

LAB 2(B)

Name: Swarad Ganesh Gat

SBU ID: 114833402

Dataset: <https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset>

Description: The dataset contains numerical data regarding breast cancer. It contains 31 attributes, which describe the various properties of the tumour. The label of this dataset is 'diagnosis', which has two labels. The tumour is either malignant or benign, malignant meaning that it is cancerous and benign meaning that it is non-cancerous or harmless. There are 31 attributes which describe the tumour, including 'radius_mean', 'texture_mean', 'area_mean', etc.

Snapshot of data:

Table data was imported. Adjust									
id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	
84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	
84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	
84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	
843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	
844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	
84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	
844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	
84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	
845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	
84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	
846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	

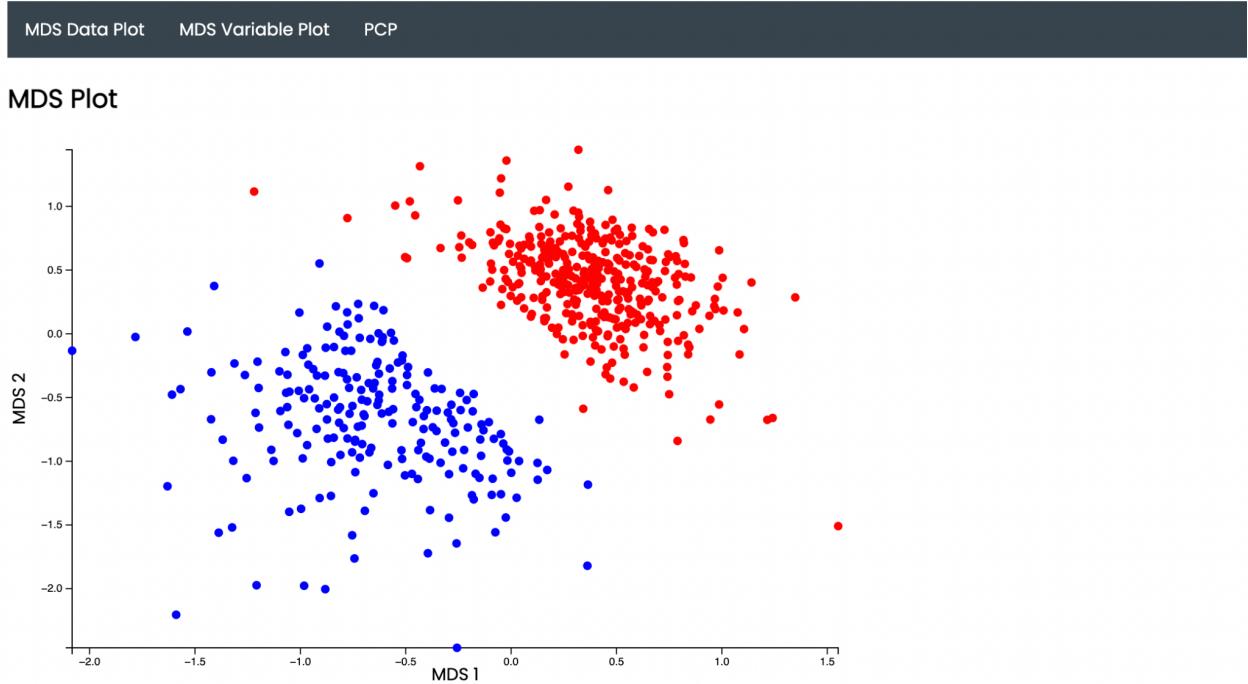
I have performed Multi Dimensional Scaling on this dataset and visualised the results.

Visualisations I have performed include:

- MDS Data Plot
- MDS Variables Plot (With User Interaction + PCP)
- PCP

Visualizations:

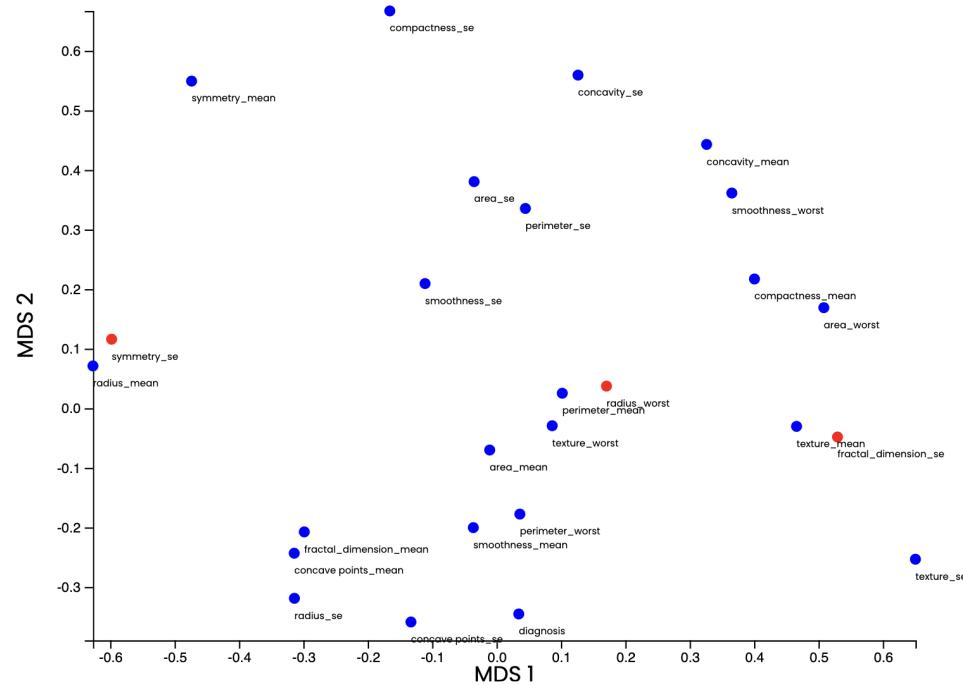
1) MDS Data Plot:



→ I have a dataset with 31 attributes. After performing MDS using sklearn (`n_components = 2`), I rendered the data from Flask to the HTML file. I've visualised the results usind D3.js. The axes represent the 2 components. The colors represent the clusters. I have performed K-Means on the N-Dimensional space and visualised the clusters here.

2) MDS Variable Plot:

MDS Variable Plot

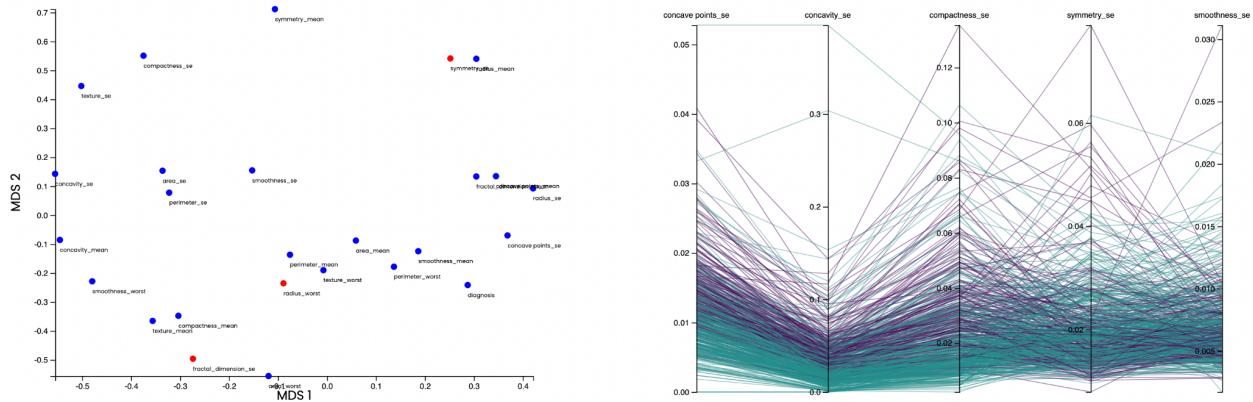


→ This represents the MDS Variables' plot. Each point on the scatterplot represents a numerical attribute. I have used $1-|\text{corr}|$ as the distance matrix instead of the euclidian distance. I have added an interaction element on each scatterplot circle. On clicking 5 different attributes on the scatterplot, a PCP is generated besides it.

- User Interaction in MDS Variables' plot:

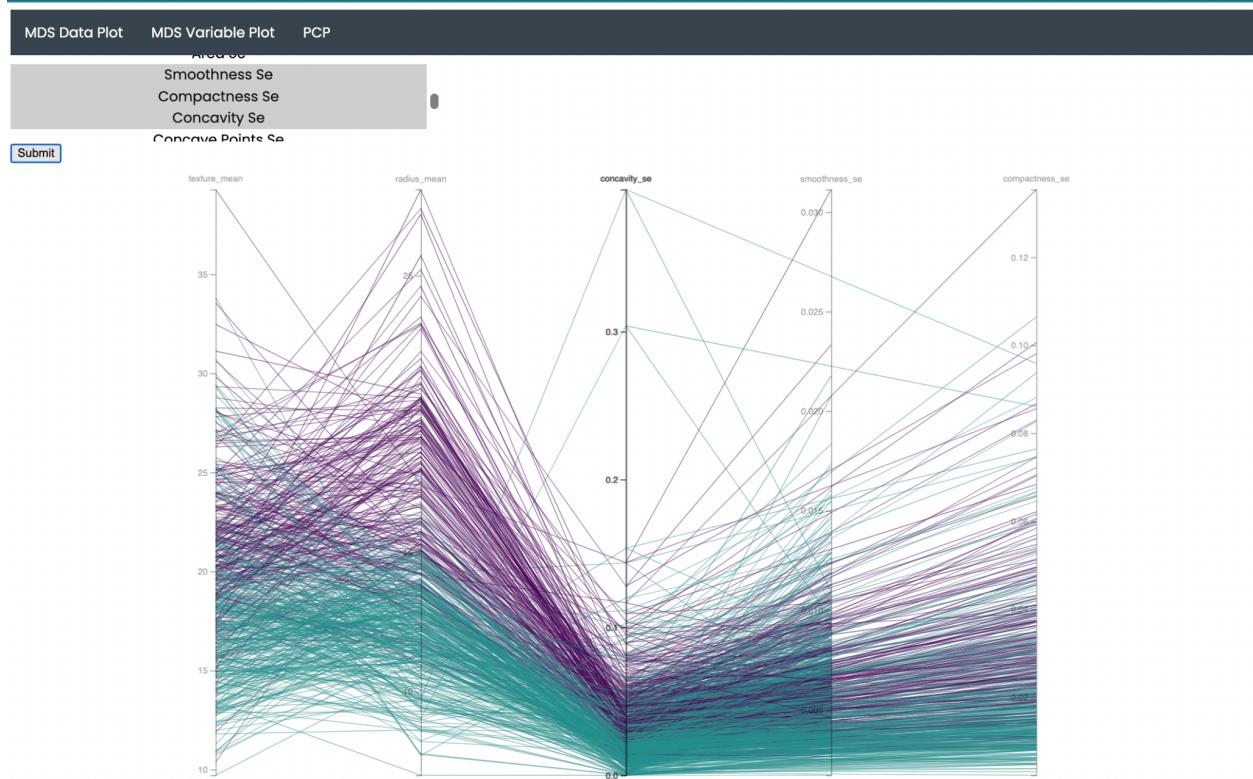


MDS Variable Plot



→ After clicking on the attributes, a PCP is generated besides it, as can be seen in the above image. The PCP is coloured according to the K-means labels.

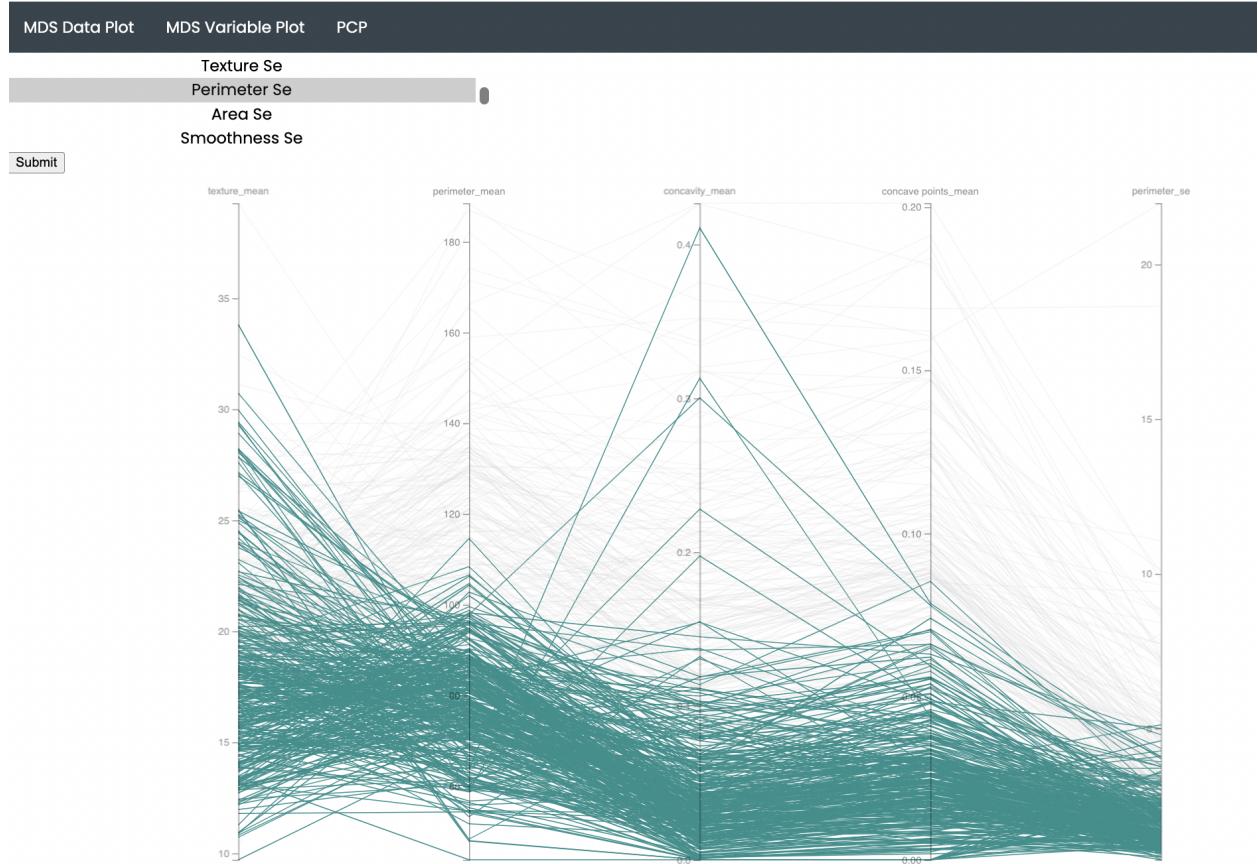
3) Parallel Coordinates Plot:



→ I have given the user the choice to select the variables. The order in which the user selects the attributes is how the PCP is generated. The user can select any number of attributes. The

user can hover over an axis to see the values clearly. They can also hover over a given cluster and can see only that clusters' values.

Hovering Effect:



→ **CONCLUSION:** Thus, I have successfully implemented MDS (Data and Variable) and visualised it as a scatterplot. I added user interaction in the MDS Variables' plot where the user selects 5 attributes and a corresponding PCP is plotted which includes the selected attributes as axes. I have also implemented PCP where the axes can be selected in any order from a given dropdown menu. The order in which the user selects the attributes is how they are displayed in the PCP.