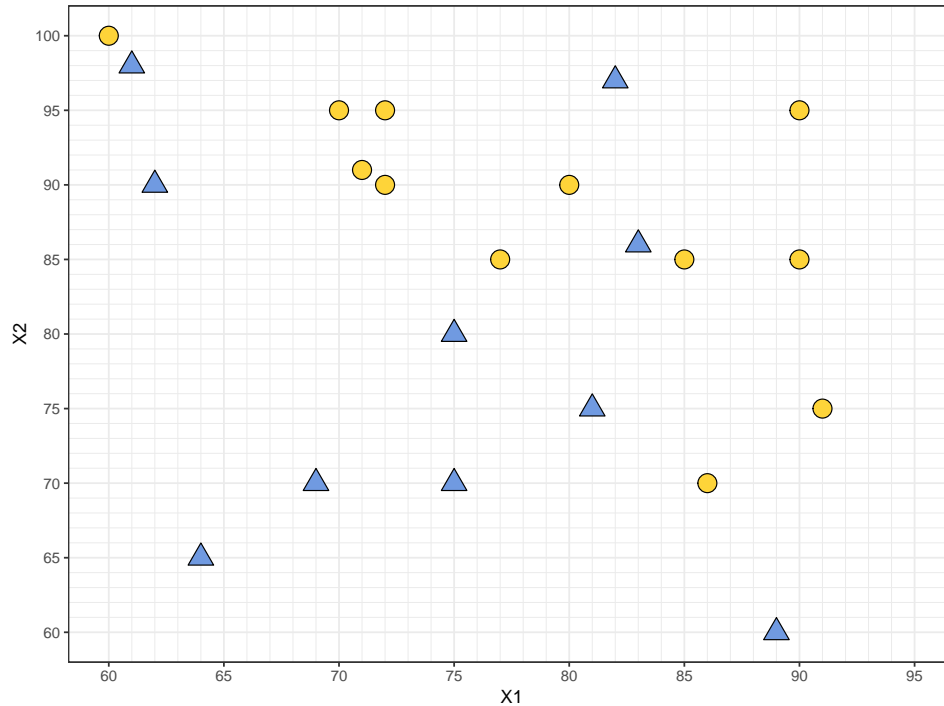


# Recursive Partitioning and Pruning: Part II

2019-04-08

Consider the following data that show an outcome (circles OR triangles) based on values for two predictors, X1 and X2.



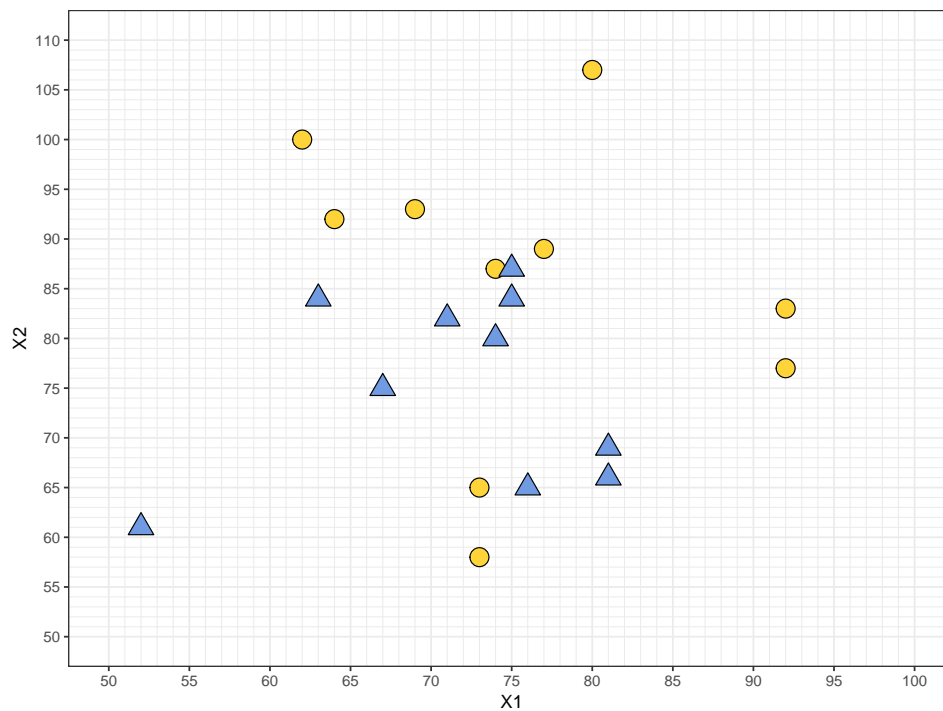
1. Draw a *single horizontal or vertical line* through the data that results in the highest possible classification accuracy for the data. Indicate the side of the line that is to be classified as “triangles” and that which will be classified as “circles”. Label this line as Decision Rule 1.
2. Consider the cases that were classified as circles by your first decision rule. Draw a *single horizontal or vertical line* through the data that results in the highest possible classification accuracy for those cases. Indicate the side of the line that is to be classified as “triangles” and that which will be classified as “circles”. Label this line as Decision Rule 2.1.
3. Consider the cases that were classified as triangles by your first decision rule. Draw a *single horizontal or vertical line* through the data that results in the highest possible classification accuracy for those cases. Indicate the side of the line that is to be classified as “triangles” and that which will be classified as “circles”. Label this line as Decision Rule 2.2.

4. For each of the Stage 2 rules, consider the cases that were classified as circles. Can you draw *a single horizontal or vertical line* through the data that improves the classification accuracy for those cases? What about for those cases that were classified as triangles? Continue the partitioning and labeling of stages until the classification accuracy can no longer be improved. (This may happen prior to reaching 100% classification accuracy.)
5. Draw out the decision tree based on your partitioning.

## Tree Trimming/Pruning

Below we provide you with 20 new validation cases to use in the model evaluation process.

X1	X2	Outcome	X1	X2	Outcome
73	58	Circle	92	83	Circle
52	61	Triangle	63	84	Triangle
73	65	Circle	75	84	Triangle
76	65	Triangle	74	87	Circle
81	66	Triangle	75	87	Triangle
81	69	Triangle	77	89	Circle
67	75	Triangle	64	92	Circle
92	77	Circle	69	93	Circle
74	80	Triangle	62	100	Circle
71	82	Triangle	80	107	Circle



6. Begin the process of evaluation by computing the classification accuracy for the model that classified 100% of the training cases correctly. Note: *Do not update or change any of the decision rules. Just use them to compute the classification accuracy.*

7. Prune the tree using a threshold for improvement of at least 3 percentage points. Continue pruning rules until you adopt a “final” tree. Draw out the “final” decision tree based on your pruning and report the classification accuracy for this model.