Published in final edited form as:

J Appl Stat. 2009 July 7; 36(7): 769–778. doi:10.1080/0266476YYxxxxxxx.

Approximating the risk score for disease diagnosis using MARS

Binbing Yu

Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, Bethesda, Maryland 20892, U.S.A.

Abstract

In disease screening and diagnosis, often multiple markers are measured and they are combined in order to improve the accuracy of diagnosis. McIntosh and Pepe (2002, *Biometrics* **58**, 657-644) showed that the risk score, defined as the probability of disease conditional on multiple markers, is the optimal function for classification based on the Neyman-Pearson Lemma. They proposed a two-step procedure to approximate the risk score. However, the resulted ROC curve is only defined in a subrange (L, h) of the false-positive rates in $(0,1)$ and determination of the lower limit L needs extra prior information. In practice, most diagnostic tests are not perfect and it is usually rare that a single marker is uniformly better than the other tests. Using simulation, I show that multivariate adaptive regression spline (MARS) is a useful tool to approximate the risk score when combining multiple markers, especially when the ROC curves from multiple tests cross. The resulted ROC is defined in the whole range of $(0,1)$ and it is easy to implement and has intuitive interpretation. The sample code of the application is shown in the appendix.

Keywords

Multivariate adaptive regression spline (MARS); Neyman-Pearson Lemma; risk score; ROC

1. Introduction

The effectiveness of continuous markers in distinguishing between healthy and diseased subjects is generally assessed through the use of the Receiver Operating Characteristic (ROC) curve [11]. A subject is assessed as diseased (positive) or healthy (negative) according to whether the subject's marker value is greater than or less than or equal to a specified threshold value. Associated with any threshold value is the probability of a true positive (sensitivity) and the probability of a true negative (specificity). The resulting theoretical ROC curve is the plot of sensitivity versus 1-specificity for all possible threshold values. The area under the ROC curve (AUC) is a widely accepted summary index of the overall performance of diagnostic procedures.

In the screening and diagnosis of disease, often multiple tests are performed or multiple markers are routinely measured. For example, the neuropathologic diagnosis of Alzheimer's disease is typically based on the densities and distributions of brain lesions such as senile plaques and neurofibrillary tangles from different brain regions [17]. Strategies for combining information from multiple tests are therefore important since a combination of the tests may provide more accurate diagnosis than any single test alone [13]. When the multiple markers are multivariate normal in both diseased and healthy populations, Su and Liu [16] derived an optimal linear discriminant function that maximizes the AUC. Pepe and Thompson [13] considered how to choose linear combinations of markers in order to optimize diagnostic accuracy. Pepe et al. [12] proposed to use the empirical AUC as the objective function for combining multiple markers. The literature in the field of combining multiple biomarkers (tests) is vast. For the

other approaches of combination, see Baker [1], Ma and Huang [9], Yang and Jin [18], Zou et al. [20], among others.

Let D be a binary indicator for disease status, with $D = 1$ for diseased subjects and $D = 0$ for nondiseased subjects and let the multiple diagnostic markers denoted by $\mathbf{Y} = (Y_1, \dots, Y_K)$. Several methods have been proposed to find the optimal linear combination of the multiple markers, i.e., to find a vector $\beta = (\beta_1, \beta_2, \dots, \beta_K)$ such that the linear score

$$L_{\beta}(\mathbf{Y}) = \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_K Y_K \quad (1)$$

maximizes the objective function, e.g. the AUC. For standardization, one can restrict $\beta_1 = 1$. Based on the Neyman-Pearson Lemma, McIntosh and Pepe [10] showed that the likelihood ratio (LR) rule is the uniformly most sensitive (UMS) screening test to combine multiple markers in the sense that the ROC curve is maximized at every point. They also cleverly used Bayes' theorem to show that the screening rules based on risk scores, i.e., probabilities of being diseased, are equivalent to the LR rule conditional on multiple markers.

They proposed a two-step screening procedure to approximate the risk score, which involves refitting a regression model for a restricted region of the observations. Although the ROC curve from the two-step procedure is substantially closer to the optimal ROC curve, it is only estimable for a sub range of $(0, 1)$. The choice of restricted region needs extra information. I will show later that the Multivariate Adaptive Regression Spline (MARS) is a useful alternative tool to approximate the risk score.

2. Methods

2.1. Reviewing the Existing Method of Approximating Risk Score

Under the likelihood ratio rule, a decision of positive test is made whenever

$$LR(\mathbf{Y}) \equiv \frac{P(\mathbf{Y}|D=1)}{P(\mathbf{Y}|D=0)} > c_{\alpha}, \quad (2)$$

where c_{α} is chosen so that the false positive rate (Type I error) is $\alpha = P\{LR(\mathbf{Y}) > c_{\alpha} | D = 0\}$. According to the Neyman-Pearson Lemma, the LR decision rule yielded the highest (optimal) ROC curve among all classification rules, and thus is the UMS test [10]. This is a powerful result that shows the optimality property of the likelihood ratio rule. Based on a simulation of one million marker pairs for ovarian cancer, they confirm the conclusion.

The Neyman-Pearson Lemma proves that the LR test is the most power test under the simplest situation where both the null and alternative hypothesis are simple, i.e., each contains a single known distribution. The uniform powerful test (UMP) also exists if both the hypothesis are from a real-parameter family with monotone likelihood ratio [8]. However, it is rarely true that the marker distributions $P(\mathbf{Y}|D)$ are simple. They usually involve more than one unknown parameter. For example in McIntosh and Pepe [10], the joint distribution of the two ovarian cancer markers is multivariate normal in the control group and is mixture multivariate normal in the diseased group, respectively. In practice, the parameters are unknown and have to be estimated from sample. The optimality of the LR rule relies on the correct specification of the distributions $P(\mathbf{Y}|D)$, a misspecified model could lead to a less powerful, or even biased, test. A more general class of models should provide a better approximation to the risk scores than the linear combination of the multiple tests.

By Bayes' rule, McIntosh and Pepe [10] show that the risk score is a monotone function of the LR function. The risk score is defined as

$$P(D=1|\mathbf{Y}) = \frac{LR(\mathbf{Y})q}{LR(\mathbf{Y})q+1},$$

where $q = P(D=1)/P(D=0)$ is the odds of disease in the population. They then propose a two-step screening procedure to approximate the risk score rule. Although the two-step procedure yields the AUC close to that from the LR rule, it has several drawbacks. First, the estimation of ROC in the two-step procedure is restricted to a sub-range of false positive rate $\alpha \in (L, h) \subset (0, 1)$. Often, the researchers are interested in the ROC in the whole range $(0, 1)$, especially both ends where false positive (negative) rate is low. Second, the choice of the lower bound L needs some information a priori. It is suggested to use $L = 5/n_D$, where n_D is the number of nondiseased subjects. For a common number $n_D = 50$, the lower bound $L = 0.1$, which means that the ROC is not estimable for false positive rate $\alpha < 0.1$. Third, the two-step procedure involves fitting a second logistic regression in the region where roughly the false positive rate $\alpha < h$ and h should be a relatively small number, say 0.15. If h is close to 1, then the logistic regression in the second step is likely to fail because of the small number of observations in the second step.

2.2. Proposed Method

Pepe et al. [12] mention that the linear predictor class for approximating the risk score is quite large, which includes generalized additive model, smoothing and regression spline, kernel method etc. They subsequently model the risk score using a generalized linear model with an unspecified link $g(\cdot)$, where

$$P(D=1|\mathbf{Y}) = g(L_\beta(\mathbf{Y})).$$

This relaxes the logistic assumption for $g(\cdot)$. In the literature by far, the linear score $L_\beta(\mathbf{Y})$ is assumed to be a linear combination of the markers with constant weights $(\beta_1, \beta_2, \dots, \beta_K)$. Here, I take a different perspective and explore the use of multivariate adaptive regression splines (MARS) to model the score $L_\beta(\mathbf{Y})$.

The MARS, developed by Friedman [4], is a multivariate non-parametric regression procedure. The MARS procedure builds flexible regression models by fitting separate splines (or basis functions) to distinct intervals of the predictor variables. Kooperberg et al. [6] extend the MARS to categorical outcome. The procedure uses linear spline and their tensor products to fit a regression model for polychotomous response. Based on simulations, Kooperberg et al. [6] show that MARS performs at least as well as other established methods for binary response, e.g., logistic regression, classification and regression trees etc., but yields the advantage of being much faster in fitting huge datasets. The procedure is implemented by the functions POLYCLASS and POLYMARS in the library `polyspline` on R 2.7.0 [14]. To avoid over-fitting and for easier interpretation, I only consider the model with linear splines but no interaction (product) terms. As a result, the risk score is modeled as

$$\text{logit}\{P(D=1|\mathbf{Y})\} = g_0 + \sum_{i=1}^K g_i(Y_i), \quad (3)$$

where

$$g_i(Y_i) = b_{i0}Y_i + \sum_{j=1}^{J_i} b_{ij}(Y_i - \tau_{ij})^+,$$

with $(z - \tau)^+ = \max(z - \tau, 0)$. This model is also called a piece-wise linear model and τ_{ij} , $j = 1, \dots, J_i$, $i = 1, \dots, K$, are called the change points and are estimated from the MARS. This model has been used to in many settings, for example modeling the trends of cancer incidence and mortality rates [5].

There are several advantages of using MARS to approximate the risk score. First, the implementation is simpler than the two-step procedure and the resulted ROC is defined in the whole range (0,1). The combination in (3) has an easy interpretation. For example, for two markers Y_1 and Y_2 with one change point τ_1 for Y_1 , the linear combination is $b_{10}Y_1 + b_{20}Y_2$ when $Y_1 < \tau_1$ and the combination is $(b_{10} + b_{11})Y_1 + b_{20}Y_2$ when $Y_1 \geq \tau_1$. Based on this interpretation, the procedure based on MARS yields similar piecewise screening rule by the two-step procedure as shown in Figure 3a in McIntosh and Pepe [10], but the calculation is simpler and less likely to fail. Next, we will use simulation to compare the proposed method with the linear combination $L_\beta(\mathbf{Y})$ method and the LR rule. In the simulation and example, we limit the maximum of the total number of change points to K , the total number of tests, in order to avoid the problem of over-fitting.

3. Simulation

Here I consider two markers Y_1 and Y_2 . In all cases the marker values for the nondiseased subjects are generated from a standard normal $N(0, 1)$ distribution, while the marker values for the diseased subjects are generated from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ for tests Y_1 and Y_2 respectively. The marker values could be independent ($\rho = 0$) or correlated with $\rho = 0.5$.

The parameters of the distributions for the two markers are shown in Table 1. For Cases 1-3, marker Y_2 has better diagnostic accuracy than marker Y_1 . For Cases 4-6, the ROC curves for the two markers cross. Figure 1 shows the ROC curves for two typical cases. The sample size (n) in each of the control and diseased groups $n_D = n_{\bar{D}} = 50$ or 100.

The results are based on 1000 simulations. The AUCs of the ROC curves from different methods are shown in Table 2. The columns Y_1 , Y_2 and $L_\beta(\mathbf{Y})$ show the AUCs for markers 1 and 2 and their linear combination. The weights of the linear combination are derived from Su and Liu's method. The columns for MARS and ELR are the AUCs based on the MARS and the empirical likelihood ratio (ELR),

$$ELR(\mathbf{Y}) = \frac{\hat{P}(\mathbf{Y}|D=1)}{\hat{P}(\mathbf{Y}|D=0)},$$

where $\hat{P}(\mathbf{Y}|D)$ are the estimated normal distributions $N(\hat{\mu}_k, \hat{\sigma}_k^2)$, $k=1, 2$. Because the LR rule is the optimal classification rule, the AUC based on the LR rule is the largest and is used as benchmark. Because the simulated data are normally distributed, the AUCs from the normal linear discriminant (NLD) rule [16] is shown in the last column. For all cases, the AUCs based on $L_\beta(\mathbf{Y})$ and the NLD rule are very close. We see that all four combination methods, i.e., the linear combination $L_\beta(\mathbf{Y})$, the MARS, ELR and NLD rules, improve the AUC value. For Cases 1-3, the AUC from all four methods are close and the ELR rule has the largest AUC values.

However, for Cases 4-6 where the two ROC curves cross, MARS performs much better than the linear score $L_{\beta}(\mathbf{Y})$ and is very close to the ELR rule.

When the correlation ρ increases from 0 to 0.5, the relative gain for the MARS and the ELR methods compared to the linear score remains similar. Compared to the linear score method, the increase of AUC by using the MARS is slightly more for the smaller sample size. I also examined the case of markers with high correlation $\rho = 0.8$, which is often the case when exposures are mixtures. The AUC estimates for $\rho = 0.8$ are similar to those for $\rho = 0.5$. Overall, the MARS method performs well and yields AUCs close to the optimal LR rule, especially when two ROC curves cross. This is because neither methods rely on the linear combination of the markers. Thus, the MARS appears to be a useful alternative of approximating the risk score.

4. Examples

4.1. Ovarian Cancer Data in DeLong et al. (1988)

DeLong et al. [3] compare the discriminating ability of three screening scores on 49 consecutive ovarian cancer patients undergoing correction of intestinal obstruction at Duke University Medical Center. The 12 patients surviving more than 2 months postoperatively were considered surgical successes and the remaining 37 are considered failures. The failure cases are considered as test positive, so $n_D = 37$ and $n_{\bar{D}} = 12$. The three scores are the Krebs-Goplerud (K-G) score, total protein (TP) and albumin (ALB) and the latter two are positively associated with the patient's nutritional status.

The AUCs are 0.622 for ALB, 0.543 for TP and 0.687 for the K-G score, respectively. One would like to see whether the combination could improve the test accuracy. Figure 2a shows the ROC curves for the three scores and their linear combination. The K-G score has higher sensitivity than the other two methods when the false positive rate is greater than 0.5, but has lower sensitivity than the other two when the false positive rate is low. Thus, their ROC curves cross.

When the Su and Liu's linear combination is used, the vector of weights is $(\beta_{ALB}, \beta_{TP}, \beta_{KG}) = (0.051, -0.066, 0.615)$. The negative value for β_{TP} is due to the positive correlation between ALB and TP. The AUC for the NLD method is 0.687, which is very close to the AUC for the K-G score. In Figure 2a, we see no improvement in ROC of the linear combination versus the other three scores.

Figure 2b compares the ROC curves from the NLD method, the two-step method and the MARS. Because $n_{\bar{D}} = 12$, the lower bound $L = 5/n_{\bar{D}} = 0.417$, which means the ROC from the two-step method can only be estimated for false positive rate greater than 0.417. Consequently, the second-step analysis is restricted to the marker range $A = \{\mathbf{Y} : Y_k < y_k(L/3), k = 1, 2, 3\}$, where $y_k(x)$ is the $(1-x)$ empirical quantile for the marker Y_k in the controls. The individuals outside the range A are classified as cases. As shown in Figure 2b, the ROC curve from the MARS is above those from the other two methods. The two-step method cannot estimate the ROC for false positive rate below 0.417.

When the MARS is used to fit the risk score, one change point is found for the K-G score at $\tau = 3$. The vector of weights are $(\beta_{ALB}, \beta_{TP}, \beta_{KG}, \beta_{(KG-\tau)^+}) = (0.083, -0.083, 2.898, -3.030)$. The interpretation is intuitive. When the K-G score is higher, it is more sensitive than the other two tests, so a larger weight should be used for the K-G score. In contrast, a smaller weight is used when K-G score is lower than 3. As shown in Figure 2b, the ROC from the MARS method has the AUC 0.773 and is much better than that from the linear combination. The likelihood ratio test with 2 degrees of freedom for comparing the MARS method and the linear

combination is borderline significant ($p = 0.057$). The lack of power is probably due to the small sample size.

4.2. Simulated Ovarian Cancer Marker Data

Motivated by a biological model, McIntosh and Pepe [10] simulate two biomarkers, e.g., CA 125 (marker 1) and Her 2/neu (marker 2). For the control group, both markers have standard normal marginal distributions. For the diseased group, marker 1 has a mixture normal distribution $Y_1 \sim \mathbf{I}(W_1 = 0)N(0, 1) + \mathbf{I}(W_1 = 1)N(1.68, 2)$ and marker 2 (Her2/neu) has a mixture normal distribution $Y_2 \sim \mathbf{I}(W_2 = 0)N(0, 1) + \mathbf{I}(W_2 = 1)N(2.54, 2)$, where (W_1, W_2) are binary indicators to show that the markers elevate with the onset of cancer. Let $p_{ij} = P(W_1 = i, W_2 = j)$, $i, j = 0, 1$. In the example of the simulation for large sample, they assume that the odd ratio $\Psi = p_{00}p_{11}/(p_{10}p_{01}) = 0.1$ and the correlation between Y_1 and Y_2 is $\rho = 0.6$ when $W_1 = W_2 = 1$. I use a moderately large sample size with $n_D = n_{\bar{D}} = 10,000$.

The AUCs for the two markers are 0.640 and 0.614, respectively. The AUCs from the linear combination, the MARS and LR rule are 0.715, 0.739 and 0.745, respectively. As shown in Figure 3, the bottom two ROC curves are for the two markers and they cross at false positive rate 0.8. The ROC curve from the MARS method is very close to that from the LR rule and is almost indistinguishable for false positive rate less than 0.4. The ROC from the MARS is above that from the linear combination for the whole range (0,1). The difference of the AUCs between the MARS and the linear combination is highly statistically significant because of the large sample size. If we look closely at the range (0, 0.15), the figure is very similar to Figure 2b in [10].

5. Discussion

In this article, the MARS is proposed as a tool to approximate the risk score in disease diagnosis. This approach extends the linear combination of the markers by adding linear spline terms. The resulted ROC curve is defined in the whole range of false positive rate (0,1) and the implementation is easier than the two-step procedure. Based on simulation, the AUCs based on the MARS are very close to those based on the LR rule.

The MARS technique is now widely used as one of the classification tools in the problems with a large number of predictors, for example gene expression data [7]. MARS has been previously shown to be potentially useful for gene association studies and the detection of variable interactions [2,19]. As a new basis of diagnostic tests is being provided by the vast amount of data on gene expression that are now becoming available through large-scale measurement of mRNA abundance [15], we expect that MARS should become a useful tool for disease diagnosis using gene expression data. To facilitate the employment of the proposed method, the sample code for the analysis of DeLong et al's data is provided in the appendix.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institute on Aging.

Appendix A. Sample code for the analysis of DeLong et al's data

```
#####
# This program compares different methods of approximating risk
# scores using the data from DeLong et al. (1988)
#Data:http://ssc.utexas.edu/docs/sashelp/sugi/24/Posters/p250-24.pdf
#####
```

```

# Include R libraries
library(ROCR)
library(caTools)
library(polspline)
# Read DeLong et al's data
YD<-read.table("DeLong.txt",na.string=".",sep="\t",header=TRUE)
YD$y1<-as.numeric(YD$y1)
YD$y2<-as.numeric(YD$y2)
# Figure 2a pred.y1<-prediction(YD$y1, YD$d)
perf.y1<-performance(pred.y1, 'tpr', 'fpr')
pred.y2<-prediction(YD$y2, YD$d)
perf.y2<-performance(pred.y2, 'tpr', 'fpr')
pred.y3<-prediction(YD$y3, YD$d)
perf.y3<-performance(pred.y3, 'tpr', 'fpr')
pred4<-glm(d y1+y2+y3,family=binomial(link="logit"),data=YD)
pred.y4<-prediction(pred4$fitted.value, YD$d)
perf.y4<-performance(pred.y4, 'tpr', 'fpr')
plot(perf.y1, lty=2, lwd=2, main="",xaxis.lwd=2, yaxis.lwd=2,
      yaxis.cex.axis=1.5,xaxis.cex.axis=1.5,box.lwd=2,
      cex.label=2,cex.main=2)
plot(perf.y2, lty=1, lwd=2,add=TRUE)
plot(perf.y3,lty=3,lwd=2,add=TRUE)
plot(perf.y4,lty=1,lwd=3,add=TRUE)
legend(0,1.0,c("ALB", "TP", "K-G Score", "Linear
Combination"),lty=c(1,2,3,1),lw=c(1,1,1,2),cex=1.25)
# Figure 2b
plot(perf.y4,lty=1,lwd=2,main="",xaxis.lwd=2,yaxis.lwd=2,
      yaxis.cex.axis=1.5,xaxis.cex.axis=1.5,box.lwd=2,cex.label=2,cex.main=2)
#----- Approximate risk score using POLYCLASS (MARS)-----
fmars<-polyclass(YD$d,YD[,c("y1","y2","y3")],silent=TRUE,maxdim=7)
nk<-fmars$knots[,1] na<-sum(nk)
aY<-matrix(0,nrow=nrow(YD),ncol=na)
j<-1
for (k in 1:3) {
  if (nk[k]>0) {
    for (l in 1:(nk[k])) {
      tau<-fmars$knots[k,l+1]
      aY[,j]<-pmax(YD[,paste("y",k,sep="")]-tau,0)
      j<-j+1
    }
  }
}

```



```

}
aY<-data.frame(aY) if (na>0) {
  names(aY)<-paste("a",1:na,sep="")
  YD<-cbind(YD,aY)
  if (na==1) pred5<-glm(d y1+y2+y3+a1, family=binomial, data=YD)
  if (na==2) pred5<-glm(d y1+y2+y3+a1+a2, family=binomial, data=YD)
  if (na==3) pred5<-glm(d y1+y2+y3+a1+a2+a3, family=binomial, data=YD)
  print(summary(pred5))
  print("Chi-square test (linear combination vs. MARS)=")
  print(round(1-pchisq(deviance(pred4)-deviance(pred5),df=2,ncp=0),3))
  pred.y5 <- prediction(pred5$fitted.value, YD$d)
  perf.y5 <- performance(pred.y5, 'tpr', 'fpr')
  plot(perf.y5, lty=2,lwd=3,add=TRUE)
}
if (na==0) print("no improvement of MARS")
legend(0,1.0,c("Linear-Combination","MARS"),lty=c(1,2),lw=c(3,3),cex=1.25)

```

References

1. Baker S. Identifying combination of cancer markers for further study as triggers of early intervention. *Biometrics* 2000;56:1082–1087. [PubMed: 11129464]
2. Cook NR, Zee RYL, Ridker PM. Tree and spline based association analysis of gene-gene interaction models for ischemic stroke. *Statistics in Medicine* 2004;23:1439–1453. [PubMed: 15116352]
3. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845. [PubMed: 3203132]
4. Friedman JH. Multivariate adaptive regression splines (with discussion). *Annals of Statistics* 1991;19:1–141.
5. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* 2000;19:335–351. (correction: 2001 20, 655). [PubMed: 10649300]
6. Kooperberg C, Bose S, Stone CJ. Polychotomous Regression. *Journal of the American Statistical Association* 1997;92:117–127.
7. Lee JW, Lee JB, Park M, Song SH. An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics and Data Analysis* 2005;48:869–885.
8. Lehman, EL. *Testing Statistical Hypothesis*. 2nd. Chapman and Hall; New York: 1994.
9. Ma S, Huang J. Combining multiple markers for classification using ROC. *Biometrics* 2007;63:751–757. [PubMed: 17825007]
10. McIntosh MW, Pepe MS. Combining several screening tests: optimality of the risk score. *Biometrics* 2002;58:657–664. [PubMed: 12230001]
11. Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; 2002.
12. Pepe MS, Cai T, Longton G. Combining predictors for classification using the area under the Receiver Operating Characteristic curve. *Biometrics* 2006;62:221–229. [PubMed: 16542249]
13. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. *Biostatistics* 2000;1:123–140. [PubMed: 12933515]
14. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: 2008. URL <http://www.R-project.org>

15. Sandvik AK, Alsberg BK, Norsett KG, Yadetie F, Waldum HL, Lagreid A. Gene expression analysis and clinical diagnosis. *Clinica Chimica Acta* 2006;363:157–164.
16. Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association* 1993;88:1350–1355.
17. Xiong C, McKeel DW, Miller JP, Morris JC. Combining correlated diagnostic tests: application to neuropathologic diagnosis of Alzheimer's disease. *Medical Decision Making* 2004;24:659–669. [PubMed: 15534346]
18. Yang Y, Jin Z. Combining dependent tests to compare the diagnostic accuracies - a nonparametric approach. *Statistics in Medicine* 2006;25:1239–1250. [PubMed: 16158406]
19. York TP, Eaves LJ. Common disease analysis using multivariate adaptive regression splines (MARS): Genetic Analysis Workshop simulated sequence data. *Genetic Epidemiology* 2001;21:S649–S654. [PubMed: 11793755]
20. Zou KH, Bhagwat JG, Carrino JA. Statistical combination schemes of repeated diagnostic test data. *Academic Radiology* 2006;13:566–572. [PubMed: 16627197]

Figure 1.
Plots of the ROC curves for markers 1 and 2

a. Three scores vs. linear combination

b. NLD, two-step method vs. MARS

Figure 2.
Plots of the ROC curves for the three scores, their linear combination and for the MARS method

Figure 3.
The ROC curves for the two ovarian cancer markers and three combination methods

Table 1

Parameters of the distributions for markers 1 and 2

Case	μ_1	σ_1	μ_2	σ_2
1	0.5	1.0	0.8	1.0
2	0.5	1.0	1.2	1.0
3	0.5	1.0	1.5	1.0
4	0.5	1.0	0.8	2.0
5	0.5	1.0	1.2	2.0
6	0.5	1.0	1.5	2.0

Table 2

Comparison of the AUCs from different methods

Case	ρ	n	Y_1	Y_2	$L_p(Y)$	MARS	ELR	NLD
1	0.0	50	0.637	0.715	0.754	0.762	0.763	0.757
1	0.0	100	0.638	0.714	0.751	0.755	0.756	0.753
1	0.5	50	0.637	0.715	0.754	0.762	0.763	0.756
1	0.5	100	0.636	0.711	0.748	0.752	0.754	0.750
2	0.0	50	0.635	0.801	0.823	0.828	0.829	0.826
2	0.0	100	0.639	0.800	0.822	0.824	0.825	0.824
2	0.5	50	0.634	0.802	0.825	0.830	0.831	0.827
2	0.5	100	0.637	0.803	0.823	0.826	0.827	0.825
3	0.0	50	0.638	0.855	0.871	0.875	0.875	0.873
3	0.0	100	0.637	0.855	0.869	0.870	0.872	0.871
3	0.5	50	0.634	0.856	0.872	0.875	0.876	0.874
3	0.5	100	0.635	0.856	0.870	0.871	0.872	0.871
4	0.0	50	0.636	0.675	0.724	0.753	0.753	0.726
4	0.0	100	0.636	0.679	0.723	0.746	0.745	0.725
4	0.5	50	0.636	0.678	0.726	0.755	0.754	0.729
4	0.5	100	0.637	0.679	0.724	0.747	0.746	0.726
5	0.0	50	0.636	0.754	0.785	0.802	0.800	0.787
5	0.0	100	0.635	0.754	0.782	0.795	0.794	0.784
5	0.5	50	0.636	0.755	0.785	0.802	0.801	0.788
5	0.5	100	0.638	0.752	0.782	0.795	0.794	0.784
6	0.0	50	0.639	0.804	0.827	0.839	0.837	0.829
6	0.0	100	0.637	0.806	0.826	0.834	0.833	0.827
6	0.5	50	0.640	0.804	0.827	0.839	0.838	0.829

Yu

Case	ρ	n	Y_1	Y_2	$L_{\rho}(Y)$	MARS	ELR	NLD
6	0.5	100	0.638	0.807	0.827	0.834	0.834	0.829

y1	y2	y3	d	d1	id
3.0	5.8	0	0	2	1
3.2	6.3	5	1	1	2
3.9	6.8	7	1	1	3
2.8	4.8	4	0	2	4
3.2	5.8	7	1	1	5
0.9	4.0	5	0	2	6
2.5	5.7	2	0	2	7
1.6	5.6	5	1	1	8
3.8	5.7	5	1	1	9
3.7	6.7	4	1	1	10
.	.	4	1	1	11
3.2	5.4	6	1	1	12
3.8	6.6	4	1	1	13
4.1	6.6	5	1	1	14
3.6	5.7	5	1	1	15
4.3	7.0	6	1	1	16
3.6	6.7	6	0	2	17
2.3	4.4	4	1	1	18
4.2	7.6	6	0	2	19
4.0	6.6	4	0	2	20
3.5	5.8	4	1	1	21
3.8	6.8	3	1	1	22
3.0	4.7	2	0	2	23
4.5	7.4	5	1	1	24
3.7	7.4	5	1	1	25
3.1	6.6	4	1	1	26
4.1	8.2	4	1	1	27
4.3	7.0	5	1	1	28
4.3	6.5	6	1	1	29
3.2	5.1	5	1	1	30
2.6	4.7	4	1	1	31
3.3	6.8	4	0	2	32
1.7	4.0	3	0	2	33
.	.	4	1	1	34
3.7	6.1	5	1	1	35
3.3	6.3	3	1	1	36
4.2	7.7	4	1	1	37
3.5	6.2	5	1	1	38
2.9	5.7	1	0	2	39
2.1	4.8	3	1	1	40
.	.	2	1	1	41

y1	y2	y3	d	d1	id
2.8	6.2	2	0	2	42
.	.	3	1	1	43
.	.	3	1	1	44
4.0	7.0	3	1	1	45
3.3	5.7	4	1	1	46
3.7	6.9	5	1	1	47
2.0	.	3	1	1	48
3.6	6.6	5	1	1	49