Dataset: Lung Cancer Data Set

Dataset link: https://archive.ics.uci.edu/ml/datasets/Lung+Cancer

Data About Data :

Data Set Characteristics:	Multivariate	Number of Instances:	32	Area:	Life
Attribute Characteristics:	Integer	Number of Attributes:	56	Date Donated	1992-05-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	424713

Data Set Information:

This data was used by Hong and Young to illustrate the power of the optimal discriminant plane even in ill-posed settings. Applying the KNN method in the resulting plane gave 77% accuracy. However, these results are strongly biased (See Aeberhard's second ref. above, or email to stefan '@' coral.cs.jcu.edu.au). Results obtained by Aeberhard et al. are :

RDA : 62.5%, KNN 53.1%, Opt. Disc. Plane 59.4%

The data described 3 types of pathological lung cancers. The Authors give no information on the individual variables nor on where the data was originally used.

Notes:

- In the original data 4 values for the fifth attribute were -1. These values have been changed to "?" (unknown). (*)

- In the original data 1 value for the 39 attribute was 4. This value has been changed to "?" (unknown). (*)

Attribute Information:

Attribute 1 is the class label.

All predictive attributes are nominal, taking on integer values 0-3