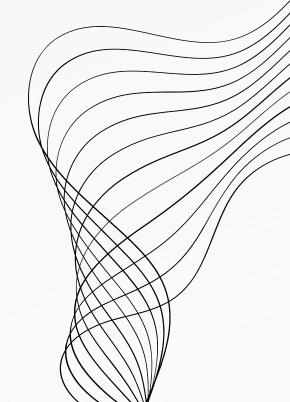# KEY IDENTIFIERS FOR A MALIGNANT BREAST MASS

LILY VOGEL

METHODS

Used Decision Trees to find the important variables for predicting a malignant breast mass

Then used clustering to verify that these features were significant

KEY IDENTIFIERS

Radius of grade 3 cells

Number of concave points

on contour of grade 3 cells

Decision Tree with Optimal Alpha

PATHS

From Decision Tree with Optimal Alpha

Outcome 1

small # concave points

small radius

benign
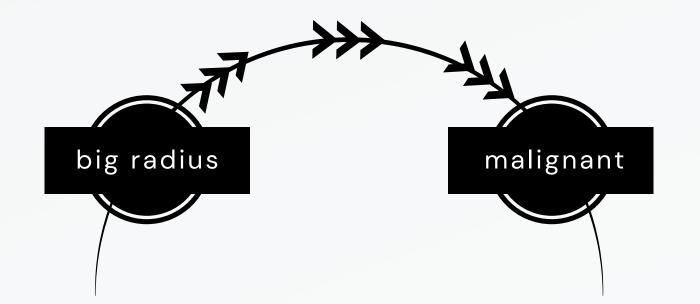
If the cell nucleus has a small radius and a small number of concave points, the mass is likely benign
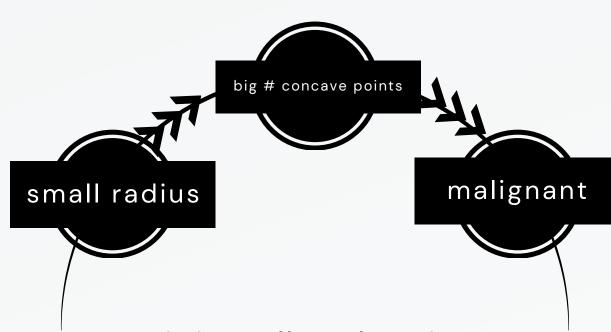
Outcome 2

If the cell nucleus has a big radius, then the mass is likely malignant

big radius

malignant

Outcome 3

big # concave points

small radius

malignant

If the cell nucleus has a small radius and a big number of concave points, the mass is likely malignant
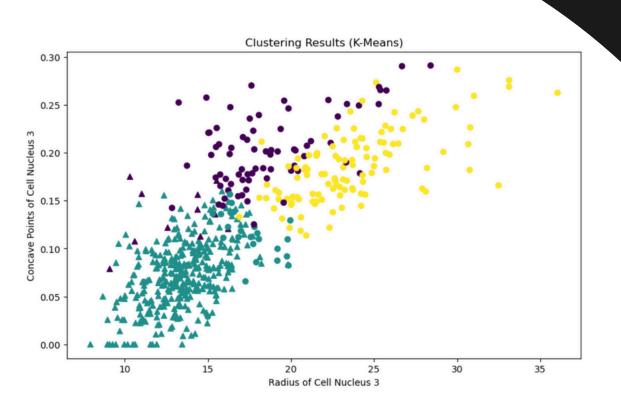
SUPPORTING EVIDENCE

- These paths came from the interpretations of my decision tree with optimal cost complexity pruning hyperparameter (alpha – α)

- I used cross validation to determine the optimal alpha value

DECISION TREE

CLUSTERS

The different colors correspond to different clusters

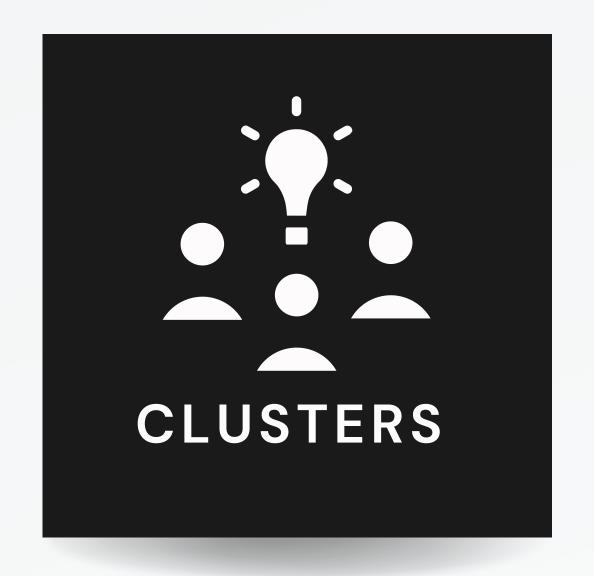The triangle shaped points are benign masses & the circle ones are malignant

Interpretation:

- distinct clusters indicates that these features are significant
- given the benign masses are those that are low in both concave points and radius, it supports our findings from earlier

Clustering Results (K-Means)

SUPPORTING EVIDENCE

- From clustering, there is evidence that supports our conclusions from the decision trees

- I used the elbow method to find the optimal number of clusters to use

CLUSTERS

CONCLUSION

Two important factors to consider when diagnosing a breast mass as malignant or benign is the radius of grade 3 cell nuclei and the number of concave points on grade 3 cell nuclei

NEXT STEPS

Given more time, I would be interested in looking into the following things to continue my exploration on breast mass diagnosis

Explore other features of grade 3 cell nuclei

NEAR FUTURE

Specifically, explore grade 1 texture of cell nucleus since it appeared as splitting criteria in one of the trees

FUTURE

Explore each grade of cell nuclei separately and see if radius and concave points are still the most important features

FAR FUTURE