

Unraveling Multicollinearity between Predictors with PCLR and PLSLR Techniques

Bindubritta Acharjee

PROBLEM STATEMENT

THE ISSUES

- ❖ The usual dimension reduction techniques avoid multicollinearity
- ❖ They undermine the statistical significance of independent variables



PROBLEM

THE RESOLUTIONS

- ❖ Need to follow Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) to solve this kind of situation



SOLUTION

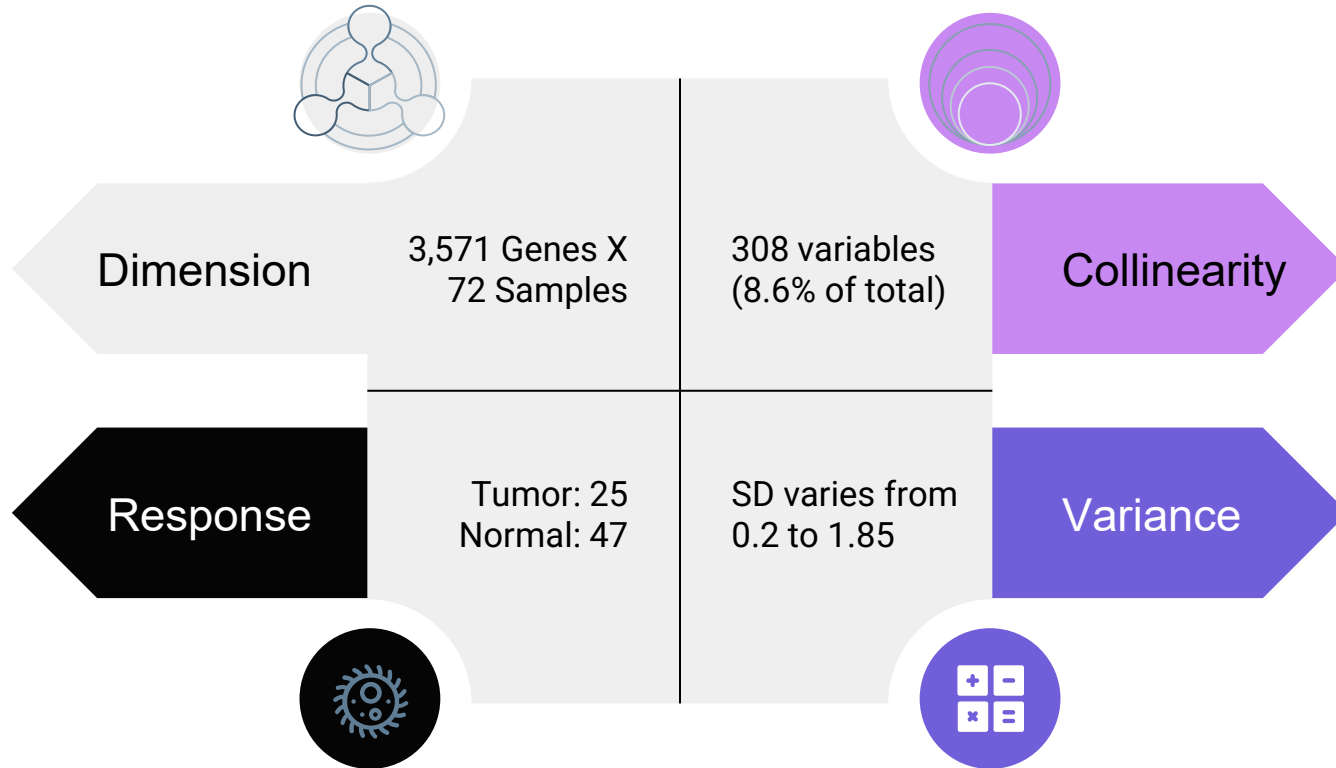
THE OUTCOMES

- ❖ Better prediction model with fewer components for regression analysis

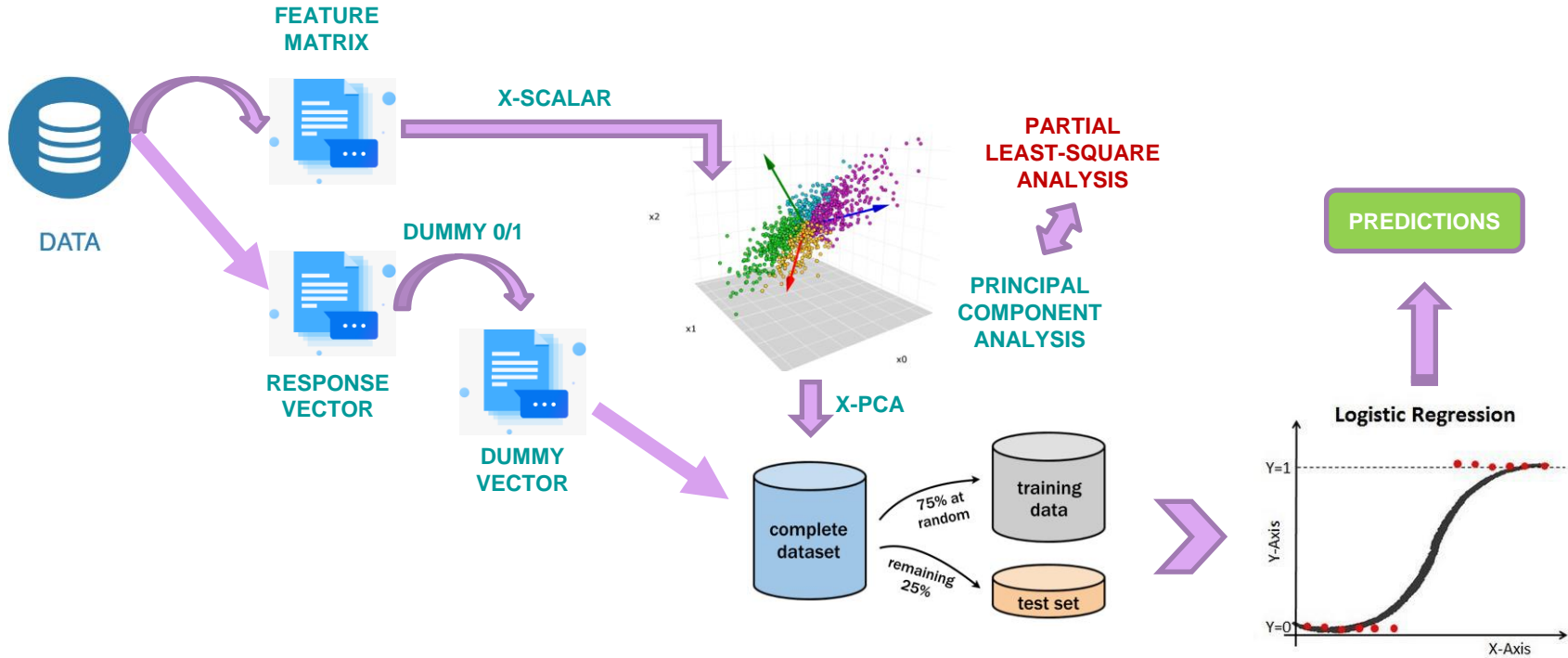


RESULT

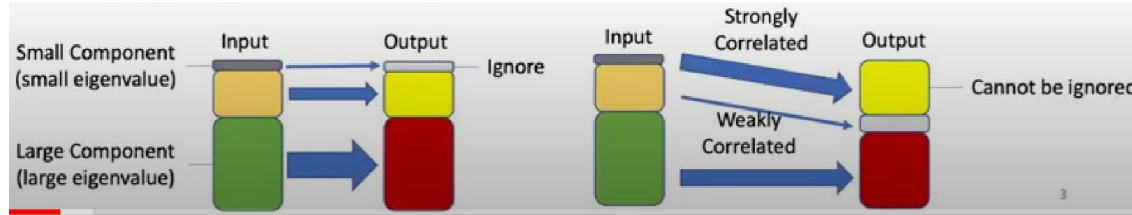
Tumor Dataset for Analysis



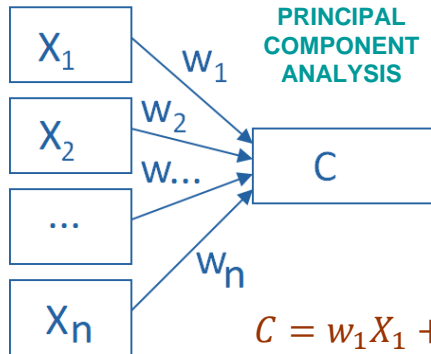
PCLR Method



PCLR vs PLSLR \Rightarrow PCA vs PLS



$$y_{n \times 1} = X_{n \times (p+1)} \cdot \beta_{(p+1) \times 1} + \varepsilon_{n \times 1}$$



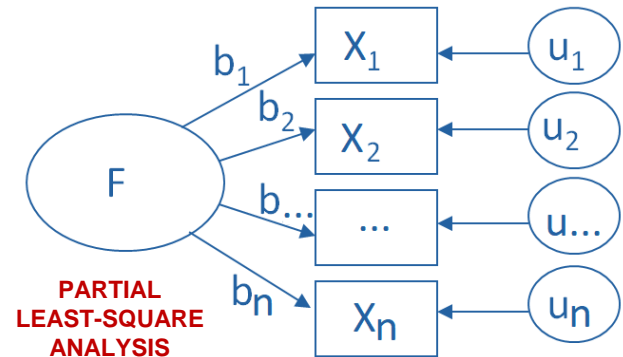
$$C = w_1X_1 + w_2X_2 + \dots + w_nX_n$$

$$X_1 = (b_1 \times F) + u_1$$

$$X_2 = (b_2 \times F) + u_2$$

\vdots

$$X_n = (b_n \times F) + u_n$$



Analysis and Results

Fig. 1
Scree Plot for PCA

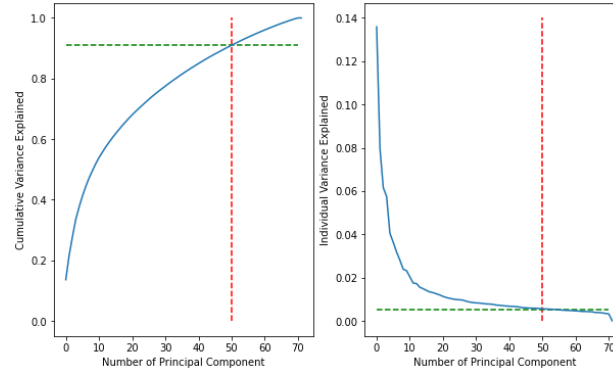


Fig. 2
KPIs vs cut-off points
for PCLR Model

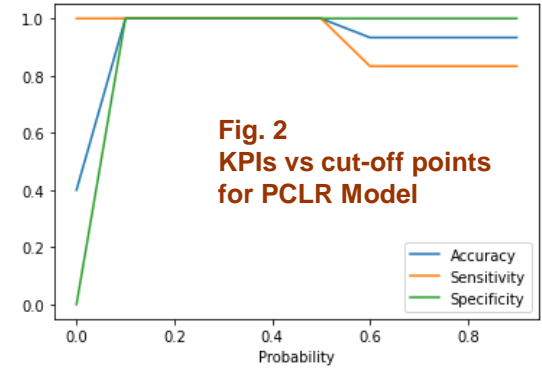


Fig. 3
Scree Plot for PLS

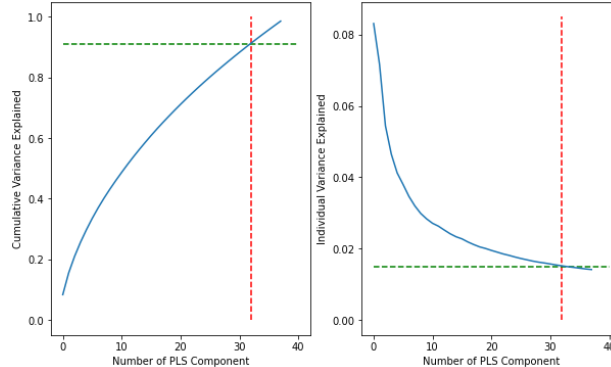
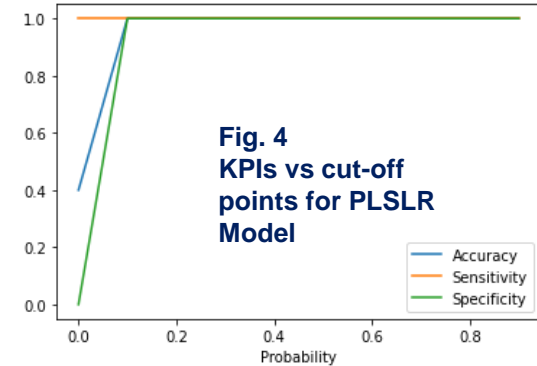


Fig. 4
KPIs vs cut-off
points for PLSLR
Model



Findings and Future Scopes

Criterial	PCLR	PLSLR
Total components for 91% explained variance	50	32
1 st component explains	13.58%	8.31%
2 nd component explains	7.96%	7.17%
3 rd component explains	6.15%	5.45%
Cut-off bandwidth for 100% accuracy	0.1-0.5	0.1-0.9



Based on the results from the analysis, we can conclude that PLSLR performs better over PCLR.

01



Overfit Model

To further analyse the model for overfitting

02



PLS-DA

Partial least squares discriminant analysis

FUTURE SCOPES:

THANK YOU!