bayes classifier from training dataset in order to predict the patient's disease status.

3.1 Data Manipulation

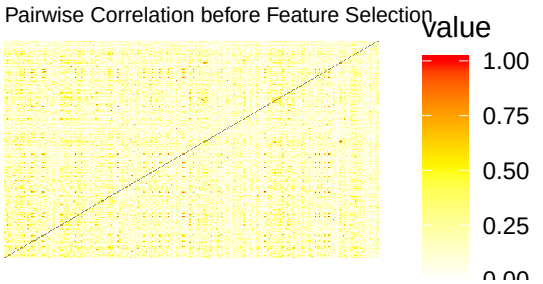
The feature dataset is a collection of expression levels of 309 probesets/genes from the 126 individuals. We obtained the data from a publically available download website at Statistical Society of Canada. (https://ssc.ca/en/meeting/annual/2017/case-study-2) Class label is the disease status with three different categories. 41 out of 126 samples are in healthy status. The rest of them are inflammatory bowel disease patients, 59 of which are having an Crohn's disease and the remaining 26 are Ulcerative Colitis patients. Two-thirds of the observations are being used for training purposes. Dataset is divided into three subgroups according to their disease status. A stratified sampling method is applied and the overall training samples consists of 2/3 of members from each subgroup.

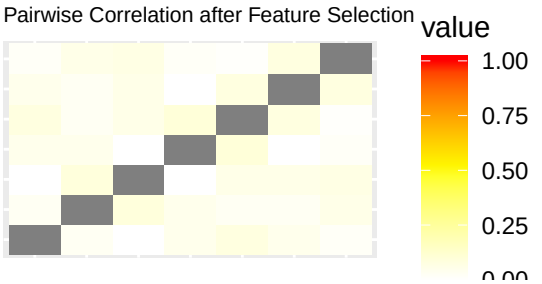
3.2 Feature Selection Results

The objective of feature selection process is reducing the correlation between features to conform the method's assumption.

Correlation-based Selection Result

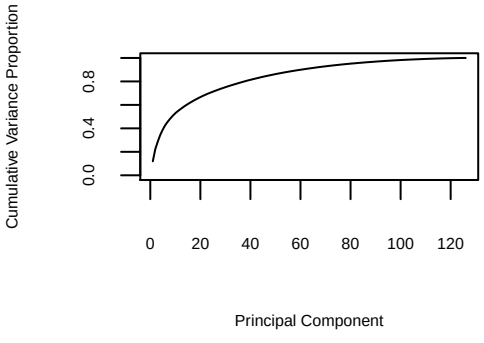
Heatmap helps to visulize the sample correlation between each pair of variables. (See Section 2.2 for the specific formula) The direction of association (i.e. the sign of correlation coefficient) is not in the consideration, only the strength of assocation. Color changes from white to red as the strength of association increases. The diagonal blocks are grey because the relation between a variable and itself is not meaningful. It can be shown that the data after processing feature selection reduces the internal correlation significantly.

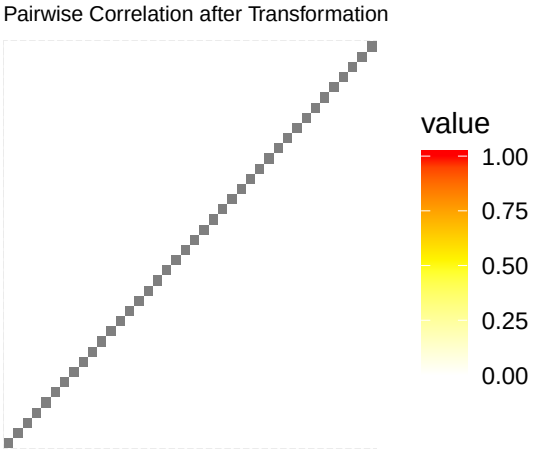




Principal Component Analysis

We finally selected the first 40 principal components (sorted by their eigenvalues) which have explained 80% of the total data varibility. After generating the heatmap, the transformed features are perfectly uncorrelated as expected.





3.3 Classification Evalutaion

We define the classification accuracy measure to be the proportion of test samples correctly classified:

$$classification\ accuracy\ =\ \frac{\sum_{i=1}^{n} 1(class_{predicted} = class_{actual})}{n} \tag{8}$$

Prediction results from nonparametric Bayes classifier using PC features

The first 40 principal components obtained in section 3.2 are utilized to train the density function for each pc. Simulation results of the first principal component are shown in the figures below. PC1 samples for each disease status are shown in the histograms. Figure 2 are the estimated densities learned from the training samples.

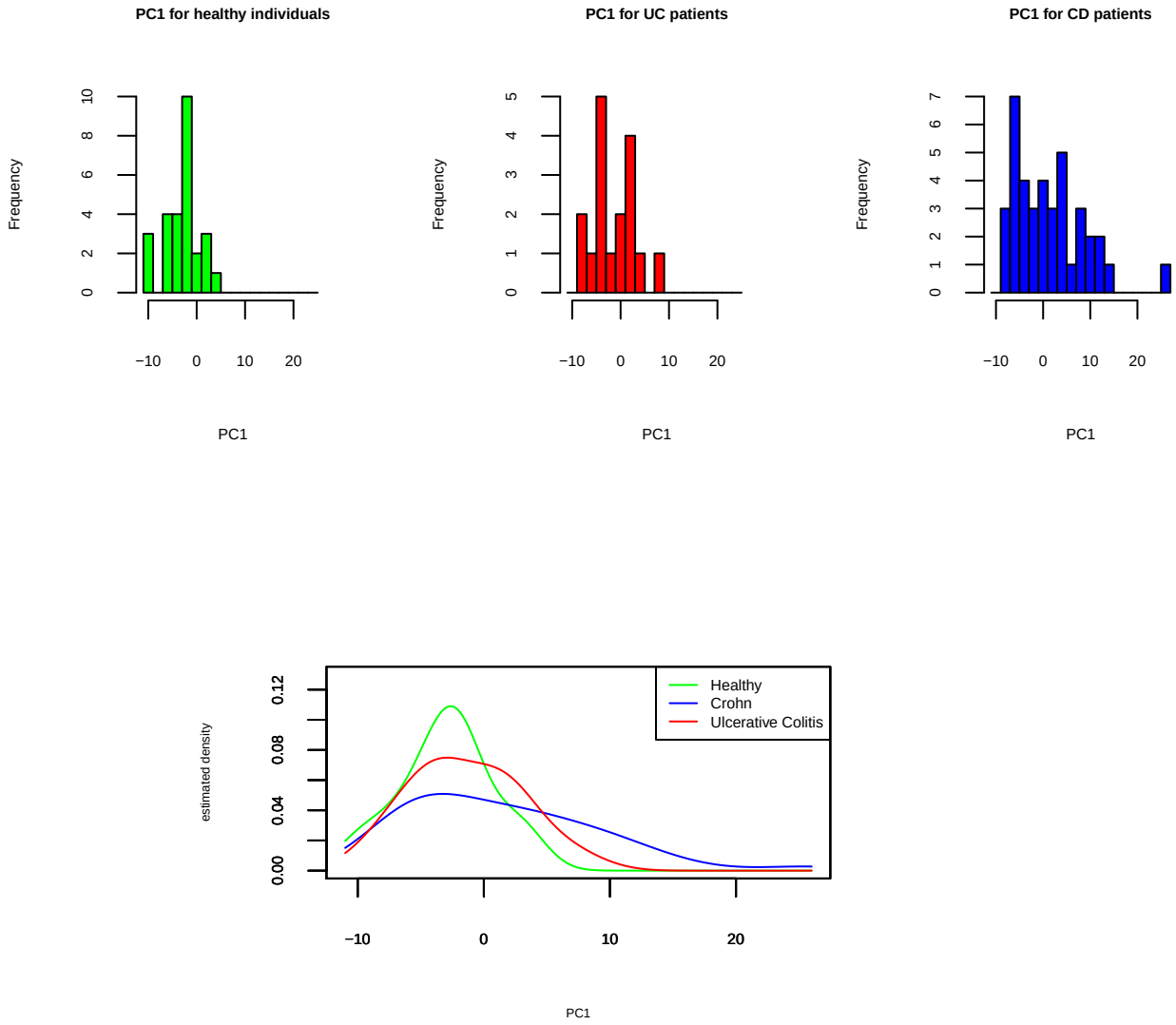


Figure 2: estimated density function of PC1

Nonparametric NB classifier makes prediction based on the features of test samples. For this split, the prediction results are shown in table 1:

Table 1		Actual Status		
		Normal	Crohn's Disease	Ulcerative Colitis
Predicted Status	Normal	14	4	2
	Crohn's Disease	0	13	2
	Ulcerative Colitis	0	3	5

Prediction Accuracy = (14+13+5)/43 = 0.744

Prediction results using features selected from Correlation-based algorithm

In this split, seven features (probeset ID:200779_at, 206072_at, 206336_at, 206485_at, 207257_at, 209664_x_at, 210171_s_at) are selected in Section 3.2. Figure 3 shows the estimated density of 206336_at and figure 4 shows the estimated density of 206072_at.

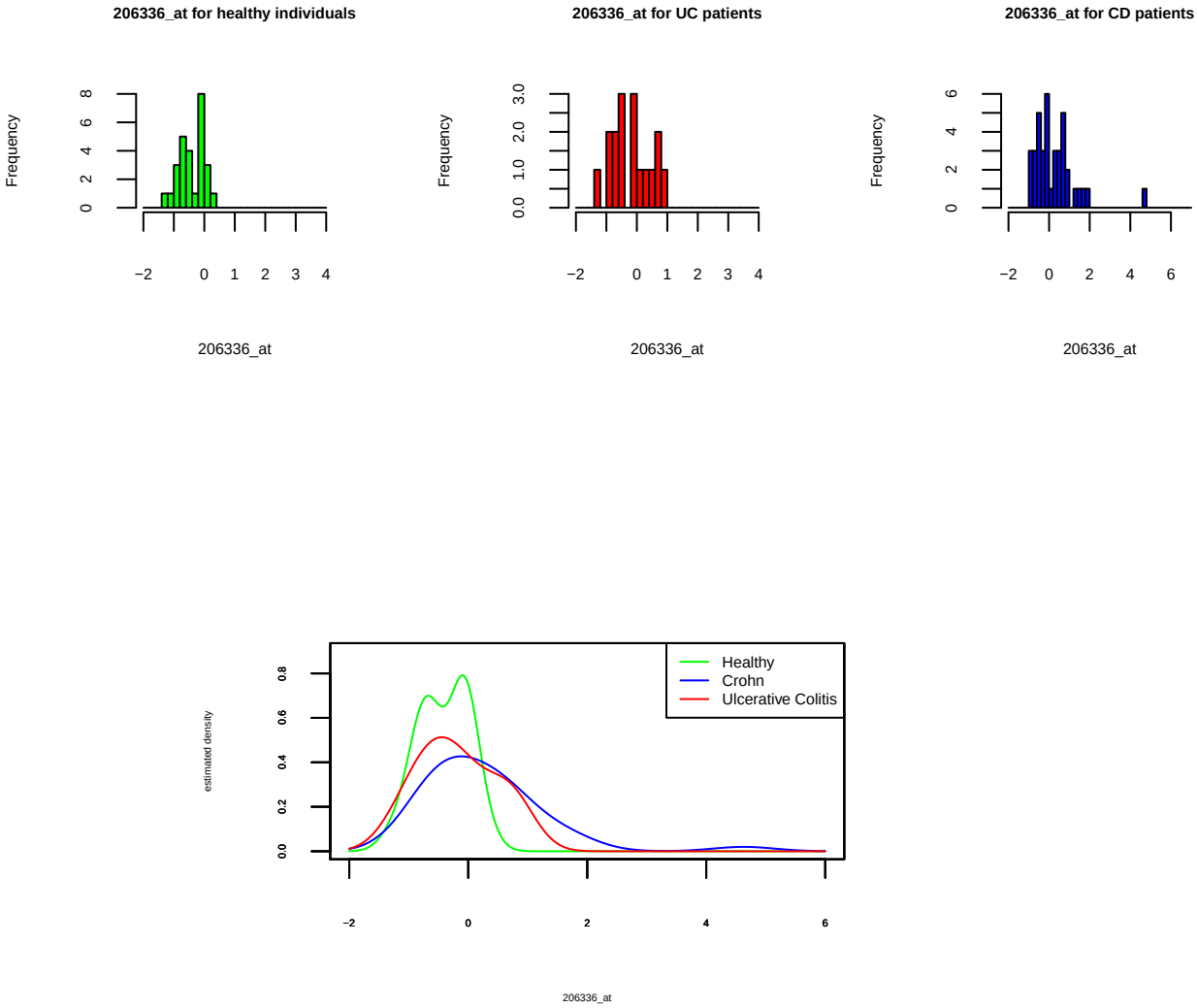
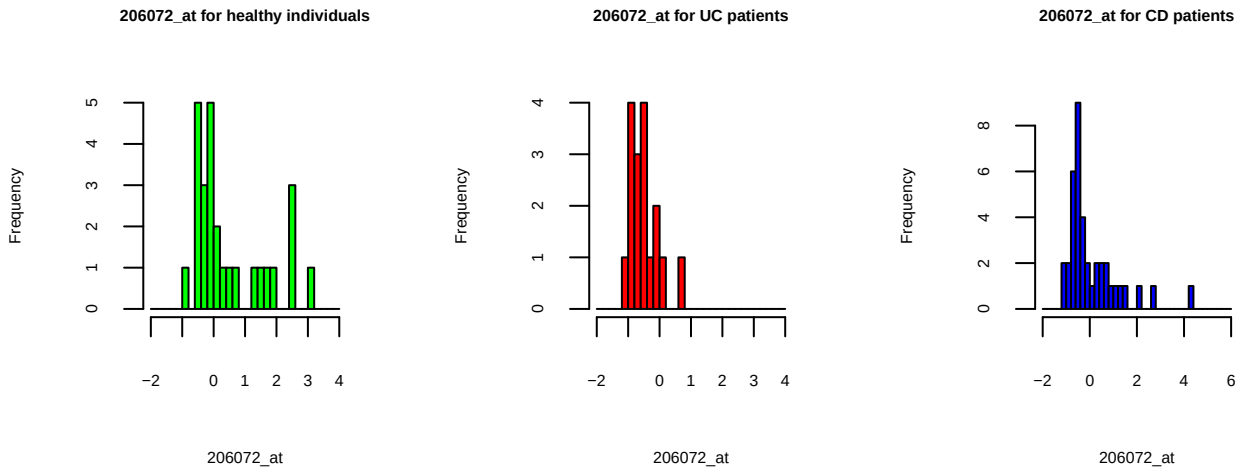


Figure 3: estimated density of 206336



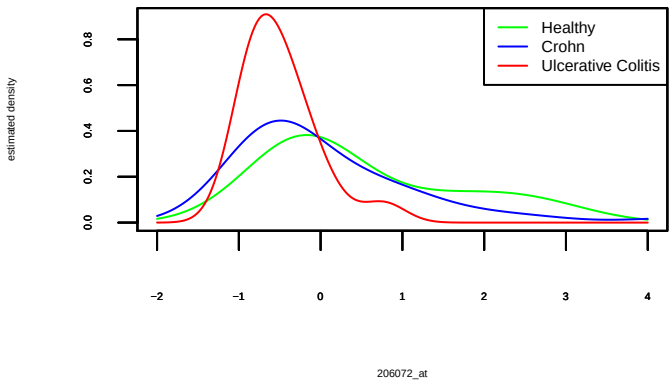


Figure 4: estimated density of 206072

Prediction result in this split:

Table 2		Actual		
		Normal	Crohn's Disease	Ulcerative Colitis
Predicted	Normal	6	6	2
	Crohn's Disease	6	11	4
	Ulcerative Colitis	2	3	3

Prediction Accuracy = (6+11+3)/43 = 0.4418

As shown in the figures, the density function generated by nonparametric NB methodology looks very similar to the real distribution in both shape and density. In terms of prediction accuracy, pc performs much better

than correlated-based method. (but still greater than the random probability when there are three classes, 33%) Exactly, it is true that the difference between UC and CD in many genes are not that obvious.

3.4 Comparing Methods

Our reference method is a multinomial regression method. Since the proportional odds assumption can not be justified, we choose the baseline category logit model.[15] Our baseline category is normal status. Structure of multinomial logit models:

$$log(\frac{\pi_c}{\pi_n}) = \alpha_1 \ + \beta_1^T x \ and \ log(\frac{\pi_u}{\pi_n}) = \alpha_2 \ + \beta_2^T x \tag{9}$$

where x is a k-length feature vector and β is a k-length coefficient, predicted status is the one with maximum estimated probability. (i.e.$\underset{j=n,c,u}{\text{argmax}}\ \hat{\pi}_j$) We apply both regression model and NB classifier to train from the same training samples and make prediction, and repeat 1000 times. The dataset is splitted in a different way in each iteration. The average accuracy is shown in the table below:

Method	Feature Selection Method	Classification Accuracy
nonparametric Bayes	Principal component	0.670930
nonparametric Bayes	Correlation-based	0.497674
multinomial regression	Principal component	0.651163
multinomial regression	Correltaion-based	0.503488

Nonparametric Bayes classification still maintains the advantage of the traditional NB. As long as the features are mutually unassociated, the prediction accuracy from nonparametric Bayes would be very high, even higher than regression models. Nevertheless, nonparametric Bayes looks very sensitive to feature correlation. Recall that features selected from the correlated-based selection are not perfectly uncorrelated but having corrlation lower than the threshold. In contrast, the regression method is more flexible especially when there is a multicolliearity issue. The regression model resolves in the way that one of the correlated variables' coefficient shrinks to 0 which doesn't affect the calcuated probability.

IV. Discussion

An optimized correlated-based feature selection method

In section 3.3, results show that the feature set obtained by correlated based selection algorithm is not very suitable for Bayes classification technique due to a relative low prediction accuracy. One conjecture to solve this problem is the value of threshold should be smaller. In our analysis, we set the value of threshold as high as 0.1 which is not a small value. On the other hand, only 7 features are selected at this point. A smaller threshold value is likely to trigger a less number of features so that the key features may be excluded. Constructing a better searching algorithm is also a proposed solution to this problem, such as finding multiple qualified feature combinations but only choosing the combination with the most number of features.

Rationality of the testing method - prior distribution

In section 3.4, even though nonparametric NB demonstrates a more robust and accurate prediction than regression models on this datatset. However, these are all limited to one premise: prior distribution of training set must be similar to the populational prior distribution. NB classifier relies heavily on prior distribution! When the prior probability of one class is high, NB tends to predict that class. Consequently, it is more recommended to use nonparametric Bayes for predicting a small number of observations. Note that the prior distribution of the training samples might be inconsistent with the predicting samples.

REFERENCES

[1] Qianfan Wu, Adel Boueiz, Alican Bozkurt, Arya Masoomi, Allan Wang, Dawn L DeMeo, Scott T Weiss and Weiliang Qiu, Deep Learning Methods for Predicting Disease Status Using Genomic Data, *J Biom Biostat*, 9(5): 417, 2018.

[2] Nicholas Stylianides Eleni Kontou, Bayes Theorem and its recent applications, *Leicester Undergraduate Mathematical Journal*, Vol 2, 2020.

[3] J. Rennie, L. Shih, J. Teevan, and D. Karger, Tackling the Poor Assumptions of Naive Bayes Text Classifiers, *AAAI Press*, 20:616-623, 2003

[4] Tao Li, Shenghuo Zhu and Mitsunori Ogihara, Using discriminant analysis for multi-class classification: an experimental investigation, *Knowledge and Information Systems*, 10: pages453–472, 2006

[5] Xueyan Liu, Nan Li, Sheng Liu, Jun Wang, Ning Zhang, Xubin Zheng, Kwong-Sak Leung, and Lixin Cheng, Normalization Methods for the Analysis of Unbalanced Transcriptome Data, *Front BIoeng Biotechnol*, 7:358, 2019

[6] Donald W.Zimmerman, Bruno D.Zumbo and Richard H. WIlliams, Bias in Estimation and Hypothesis Testing of Correlation, *Psicologica*, 24:133-158, 2003

[7] Lu, Yijuan, Cohen, Ira, Zhou, Xiang Sean, Tian and Qi, Feature selection using principal feature analysis, *ACM Multimedia 2007*, 15:301-304, 2007

[8] Pierre Duchesne, Estimation of a Proportion with Survey Data, *Journal of Statistics Education, Vol.11:3*, 2003

[9] M.S. Waterman and D.E. Whiteman, Estimation of probability densities by empirical density functions, *INT.J.MATH.EDUC.SCI.TECHNOL.*, 9(2): 127-137, 1978

[10] M.C.Jones, The performance of kernel density functions in kernel distribution function estimation, *Statistics & Probability Letters*, 9(2): 129-132, 1990

[11] Mils-Bastian Heidenrich and Stefan Sperlich, Bandwidth Selection in Kernel Density Estimation, *AStA Advances in Statisticl Analysis*, 97(4): 403-433, 2013

[12] Daniel J.Henderson, Christopher F.Parmeter, Normal reference bandwidths for the general order, multivariate kernel density derivative estimator, *Statistics & Probability Letters*, 82(12): 2198-2205, 2012

[13] Shean-Tsong Chiu, Bandwidth Selection for Kernel Density Estimation, *The Annals of Statistics*, 19(4): 1883-1905, 1991

[14] Schuermann, Til; Hanson, Samuel Gregory, Estimating probabilities of default, *Staff Report, No.190, Federal Reserve Bank of New York*, 2004

[15] Courtney Coughenour, Alexander Paz, Hanns de la Fuente-Mella, Ashok Singh, Multinomial logistic regression to estimate and predict perceptions of bicycle and transportation infrastructure in a sprawling metropolitan area, *Journal of Public Health*, 38(4): 401-408, 2015

APPENDIX

PROOF of (2):

$$
\begin{aligned}
P(Y=j|X=x_{n+1}) &= \frac{f_{X_1,X_2,...X_k|Y=j}(x_{n+1})\pi_j}{\sum_{c=1}^{m} f_{X_1,X_2,...X_k|Y=c}(x_{n+1})\pi_c} \\
&= \frac{f_{X_1|X_2=x_{n+1,2},...X_k=x_{n+1,k},Y=j}(X_1=x_{n+1,1})...f_{X_k|Y=j}(X_k=x_{n+1,k})\pi_j}{\sum_{c=1}^{m} f_{X_1|X_2=x_{n+1,2},...X_k=x_{n+1,k},Y=c}(X_1=x_{n+1,1})...f_{X_k|Y=c}(X_k=x_{n+1,k})\pi_c} \\
\textit{under the assumption of independence}, &= \frac{f_{X_1|Y=j}(X_1=x_{n+1,1})f_{X_2|Y=j}(X_2=x_{n+1,2})...f_{X_k|Y=j}(X_k=x_{n+1,k})\pi_j}{\sum_{c=1}^{m} f_{X_1|Y=c}(X_1=x_{n+1,1})f_{X_2|Y=c}(X_2=x_{n+1,2})...f_{X_k|Y=c}(X_k=x_{n+1,k})\pi_c} \\
&= \frac{\prod_{i=1}^{k} f_{X_i|Y=j}(X_i=x_{n+1,i})\pi_j}{\sum_{c=1}^{m}\prod_{i=1}^{k} f_{X_i|Y=c}(X_i=x_{n+1,i})\pi_c}
\end{aligned}
\tag{10}
$$

PROOF of (3):

$$
f_{X_i|Y=j}(X_i=x_{n+1,i}) = F'_{X_i|Y=j}(X_i=x_{n+1,i}) = \lim_{h\to 0^+}\frac{F_{X_i|Y=j}(x_{n+1,i}+h)-F_{X_i|Y=j}(x_{n+1,i}-h)}{2h} \tag{11}
$$

$F_{X_i|Y=j}$ can be estimated by empirical distribution. $\hat{F}_{X_i|Y=j}(X_i=x_{n+1,i}) = \frac{1}{n_j}\sum_{r=1}^{n} 1(x_{r,i}\le x_{n+1,i})1(y_r = j)$, where $n_j = n\hat{\pi}_j$. After plugging in,

$$
\hat{f}_{X_i|Y=j}(X_i=x_{n+1,i}) = \frac{1}{2n\hat{\pi}_j h}\sum_{r=1}^{n} 1(x_{n+1,i}-h\le x_{r,i}\le x_{n+1,i}+h)1(y_r=j),\ \textit{for some } h \tag{12}
$$

PROOF of (5):

$$
\begin{aligned}
\frac{1}{2n\hat{\pi}_j h}\sum_{r=1}^{n} 1(x_{n+1,i}-h\le x_{r,i}\le x_{n+1,i}+h)1(y_r=j) &= \frac{1}{2n\hat{\pi}_j h}\sum_{r=1}^{n} 1(-h\le x_{r,i}-x_{n+1,i}\le h)1(y_r=j) \\
&= \frac{1}{2n\hat{\pi}_j h}\sum_{r=1}^{n} 1(-1\le (x_{r,i}-x_{n+1,i})/h\le 1)1(y_r=j) \\
&= \frac{1}{n\hat{\pi}_j h}\sum_{r=1}^{n} 1/2(-1\le (x_{r,i}-x_{n+1,i})/h\le 1)1(y_r=j) \\
&= \frac{1}{n\hat{\pi}_j h}\sum_{r=1}^{n} K(\frac{x_{n+1,i}-x_{r,i}}{h})1(y_r=j) \\
&, \textit{ where } K(\frac{x_{n+1,i}-x_{r,i}}{h}) \textit{ follows } Unif(-1,1)
\end{aligned}
$$