1. Mutual information. We are going to investigate the use of the mutual information criterion to evaluate a set of candidate features and to select an informative subset to be used as input data for a typical classifier.

First of all, you should select a suitable subset of variables and plot those with higher mutual information. Are you able to distinguish the three types of wine?

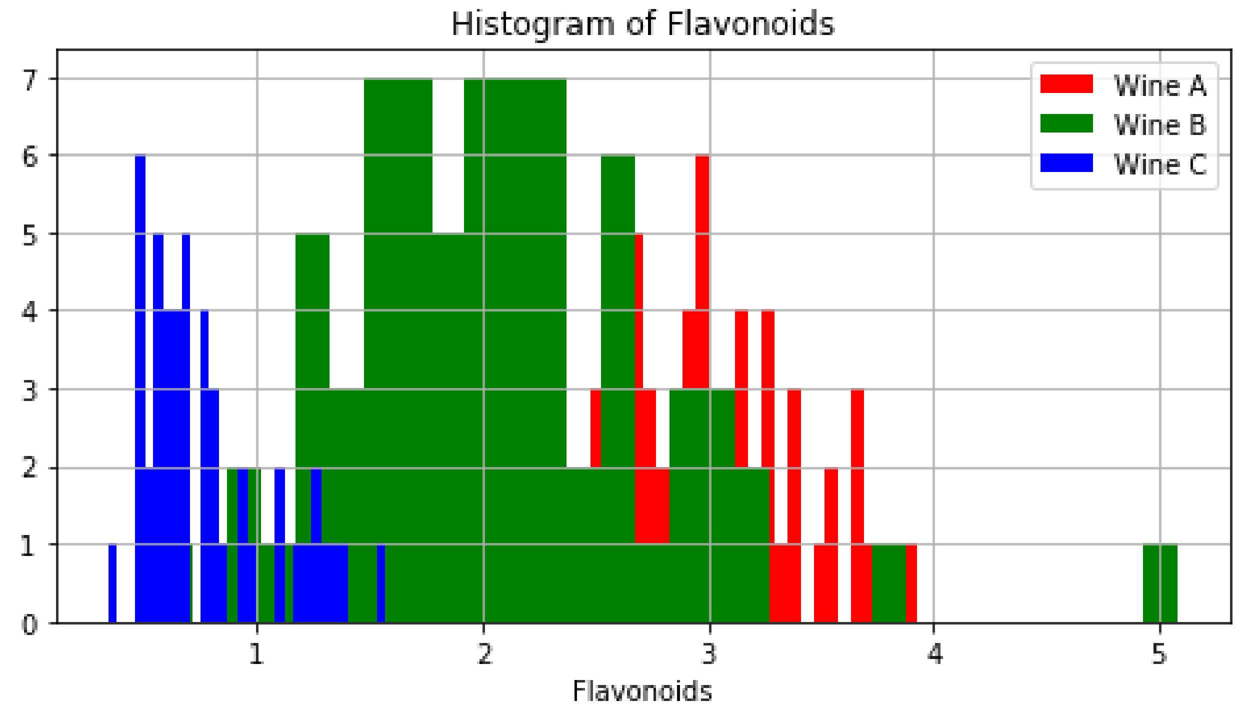
Los datos se agruparon de tal manera que se formaron grupos de entre 2 y 9 clusters para comprobar cómo la información mutual variaba a lo largo de diferentes números de clusters.

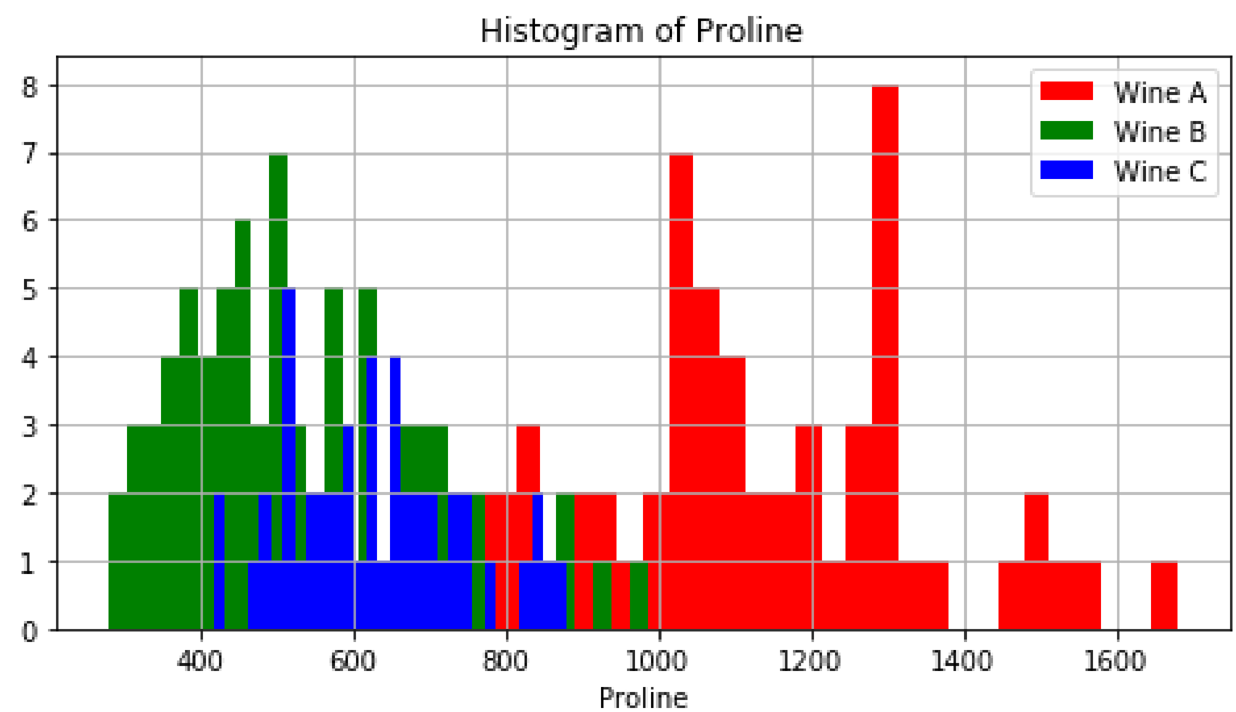
	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols	Flavanoids	Nonflavanoid.phenols	Proanth	Color.int	Hue	OD	Proline
2	0.31	0.13	0.04	0.14	0.05	0.32	0.41	0.14	0.12	0.19	0.25	0.41	0.36
3	0.38	0.20	0.11	0.20	0.10	0.34	0.57	0.15	0.18	0.43	0.35	0.43	0.47
4	0.41	0.19	0.12	0.18	0.16	0.37	0.57	0.17	0.22	0.44	0.39	0.47	0.46
5	0.41	0.31	0.12	0.21	0.18	0.37	0.60	0.18	0.23	0.55	0.43	0.49	0.54
6	0.44	0.29	0.11	0.22	0.18	0.39	0.63	0.17	0.20	0.52	0.45	0.50	0.54
7	0.45	0.31	0.12	0.25	0.18	0.40	0.67	0.17	0.23	0.50	0.46	0.51	0.55
8	0.46	0.32	0.13	0.25	0.20	0.40	0.69	0.18	0.24	0.51	0.46	0.52	0.54
9	0.46	0.34	0.14	0.21	0.21	0.41	0.70	0.19	0.25	0.52	0.45	0.51	0.55

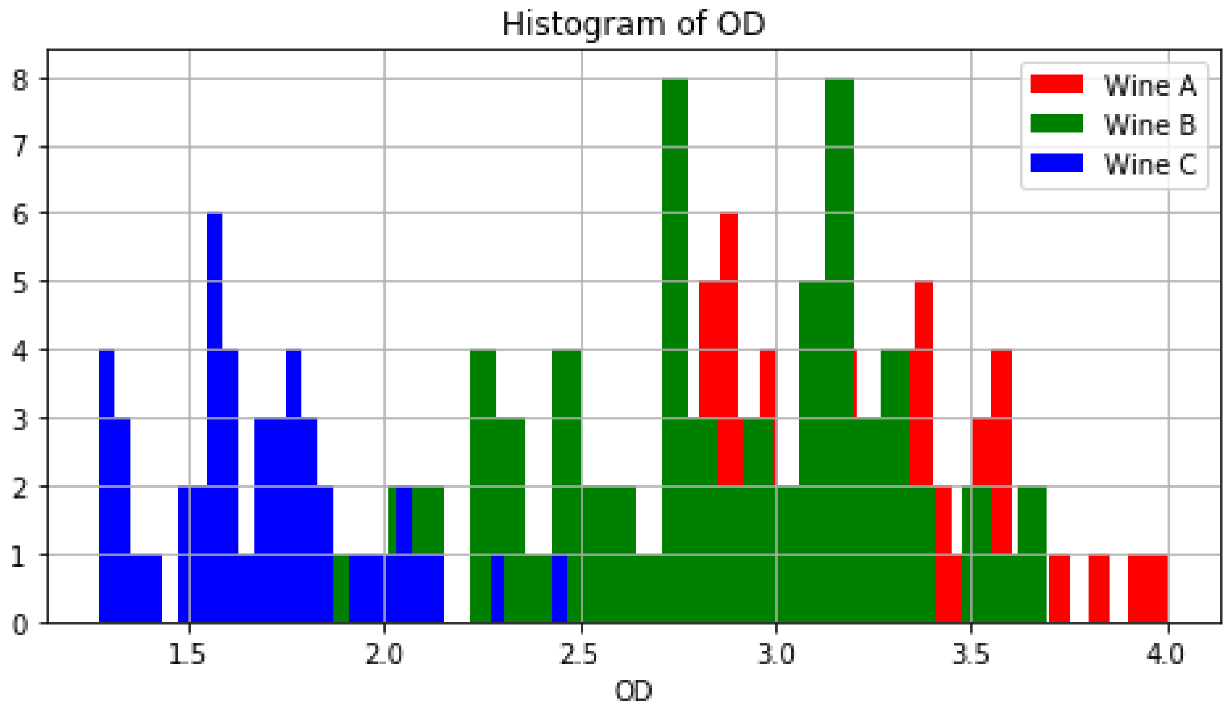
La descripción de la información mutual se ha tenido en cuenta para decider el número de clusters en los que hay que divider los datos. Dado que los datos no varían lo suficiente cuando se contempla la deviación estandar se puede asumir que *mean* es un buen indicador para elegir el major subset de variables.

	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols	Flavanoids	Nonflavanoid.phenols	Proanth	Color.int	Hue	OD	Proline
count	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00
mean	0.41	0.26	0.11	0.21	0.16	0.37	0.60	0.17	0.21	0.46	0.41	0.48	0.50
std	0.05	0.08	0.03	0.04	0.05	0.03	0.09	0.02	0.04	0.11	0.07	0.04	0.07
min	0.31	0.13	0.04	0.14	0.05	0.32	0.41	0.14	0.12	0.19	0.25	0.41	0.36
25%	0.40	0.20	0.11	0.20	0.15	0.36	0.57	0.16	0.20	0.44	0.38	0.46	0.46
50%	0.43	0.30	0.12	0.21	0.18	0.38	0.61	0.17	0.22	0.50	0.44	0.50	0.54
75%	0.45	0.32	0.13	0.23	0.19	0.40	0.67	0.18	0.23	0.52	0.45	0.51	0.54
max	0.46	0.34	0.14	0.25	0.21	0.41	0.70	0.19	0.25	0.55	0.46	0.52	0.55

Las mayores correlaciones se observaron en Flavanoids, Proline, OD, Color.int, Alcohol y Hue. Mirando la representación gráfica de las tres mayores, se puede observer que las clases están bien separadas.


Histogram of Flavonoids
Wine A
Wine B
Wine C
0
1
2
3
4
5
6
7
1
2
3
4
5
Flavonoids


Histogram of Proline
Wine A
Wine B
Wine C
0
1
2
3
4
5
6
7
8
400
600
800
1000
1200
1400
1600
Proline

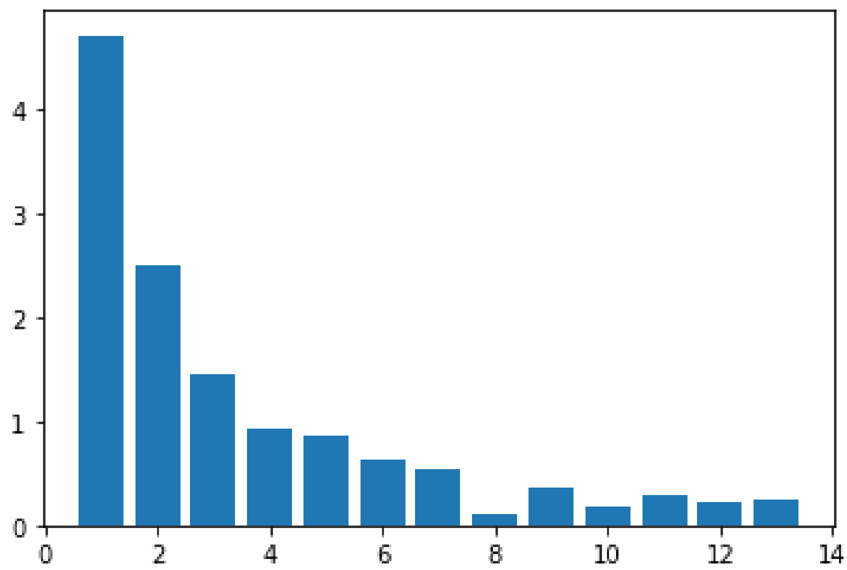


2. Chi-Square. Repeat the selection of variables with Chi-Square method. Do you get the same results as with the previous one?

No, puesto que chi-square solo se puede usar en datos categóricos.

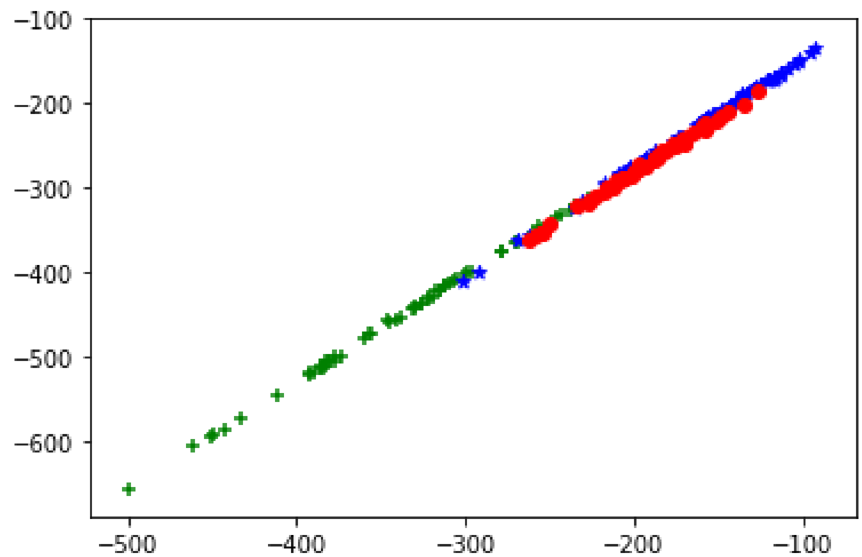
3. Principal Components Analysis (PCA). Now we are going to work with PCA as a method for dimensionality reduction. The Principal Component Analysis (PCA) was independently proposed by Karl Pearson (1901) and Harold Hotelling (1933) to turn a set of possibly correlated variables into a smaller set of uncorrelated variables. The idea is that a high-dimensional dataset is often described by correlated variables and therefore only a few meaningful dimensions account for most of the information. The PCA method finds those directions in the original dataset that account for the greatest variance in data, also called the principal components.

a) PCA without normalization:
   a. Calculate the eigenvalues and plot them. How many components do you need to explain 90% of the total variance?



Considerando la gráfica y calculando la contribución de cada componente, se necesitaría al menos 8 componentes para expresar el 90% de los datos.

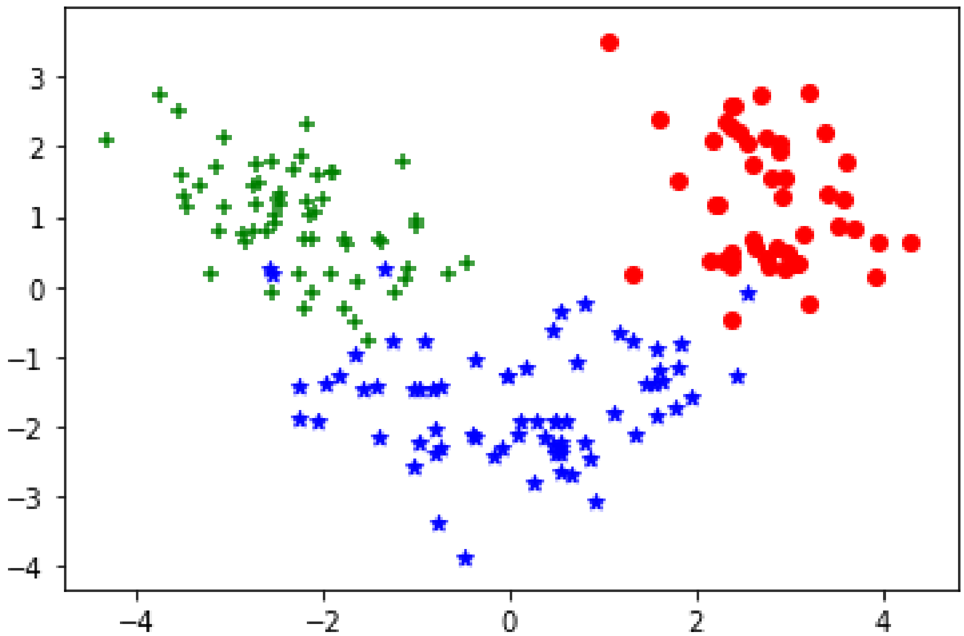
b. Plot the two first components, are the resulting clusters clearly separated?



No, los clusters están mezclados.

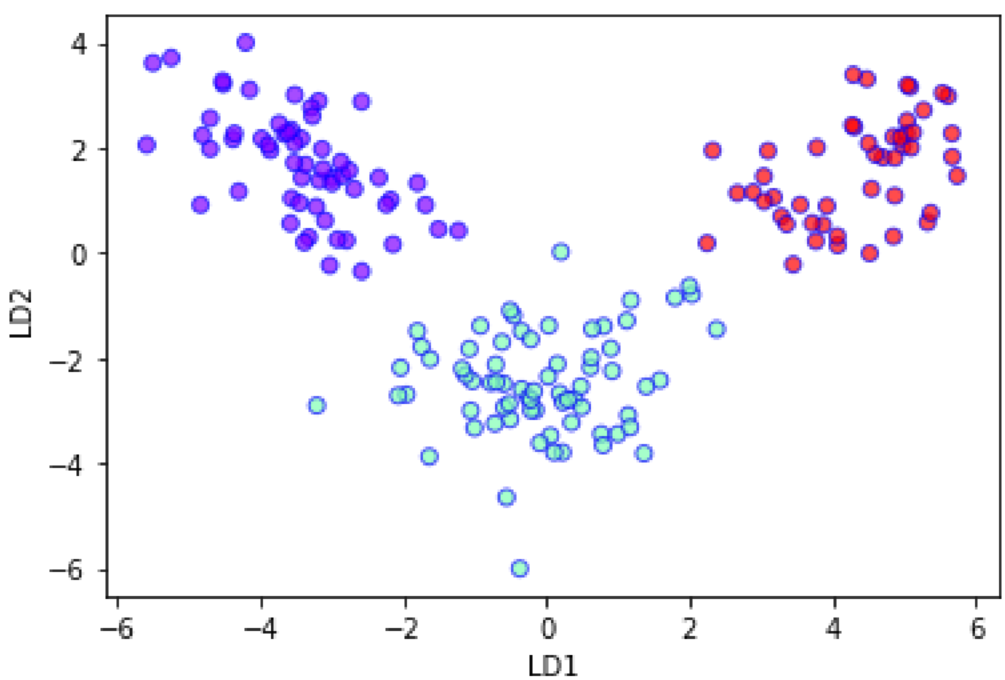
b) PCA with normalization: Repeat the two previous steps but in this case scaling the input to zero mean and unit variance N(0,1), it is also called z-scores. What do you see now? In our dataset, why does PCA without normalization perform poor?

Los eigenvalues son iguales, pero las gráficas son completamente diferentes En este caso las clases están bien definidas.



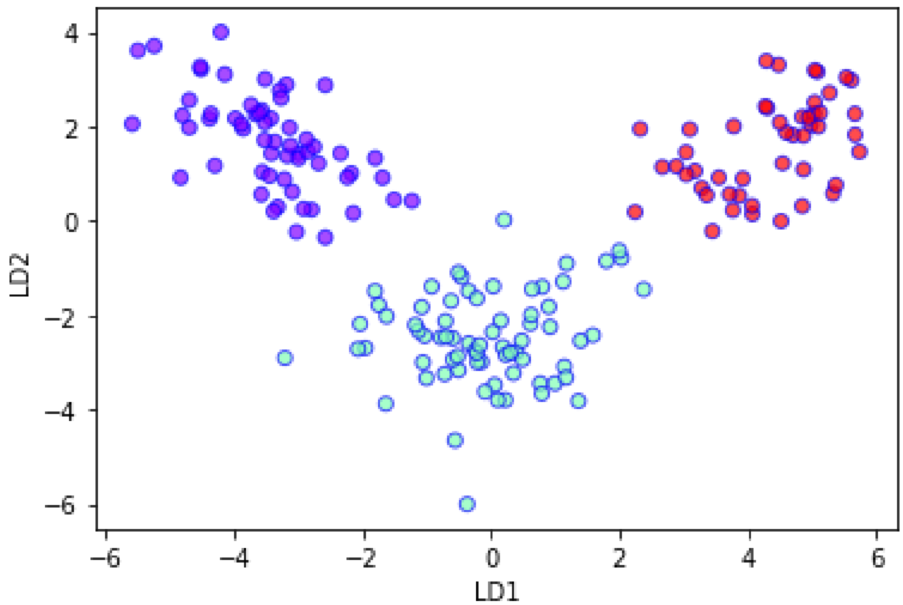
4. Linear Discriminant Analysis (LDA). What we aim for is a projection that maintains the maximum discriminative power of a given dataset, so a method should make use of class labels (if they are known a priori). The Linear Discriminant Analysis, invented by R. A. Fisher (1936), does so by maximizing the between-class scatter, while minimizing the within-class scatter at the same time.

a) LDA without normalization: Calculate the two components (C-1) and plot them, are the resulting clusters clearly separated?



Sí, están claramente separadas.

b) LDA with normalization: Repeat the previous step, what do you see now? Which is the difference with the previous one?



Tanto si se normaliza como si no, la separación es la misma. Por lo tanto, en LDA la normalización es irrelevante.

5. Logistic Regression and Model Evaluation
From the wine dataset, we are going to study the importance (or not) of reducing dimensionality. We are going to apply logistic regression as a predictive model and see the influence of increasing the number of predictor variables. In the wine dataset the dependent variable is not binary, therefore we need to perform a study two to two classes: 1 vs. 2, 1 vs. 3 and 2 vs. 3

Considering the cases as 1 vs 2, 1 vs 3 and 2 vs 3, respectively, we got:

Logistic Regression	Accuracy	Precision	Sensitivity	Specificity	AUC-ROC
			1 vs 2		
Full data	0.97	0.97	0.947	1.0	0.97
			MI		
1 var.	0.82	0.82	0.95	0.7	0.82
2 vars.	0.95	0.95	1.0	0.9	0.95
3 vars.	0.95	0.95	1.0	0.9	0.95
4 var.	0.92	0.92	1.0	0.85	0.92
5 vars.	0.92	0.92	1.0	0.85	0.92
6 vars.	0.95	0.95	1.0	0.9	0.95
7 vars.	0.95	0.95	1.0	0.9	0.95

8 vars.	0.97	0.97	1.0	0.95	0.98
9 vars.	0.97	0.97	1.0	0.95	0.98
10 vars.	0.95	0.95	1.0	0.9	0.95
11 vars.	0.97	0.97	1.0	0.95	0.98
12 vars.	0.95	0.95	1.0	0.9	0.95
13 vars.	0.95	0.95	1.0	0.9	0.95
PCA					
1 component	0.9	0.9	1.0	0.8	0.9
2 components	0.9	0.9	1.0	0.8	0.9
3 components	0.9	0.9	1.0	0.8	0.9
4 components	0.97	0.97	1.0	0.95	0.98
5 components	0.97	0.97	1.0	0.95	0.98
6 components	0.92	0.95	1.0	0.9	0.95
7 components	0.95	0.95	1.0	0.9	0.95
8 components	0.95	0.95	1.0	0.9	0.95
9 components	0.97	0.97	1.0	0.95	0.98
10 components	0.97	0.97	1.0	0.95	0.98
11 components	0.95	0.97	1.0	0.95	0.98
12 components	1.0	1.0	1.0	1.0	1.0
13 components	0.95	0.95	1.0	0.9	0.95
LDA					
1 component	1.0	1.0	1.0	1.0	1.0
	Accuracy	Precision	Sensitivity	Specificity	AUC-ROC
1 vs 3					
Full data	1.0	1.0	1.0	1.0	1.0
MI					
1 var.	1.0	1.0	1.0	1.0	1.0
2 vars.	1.0	1.0	1.0	1.0	1.0
3 vars.	1.0	1.0	1.0	1.0	1.0
4 var.	1.0	1.0	1.0	1.0	1.0
5 vars.	1.0	1.0	1.0	1.0	1.0
6 vars.	1.0	1.0	1.0	1.0	1.0
7 vars.	1.0	1.0	1.0	1.0	1.0
8 vars.	1.0	1.0	1.0	1.0	1.0
9 vars.	1.0	1.0	1.0	1.0	1.0
10 vars.	1.0	1.0	1.0	1.0	1.0
11 vars.	1.0	1.0	1.0	1.0	1.0
12 vars.	1.0	1.0	1.0	1.0	1.0
13 vars.	1.0	1.0	1.0	1.0	1.0
PCA					
1 component	1.0	1.0	1.0	1.0	1.0
2 components	1.0	1.0	1.0	1.0	1.0
3 components	1.0	1.0	1.0	1.0	1.0
4 components	1.0	1.0	1.0	1.0	1.0
5 components	1.0	1.0	1.0	1.0	1.0

6 components	1.0	1.0	1.0	1.0	1.0
7 components	1.0	1.0	1.0	1.0	1.0
8 components	1.0	1.0	1.0	1.0	1.0
9 components	1.0	1.0	1.0	1.0	1.0
10 components	1.0	1.0	1.0	1.0	1.0
11 components	1.0	1.0	1.0	1.0	1.0
12 components	1.0	1.0	1.0	1.0	1.0
13 components	1.0	1.0	1.0	1.0	1.0
LDA					
1 component	1.0	1.0	1.0	1.0	1.0
	Accuracy	Precision	Sensitivity	Specificity	AUC-ROC
2 vs 3					
Full data	0.94	0.94	0.91	1.0	0.96
MI					
1 var.	0.97	0.97	1.0	0.92	0.96
2 vars.	0.94	0.94	0.96	0.92	0.94
3 vars.	0.97	0.97	0.96	1.0	0.98
4 var.	0.97	0.97	0.96	1.0	0.98
5 vars.	0.92	0.92	0.87	1.0	0.93
6 vars.	0.92	0.92	0.87	1.0	0.93
7 vars.	0.92	0.92	0.87	1.0	0.93
8 vars.	0.94	0.94	0.91	1.0	0.96
9 vars.	0.92	0.92	0.87	1.0	0.93
10 vars.	0.92	0.92	0.87	1.0	0.93
11 vars.	0.92	0.92	0.87	1.0	0.93
12 vars.	0.92	0.92	0.87	1.0	0.93
13 vars.	0.94	0.94	0.91	1.0	0.96
PCA					
1 component	0.94	0.97	0.96	1.0	0.98
2 components	0.97	0.97	0.96	1.0	0.98
3 components	0.97	0.97	0.96	1.0	0.98
4 components	0.94	0.94	0.91	1.0	0.96
5 components	0.94	0.94	0.91	1.0	0.96
6 components	0.94	0.94	0.91	1.0	0.96
7 components	0.94	0.94	0.91	1.0	0.96
8 components	0.94	0.94	0.91	1.0	0.96
9 components	0.92	0.94	0.91	1.0	0.96
10 components	0.94	0.94	0.91	1.0	0.96
11 components	0.92	0.94	0.91	1.0	0.96
12 components	0.92	0.94	0.91	1.0	0.96
13 components	0.92	0.94	0.91	1.0	0.96
LDA					
1 component	1.0	1.0	1.0	1.0	1.0