Evaluation of Multivariate Classification Models for Analyzing NMR Metabolomics Data

Thao Vu,[1] Parker Siemek,[2] Fatema Bhinderwala,[2,3] Yuhang Xu[1] and Robert Powers[2,3,*]

[1]*Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE USA 68583-0963*
[2]*Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE USA 68588-0304*
[3]*Nebraska Center for Integrated Biomolecular Communication*

Table of Contents:

Table S1A: Pairwise ANOVA p-values calculated from the classification accuracy plots **(Figures 3A1-A3).**

	OPLS	**PLS**	**SVM**	**RF**
(A1)				
PLS	0.8008			
SVM	0.8297	0.998		
RF	0	0	0	
PC-LDA	0.0026	0.0036	0.0061	0.0027
(A2)				
PLS	0.0101			
SVM	0.1783	0.3605		
RF	0.0001	0.005	0.0074	
PC-LDA	0.0013	0.0256	0.0083	0.4982
(A3)				
PLS	0.0092			
SVM	0.0245	0.1571		
RF	0.159	0.0043	0.0126	
PC-LDA	0.4159	0.0479	0.0618	0.5751

Table S1B: Pairwise ANOVA p-values calculated from the classification accuracy plots **(Figures 3B1-B3).**

	OPLS	**PLS**	**SVM**	**RF**
(B1)				
PLS	0.3786			
SVM	0.0494	0.1045		
RF	0	0	0	
PC-LDA	0	0.0001	0.0001	0.0323
(B2)				
PLS	0.0492			
SVM	0.1036	0.5519		
RF	0	0.0002	0.0001	
PC-LDA	0.0014	0.0089	0.0065	0.1528
(B3)				
PLS	0.0004			
SVM	0.0044	0.2955		
RF	0.2528	0.0005	0.003	
PC-LDA	0.007	0.0003	0.0021	0.0004

Table S2A: Pairwise ANOVA p-values calculated from the AUROC plots **(Figures 4A1-A3).**

	OPLS	**PLS**	**SVM**	**RF**
(A1)				
PLS	0.5246			
SVM	0.6347	0.0922		
RF	0	0.0001	0.0001	
PC-LDA	0.0172	0.0333	0.0241	0.7923
(A2)				
PLS	0.0005			
SVM	0.001	0.6887		
RF	0	0.0001	0	
PC-LDA	0.0016	0.0045	0.0033	0.3599
(A3)				
PLS	0.0046			
SVM	0.0657	0.0109		
RF	0.9193	0.0661	0.2382	
PC-LDA	0.2472	0.5964	0.9555	0.0973

Table S2B: Pairwise ANOVA p-values calculated from the AUROC plots **(Figures 4B1-B3).**

AUROC	**OPLS**	**PLS**	**SVM**	**RF**
(B1)				
PLS	0.1302			
SVM	0.0413	0.0438		
RF	0.0006	0.0004	0.0013	
PC-LDA	0.0019	0.0013	0.0026	0.3616
(B2)				
PLS	0.0177			
SVM	0.1797	0.0383		
RF	0	0	0.0001	
PC-LDA	0.0024	0.0021	0.0027	0.7929
(B3)				
PLS	0.0006			
SVM	0.0212	0		
RF	0.6259	0.0038	0.0436	
PC-LDA	0.1092	0.0016	0.0214	0.2656

Table S3A: Pairwise ANOVA p-values calculated from the *RMSE* plots **(Figures 5A1-A3).**

RMSE	**OPLS**	**PLS**	**SVM**	**RF**
(A1)				
PLS	0.0027			
SVM	0.0058	0.0158		
RF	0.0059	0.0175	0.1703	
PC-LDA	0.0019	0.0006	0.0596	0.069
(A2)				
PLS	0.0076			
SVM	0.0138	0.0217		
RF	0.039	0.1014	0.0509	
PC-LDA	0.005	0.0031	0.1491	0.6717
(A3)				
PLS	0.0147			
SVM	0.2296	0.0526		
RF	0.0339	0.0657	0.0193	
PC-LDA	0.0065	0.0085	0.0029	0.7813

Table S3B: Pairwise ANOVA p-values calculated from the *RMSE* plots **(Figures 5B1-B3).**

RMSE	OPLS	PLS	SVM	RF
(B1)				
PLS	0.0026			
SVM	0.0053	0.0148		
RF	0.2114	0.9189	0.38	
PC-LDA	0.0037	0.0094	0.0119	0.1301
(B2)				
PLS	0.0036			
SVM	0.0014	0.0293		
RF	0.2347	0.6381	0.44	
PC-LDA	0.0056	0.0096	0.0136	0.336
(B3)				
PLS	0.0135			
SVM	0.0128	0.0038		
RF	0.4578	0.8312	0.8583	
PC-LDA	0.0109	0.0178	0.0193	0.3915

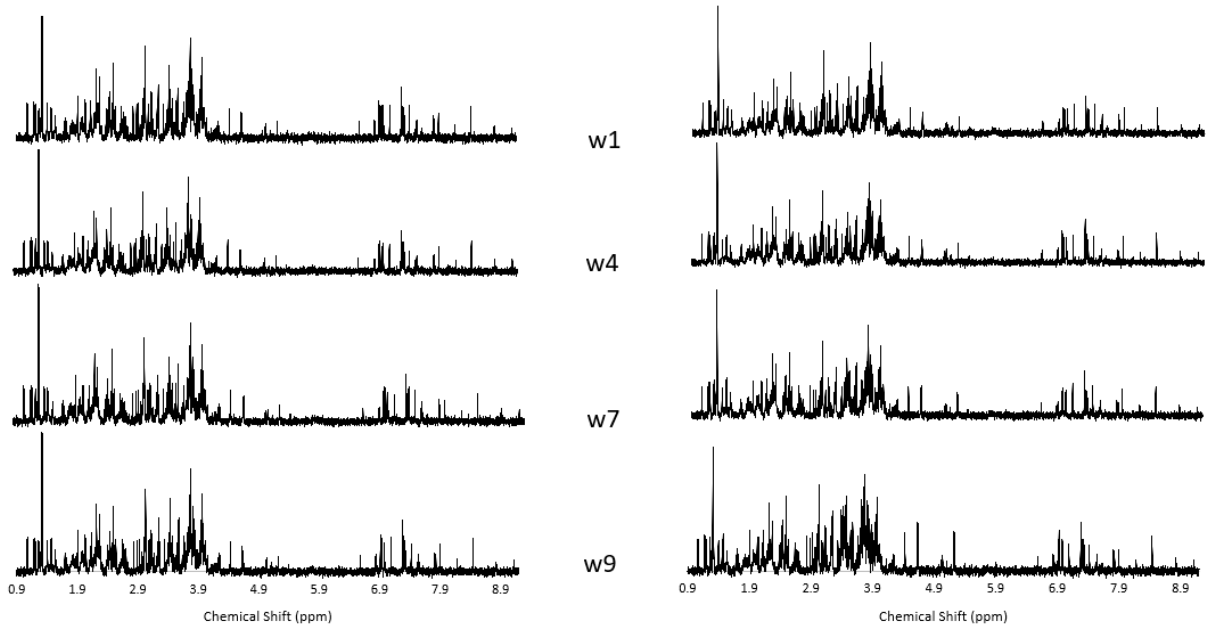

Figure S1. Representative 1D ^{1}H NMR spectra for simulated data set. Representative simulated spectra for group 1 (*left*) and group 2 (*right*) at different group separations w1, w4, w7, and w9 respectively, as shown in **Table 2**.

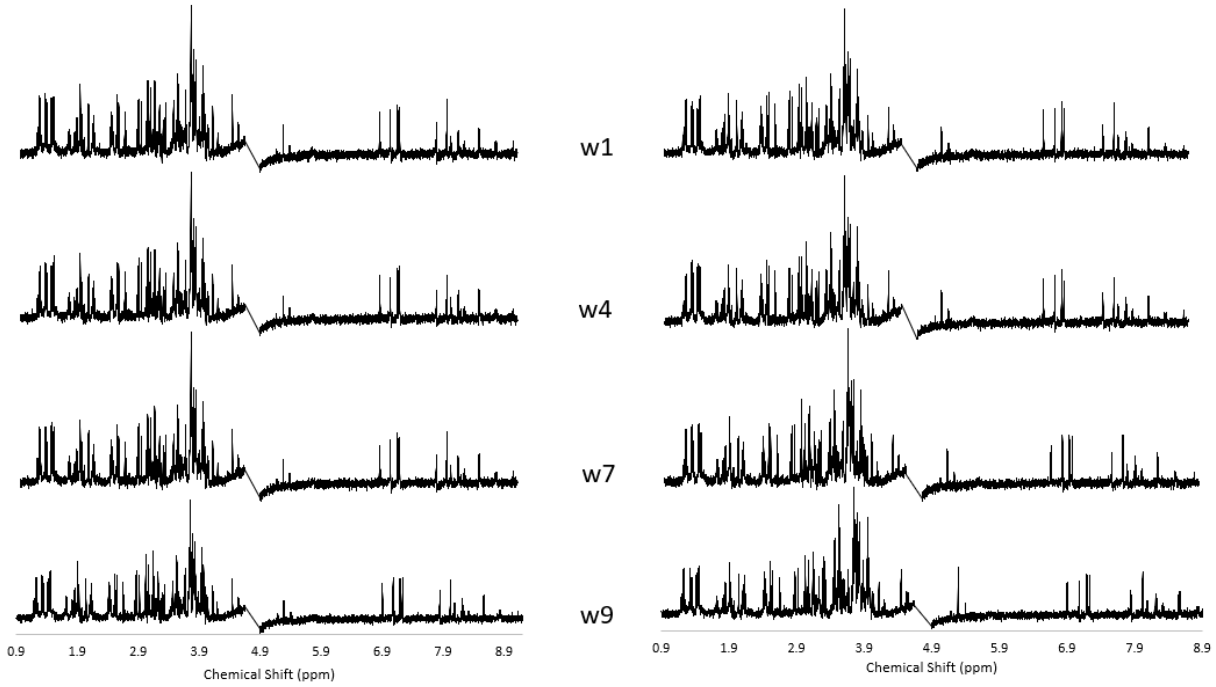

Figure S2. Representative 1D ^{1}H NMR spectra for experimental data set. Representative experimental spectra for group 1 (*left*) and group 2 (*right*) at different group separations w1, w4, w7, and w9 respectively, as shown in **Table 2**.

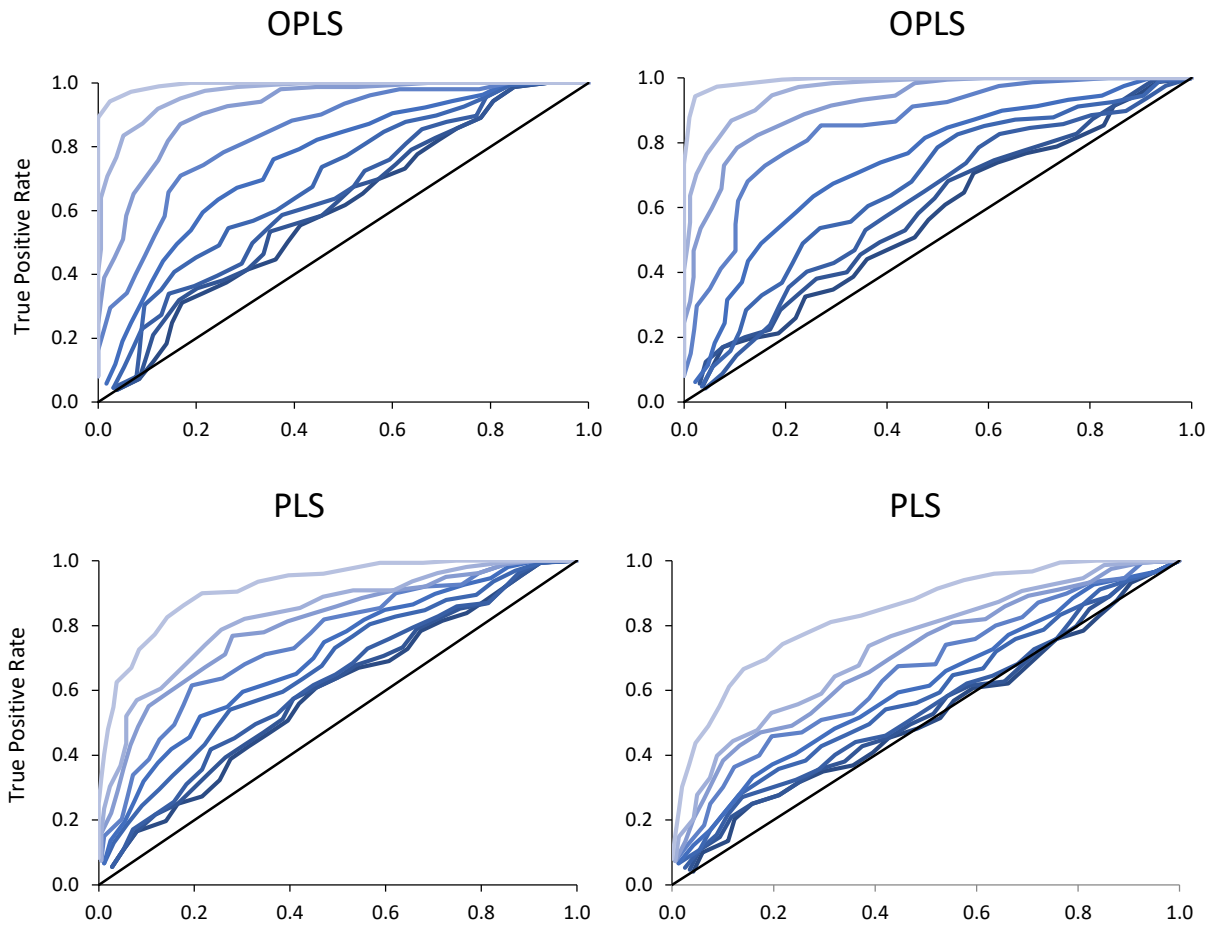

Figure S3. ROC curves from OPLS or PLS models. A set of ROC curves are plotted using simulated (*left*) and experimental (*right*) datasets for scenario 3. Each plot presents nine different ROC curves corresponding to group separations w1 to w9 as shown in **Table 2**. As group separation increases from w1 to w9, the ROC curve color changes from darker blue to lighter blue. Black diagonal line represents classification based on random guessing.

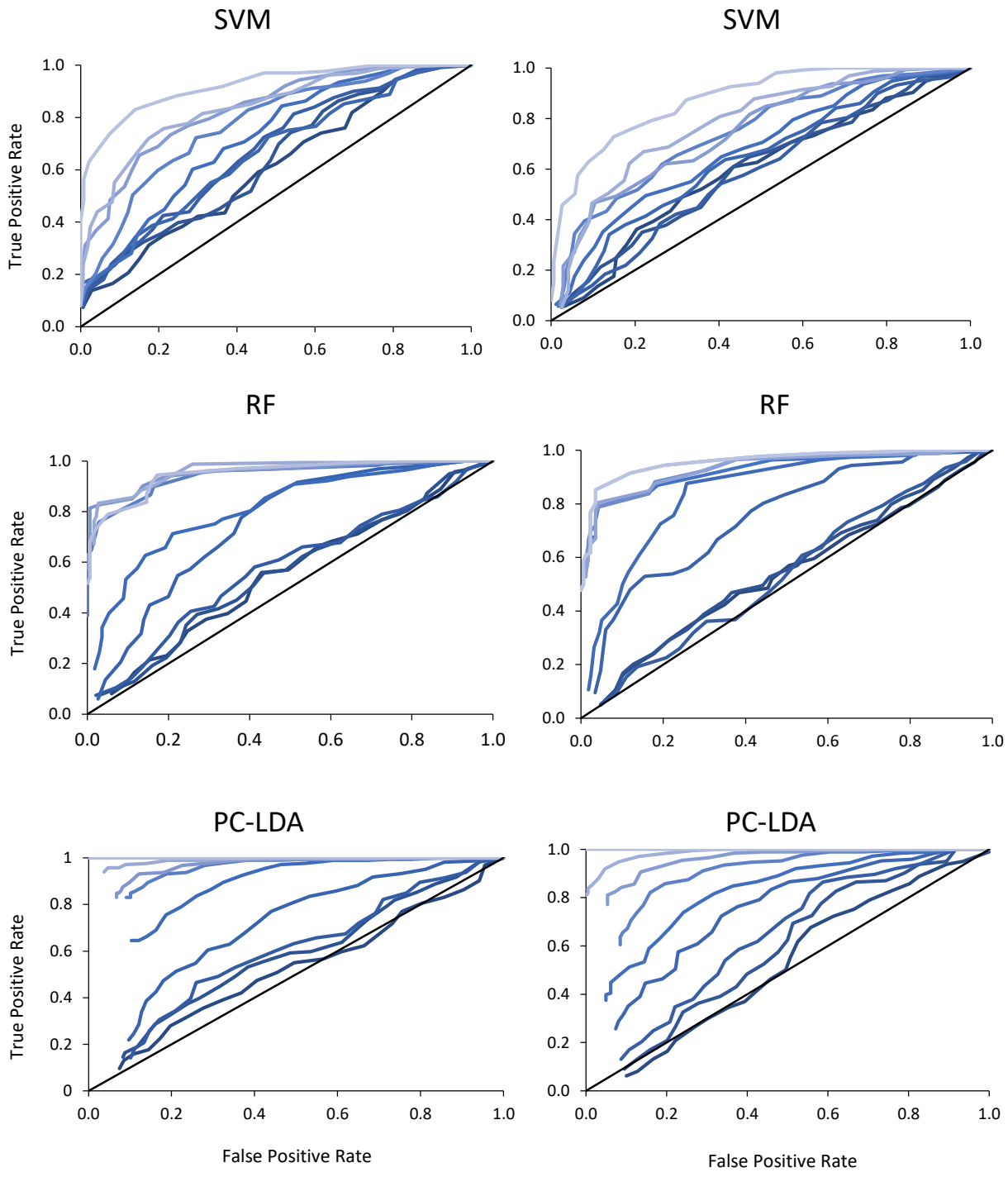

Figure S4. ROC curves from SVM, RF or PC-LDA models. A set of ROC curves are plotted using simulated (*left*) and experimental (*right*) datasets for scenario 3. Each plot presents nine different ROC curves corresponding to group separations w1 to w9 as shown in **Table 2**. As group separation increases from w1 to w9, the ROC curve color changes from darker blue to lighter blue. Black diagonal line represents classification based on random guessing.