Optimal Linear Combination of Biomarkers for Multi-category Diagnosis

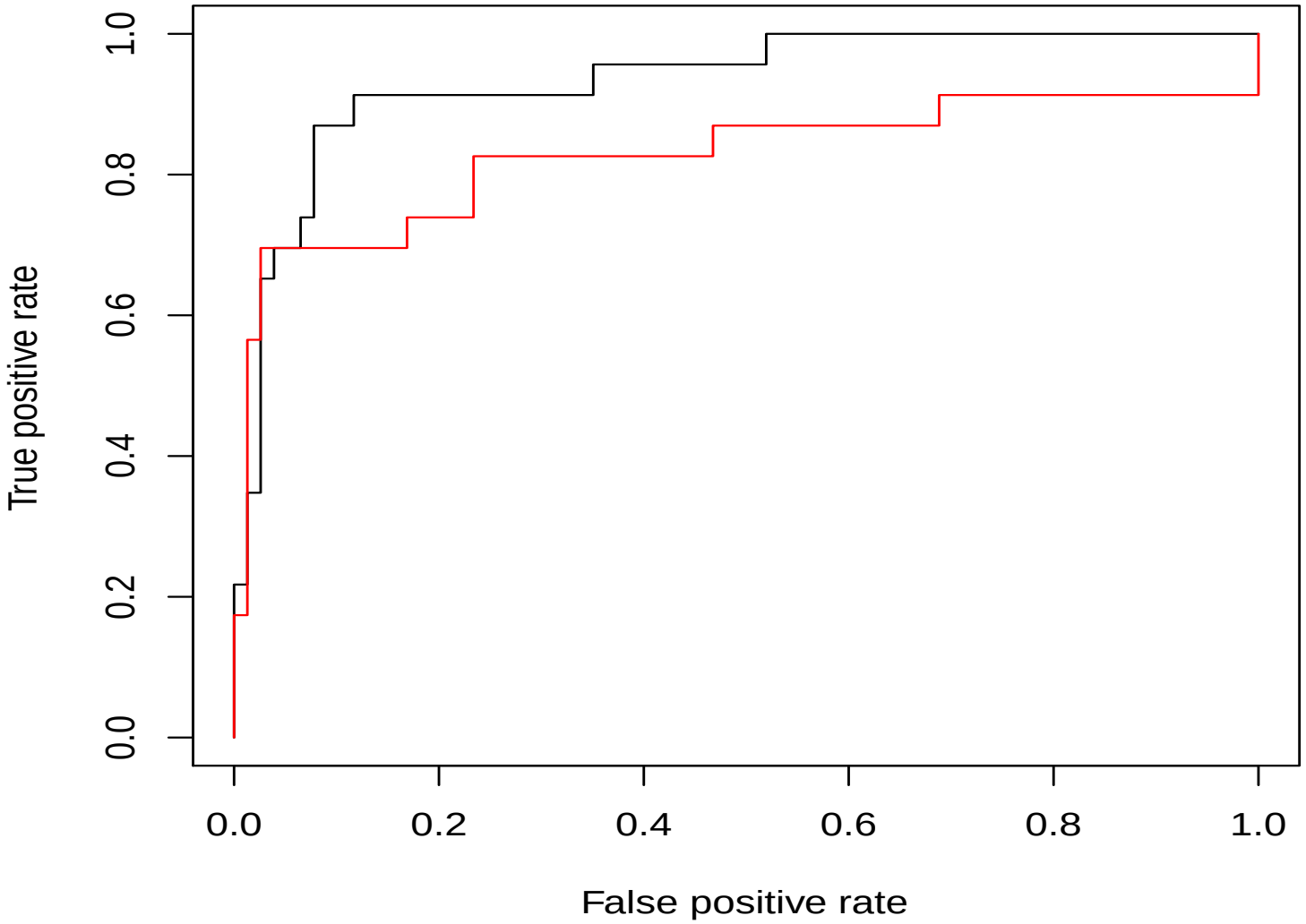
Yi-Hau Chen
Institute of Statistical Science, Academia Sinica
yhchen@stat.sinica.edu.tw

National University of Singapore
July 2017

Introduction

- the receiver operating characteristic (ROC) curve:
  a popular tool for evaluating accuracy of a diagnostic tool (biomarker)

- providing an exhaustive look at the relation between sensitivity (true positive rate) and specificity (true negative rate) over all possible decision points


True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

- the area under the ROC curve (AUC)
  summary measure derived from the ROC curve that can provide an overall assessment of diagnosis accuracy

- larger AUC implies higher diagnosis accuracy

- AUC amounts to the probability that the biomarker is concordant with the disease outcome

- X_1, X_2 the values of a biomarker X from the populations of two categories:

  $$\text{AUC} = P(X_1 > X_2)$$

  (assuming population 1 ranks higher than population 2)

- the hypervolume under the ROC manifold (HUM):
  extension of AUC to multi-category diagonosis/classification (Nakas & Yiannoutsos 2004 Stat Med)

- $X_1, \ldots, X_M$ the values of a biomarker X from the populations of M categories:

$$\text{HUM} = P(X_1 > X_2 > \cdots > X_M)$$

  (assuming populations 1 to M are ranked in descending order)

- accuracy index for multi-category diagnosis

- multiple biomarkers/diagnostic tests:
  a usual practice is to find some suitable linear combination of them, to improve the diagnostic accuracy attained by individual biomarkers/diagnostic tests (Li & Fine 2008 Biostat; Kang et al. 2013 Stat Med)

- optimal linear combination of biomarkers/diagnostic tests:
  diagnosis accuracy criterion such as AUC

- computationally challenging when using HUM as the criterion function and the number of categories is more than three

our study:

- proposing overall accuracy indices much simpler than HUM for multi-class diagnosis

- using the new accuracy measures as objective functions to derive parametric and nonparametric procedures for identifying optimal linear combinations of biomarkers in multi-category diagnosis

Diagnostic accuracy indices

- M multiple diagnostic categories; p biomarkers

- $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M$ the p-dimensional vectors of values of p biomarkers from the populations of M categories

- for a p-dimensional vector $\boldsymbol{\beta}$ and the associated linear combinations of the biomarkers, consider the probability

$$P(\boldsymbol{\beta}^T \boldsymbol{X}_M > \boldsymbol{\beta}^T \boldsymbol{X}_{M-1} > \cdots > \boldsymbol{\beta}^T \boldsymbol{X}_1)$$

  which reflects the diagnostic power for the linear combination of the biomarkers and amounts to HUM (Nakas & Yiannoutsos 2004 Stat Med)

- Let

$$\boldsymbol{P_A} = \sum_{i=1}^{M-1} P(\boldsymbol{\beta}^T \boldsymbol{X}_{i+1} > \boldsymbol{\beta}^T \boldsymbol{X}_i)/(M-1)$$

$$\boldsymbol{P_M} = \min_{1 \le i \le M-1} P(\boldsymbol{\beta}^T \boldsymbol{X}_{i+1} > \boldsymbol{\beta}^T \boldsymbol{X}_i)$$

- by Fréchet inequality for joint probability

$$\begin{aligned} \max\{0, (M-1)\boldsymbol{P_A} - (M-2)\} &\le \text{HUM} \\ &= P(\boldsymbol{\beta}^T \boldsymbol{X}_M > \boldsymbol{\beta}^T \boldsymbol{X}_{M-1} > \cdots > \boldsymbol{\beta}^T \boldsymbol{X}_1) \\ &\le \boldsymbol{P_M} \end{aligned}$$

- $P(\boldsymbol{\beta}^T \boldsymbol{X}_{i+1} > \boldsymbol{\beta}^T \boldsymbol{X}_i)$ is just the AUC of the linear combination of the markers for discriminating populations between the $(i+1)$- and i-th categories

- P_A and P_M reflect the average and worst-case diagnostic accuracies across adjacent-category discriminations

- P_A and P_M involve only pairwise comparisons between two adjacent categories, , whose calculation is much simpler than HUM

- P_A and P_M themselves may serve as convenient and useful overall accuracy indices in multi-category diagnosis

- For given β, the labeling of the diagnosis categories are defined in such a way that the corresponding PA, PM, or HUM is largest among all possible ways of labeling (Scurfield 1996 J Math Psychol)

- In general we need to evaluate the index over for all $M!$ ways of labeling to identify the correct labeling

- Finding the optimal linear combination of multiple markers thus becomes computationally challenging when there are more than three diagnosis categories

- Also, evaluating HUM would introduce further difficulty when there are more diagnosis categories since it involves M-dimensional summation or integration

- In contrast, PA or PM would substantially reduce computation burden since they just require calculations of two-category AUCs involving only 1-dimensional integration or two-fold summation

Computation procedures for optimal linear combination of biomarkers

The parametric method

- the biomarker $\boldsymbol{X}_i$ from the ith diagnostic category follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$: $\boldsymbol{X}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),\ i = 1, \ldots, M$

- The AUC comparing the distributions of the linear combinations in categories g and h:

$$AUC_{g,h}(\boldsymbol{\beta}) = P(\boldsymbol{\beta}^T \boldsymbol{X}_h > \boldsymbol{\beta}^T \boldsymbol{X}_g) = \int_0^1 \Phi\left(\frac{\boldsymbol{\beta}^T(\boldsymbol{\mu}_h - \boldsymbol{\mu}_g) - c(u)\sqrt{\boldsymbol{\beta}^T \boldsymbol{\Sigma}_g \boldsymbol{\beta}}}{\sqrt{\boldsymbol{\beta}^T \boldsymbol{\Sigma}_h \boldsymbol{\beta}}} \right) du$$

- Under the normal assumption, we can then plug the AUC formula into P_A or P_M, and solve for $\boldsymbol{\beta}$ that maximizes the resulting objective function

- In the special case

$$\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 - \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_M - \boldsymbol{\mu}_{M-1} = \boldsymbol{\delta}$$

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_M = \boldsymbol{\Sigma}$$

  the coefficient vector of the linear combination that maximizes P_A is the same as the one that maximizes P_M, and both $\propto \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$

- Also, in this special case with $M = 3$, the optimal linear combination based on P_A (and P_M) is the same as that using HUM (Zhang and Li 2011 ANZ J Stat)

- when using HUM as the criterion, the optimal linear combination of biomarkers for multi-class diagnosis under normality involves evaluation of multidimensional integration in general, which is computationally more difficult than using the P_A or P_M criterion

The nonparametric method

- a common nonparametric estimate of the AUC for comparing the distributions of $\boldsymbol{\beta^T X_g}$ and $\boldsymbol{\beta^T X_h}$ $(1 \leq g < h \leq M)$ is the Mann-Whitney statistic (Hanley & McNeil 1982 Radiology):

$$\widehat{\text{AUC}}_{g,h} = \frac{1}{n_g n_h} \sum_{s=1}^{n_g} \sum_{r=1}^{n_h} I\left(\boldsymbol{\beta}^T \boldsymbol{X}_{h,r} > \boldsymbol{\beta}^T \boldsymbol{X}_{g,s}\right)$$

  where n_i denotes the size of the sample drawn from the ith diagnosis category, and $\boldsymbol{X}_{i,t}$ denotes the vector of the biomarkers from the tth subject in the sample of the ith category

- Substitute the nonparametric estimate for pairwise AUCs in P_A or P_M and find the $\boldsymbol{\beta}$ that maximizes the resulting objective function

- the involved computation is challenging especially when the dimension p of the biomarkers $\boldsymbol{X}$ is more than three given the non-smoothness of the nonparametric estimate

- We adopt the stepwise procedures, including the step-down and step-up procedures to find the optimal linear combination based on the criterion function P_A or P_M and the nonparametric AUC estimate (Kang et al. 2013 Stat Med; Pepe & Thompson 2000 Biostat)

The step-down procedure:

Step 1 : For each of the p biomarkers, evaluate P_A (or P_M) with the nonparametric estimates of the pairwise AUCs in adjacent categories

Step 2 : Sort the p biomarkers according to the P_A (or P_M) values in descending order. That is, the first biomarker has the largest P_A (or P_M) value, the second biomarker has the second largest P_A (or P_M) value and so on

Step 3 : Combine the first two biomarkers by the linear combination

$$V = X_{(1)} + \lambda X_{(2)},$$

where $X_{(1)}$ ($X_{(2)}$) denotes the first (second) biomarker and the subscripts for category and subject have been suppressed, use the nonparametric estimate of the pairwise AUCs for the combined biomarker V when evaluating P_A (or P_M), and find the value λ that maximizes P_A (or P_M). The resulting λ gives a combined biomarker $V = X_{(1)} + \lambda X_{(2)}$

Step 4 : Run Step 3 with $X_{(1)}$ replaced by the newly obtained biomarker V and $X_{(2)}$ replaced by the third biomarker, to obtain an updated combination of the biomarkers

Step 5 : Repeat Step 4 until all p biomarkers are included in the linear combination V

The step-up procedure:

- performed in the same way as the step-down procedure, except that in Step 2 the biomarkers are sorted according to the P_A (or P_M) values in ascending order

- a common nonparametric estimate of the HUM for the linear combination of biomarkers with coefficient vector $\boldsymbol{\beta}$ is:

$$\frac{1}{n_1 \cdots n_M} \sum_{r_1=1}^{n_1} \cdots \sum_{r_M=1}^{n_M} I\left(\boldsymbol{\beta}^T \boldsymbol{X}_{M,r_M} > \boldsymbol{\beta}^T \boldsymbol{X}_{M-1,r_{M-1}} > \cdots > \boldsymbol{\beta}^T \boldsymbol{X}_{1,r_1}\right)$$

  which involves M-fold summation and hence is feasible only for small M ($M \leq 3$) (Kang et al. 2013 Stat Med)

The min-max procedure:

- An even simplified procedure using only the two most extreme biomarkers in each subject for constructing linear combinations of biomarkers (Liu et al. 2011 Stat Med)

- Consider the linear combination $\boldsymbol{X}_{h,r,max} + \lambda \boldsymbol{X}_{h,r,min}$ for the rth subject in the hth category, where $\boldsymbol{X}_{h,r,max} = \max_{1 \le j \le p} \boldsymbol{X}_{h,r,j}$ is the maximum biomarker value, and $\boldsymbol{X}_{h,r,min} = \min_{1 \le j \le p} \boldsymbol{X}_{h,r,j}$ is the minimum biomarker value for the rth subject in the hth category

- The pairwise AUCs of the min-max combination for comparing categories g and h is empirically estimated by

$$\frac{1}{n_g n_h} \sum_{s=1}^{n_g} \sum_{r=1}^{n_h} I\left(\boldsymbol{X}_{h,r,max} + \lambda \boldsymbol{X}_{h,r,min} > \boldsymbol{X}_{g,s,max} + \lambda \boldsymbol{X}_{g,s,min}\right)$$

- The λ value maximizing the corresponding P_A or P_M function then gives the optimal min-max combination based on the P_A or P_M criterion

- All the biomarkers should be standardized to have the same scale when applying the min-max procedure

Simulation studies

- Five biomarkers ($p = 5$) for three ($M = 3$) diagnosis categories

- The distribution of the biomarkers in each category is given by a multivariate normal distribution or a multivariate skew t distribution (Azzalini & Capitanio 2003 JRSSB)

- Three scenarios for the sizes of the samples from different categories: (20,20,20), (30, 40, 50), (50, 50, 50)

Multivariate normal setting

- The means of the biomarkers in the three categories are $(0.1, 0.1, 0.1, 0.1, 0.1)$, $(0.8, 1.1, 1.4, 1.7, 2.0)$, and $(1.6, 2.2, 2.8, 3.4, 4.0)$ respectively

- Biomarkers in the three categories have a common covariance matrix given by the "compound symmetric" structure with diagonal elements (marginal variances) of 1 and off-diagonal elements (pairwise correlations) of 0.5

Multivariate skew t setting (Azzalini & Capitanio JRSSB 2003)

- The location parameters of the distributions of biomarkers in the three categories are all the zero vector

- The dispersion parameters in the three categories are given by the compound symmetric matrix with diagonal elements of 1 and off-diagonal elements of 0.5

- the skewness parameters in the three categories are $(0.1, 0.1, 0.1, 0.1, 0.1)$, $(0.8, 1.1, 1.4, 1.7, 2.0)$, and $(1.6, 2.2, 2.8, 3.4, 4.0)$ respectively

- A training and an independent test samples are generated, with the training sample used to identify the optimal linear combinations, while the test sample used to evaluate the diagnosis accuracy of the identified linear combinations of the biomarkers

Methods for identifying the optimal linear combination of biomarkers

- Using the criteria P_A or P_M obtained by the parametric method based on the normality assumption, and the nonparametric method with step-down, step-up, and min-max procedures

- Using the empirical HUM criterion obtained by the nonparametric method with step-down and step-up procedures

Assessment of the diagnosis accuracy for different methods

- Nonparametric estimates of P_A, P_M, and HUM based on the test data

Multivariate normal setting

	P_A			P_M			HUM
index	Normal	Step-down	Min-Max	Normal	Step-down	Min-Max	Step-down
			$n = (20, 20, 20)$				
$\widehat{P_A}$	0.920(0.038)	0.920(0.038)	0.917(0.038)	0.920(0.039)	0.919(0.038)	0.914(0.038)	0.920(0.037)
$\widehat{P_M}$	0.915(0.039)	0.916(0.038)	0.900(0.043)	0.915(0.039)	0.914(0.039)	0.898(0.042)	0.915(0.038)
$\widehat{HUM}$	0.840(0.076)	0.841(0.074)	0.835(0.075)	0.840(0.076)	0.839(0.076)	0.829(0.075)	0.840(0.074)
			$n = (30, 40, 50)$				
$\widehat{P_A}$	0.931(0.024)	0.929(0.025)	0.908(0.022)	0.930(0.024)	0.931(0.020)	0.907(0.027)	0.929(0.025)
$\widehat{P_M}$	0.924(0.026)	0.921(0.027)	0.891(0.028)	0.923(0.026)	0.920(0.024)	0.891(0.032)	0.922(0.027)
$\widehat{HUM}$	0.863(0.048)	0.858(0.049)	0.817(0.044)	0.861(0.048)	0.863(0.039)	0.815(0.054)	0.858(0.049)
			$n = (50, 50, 50)$				
$\widehat{P_A}$	0.933(0.021)	0.930(0.022)	0.909(0.024)	0.933(0.020)	0.931(0.021)	0.907(0.024)	0.930(0.022)
$\widehat{P_M}$	0.928(0.022)	0.926(0.022)	0.891(0.026)	0.929(0.020)	0.926(0.021)	0.891(0.026)	0.926(0.022)
$\widehat{HUM}$	0.866(0.042)	0.861(0.043)	0.818(0.047)	0.866(0.039)	0.862(0.042)	0.815(0.048)	0.861(0.043)

Multivariate skew t setting

	P_A			P_M			HUM
index	Normal	Step-down	Min-Max	Normal	Step-down	Min-Max	Step-down
			$n = (20, 20, 20)$				
$\widehat{P_A}$	0.563(0.043)	0.568(0.041)	0.578(0.038)	0.532(0.054)	0.548(0.047)	0.561(0.042)	0.570(0.043)
$\widehat{P_M}$	0.502(0.021)	0.501(0.023)	0.502(0.019)	0.486(0.046)	0.493(0.035)	0.499(0.021)	0.499(0.026)
$\widehat{HUM}$	0.249(0.054)	0.256(0.051)	0.271(0.047)	0.210(0.061)	0.230(0.055)	0.247(0.053)	0.260(0.052)
			$n = (30, 40, 50)$				
$\widehat{P_A}$	0.558(0.041)	0.564(0.039)	0.576(0.038)	0.547(0.043)	0.557(0.041)	0.561(0.043)	0.566(0.040)
$\widehat{P_M}$	0.491(0.053)	0.492(0.056)	0.496(0.050)	0.486(0.054)	0.490(0.053)	0.493(0.047)	0.493(0.055)
$\widehat{HUM}$	0.240(0.053)	0.248(0.050)	0.266(0.049)	0.225(0.053)	0.238(0.051)	0.245(0.056)	0.251(0.050)
			$n = (50, 50, 50)$				
$\widehat{P_A}$	0.571(0.027)	0.576(0.024)	0.579(0.022)	0.542(0.039)	0.553(0.031)	0.559(0.033)	0.578(0.024)
$\widehat{P_M}$	0.504(0.009)	0.503(0.010)	0.502(0.009)	0.496(0.025)	0.500(0.014)	0.501(0.010)	0.501(0.010)
$\widehat{HUM}$	0.261(0.034)	0.267(0.030)	0.273(0.026)	0.219(0.047)	0.233(0.039)	0.244(0.043)	0.271(0.029)

Summary: comparison of computation methods

- When the biomarkers are multivariate normal and the P_A or P_M criterion is used for identifying the optimal linear biomarker combination, the parametric procedure based on the normality assumption and the nonparametric step-down procedure perform quite similarly, and perform better than the min-max method

- When the biomarkers follow multivariate skew t, the min-max method performs best, followed by the nonparametric step-down procedure, and the parametric procedure based on the wrong normality assumption performs worst

Summary: comparison of optimization criteria

- the optimal linear combinations obtained by the P_A, P_M and HUM criteria achieve similar diagnosis accuracy when the biomarkers follow multivariate normal

- When biomarkers follow multivariate skew t, the diagnosis accuracy of the optimal linear combination identified by the P_A criterion is very similar to that by the HUM criterion, and both of them are slightly better than the accuracy obtained by the P_M criterion

Summary: comparison of coefficient results

- The P_A criterion in general yields coefficients of the optimal linear combination closer to those obtained by the HUM criterion than the P_M criterion (results not shown)

Summary: comparison of computation time

- the P_A- and P_M-based methods are much more efficient in computation than the HUM-based method

Application: The Alzheimer's disease data

- 14 neuropsychological biomarkers as diagnostic tools for 3 categories of Alzheimer's disease

- 118 subjects of age 75 are divided into three categories: non-demented (44 subjects), very mildly demented (43 subjects), and mildly demented (21 subjects)

- According to Xiong et al. (2006), the estimated HUMs of individual biomarkers using the parametric method range from 0.347 to 0.752

- Identify the optimal coefficients using the parametric and nonparametric P_A and P_M methods, the nonparametric HUM method, and the naive method which sets equal weights to all 14 markers

- The coefficients from each method are standardized to have unit length.

- Different methods are assessed through the P_A, P_M and HUM performance indices estimated by the nonparametric method with the same dataset

Method	The optimal coefficients											
	f1	kt	kpar	kfr	zp4	zp5	zp6	zinfo	zbc	zbd	zb	zment
Naive	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267
P_A-par	0.175	0.251	0.197	0.400	0.354	-0.079	0.258	-0.226	0.402	-0.495	0.038	-0.203
P_A-npar	0.182	0.541	0.004	0.273	0.334	-0.323	0.192	-0.462	0.128	-0.286	-0.118	0.090
P_M-par	-0.127	0.247	0.360	0.080	0.147	0.123	0.411	-0.072	0.196	-0.424	0.291	0.060
P_M-npar	-0.099	0.658	0.000	0.500	0.072	-0.007	0.428	-0.007	-0.092	-0.026	-0.323	-0.020
HUM	0.188	0.500	0.354	0.373	0.355	-0.038	0.184	-0.278	0.087	-0.154	-0.164	-0.096

Method	$\widehat{P_A}$	$\widehat{P_M}$	$\widehat{HUM}$
Naive	0.886	0.882	0.792
P_A-par	0.913	0.900	0.827
P_A-npar	0.913	0.906	0.826
P_M-par	0.905	0.904	0.813
P_M-npar	0.901	0.901	0.805
HUM	0.906	0.892	0.823

Summary of the results

- The optimal linear combination identified by the parametric and nonparametric P_A criteria perform the best among all the methods; they achieve highest P_A, P_M and HUM accuracy indices

- The linear combination identified by the parametric P_M criterion performs slightly worse than those obtained by the P_A and the HUM criteria

- The naive (equal-weight) method has the worst performances in this dataset

Application: The heart disease data

- 303 heart disease patients from five categories representing different presence statuses of the heart disease
- The numbers of subjects in the five categories are 164, 55, 36, 35, and 13, respectively
- Four biomarkers are available for disease classification

- The distributions of the biomarkers do not follow a specific ordering across the five categories

- Recall that the diagnosis accuracy index P_A, P_M or HUM is defined as the maximum value of the index across all possible (5! = 120) orderings of the categories

- To find the best linear combination of the biomarkers, it may be more appropriate to conduct the calculations without assuming a specific ordering of the diagnosis categories, even though the disease categories are defined on an ordinal scale

- The need to evaluate the objective function for a large number of orderings of the categories renders the HUM-based method infeasible since it already involves five-dimensional summation

- The proposed P_A and P_M criteria involve only two-dimensional summation for evaluating pairwise AUCs over adjacent categories, hence are computationally much more efficient

	The optimal coefficients						
Method	trestbps	chol	thalach	oldpeak	$\widehat{P_A}$	$\widehat{P_M}$	$\widehat{HUM}$
Naive	0.5000	0.5000	0.5000	0.5000	0.5493	0.5226	0.0147
P_A-par	0.2548	0.1268	-0.3577	0.8894	0.6190	0.5127	0.0451
P_A-npar	0.0034	0.1563	-0.9406	0.3013	0.6166	0.5235	0.0704
P_M-par	0.1421	-0.6734	-0.6319	0.3564	0.5817	0.5209	0.0350
P_M-npar	0.2526	0.1467	-0.9287	0.2285	0.6100	0.5363	0.0624

Results

- The empirical estimates of HUMs for these four biomarkers are 0.0111 (trestbps), 0.0109 (chol), 0.0297 (thalach), and 0.0251 (oldpeak), respectively

- The P_A-based methods generally perform well in terms of the accuracy performance measures P_A, P_M, and HUM

- All the proposed linear combinations of the four biomarkers attain higher HUM than the best single biomarker, while the linear combination from the naive method does not

Thank You !!