Problem Set 1

Consider the lower case letters of the roman alphabet:

\[ a b c d e f g h i j k l m n o p q r s t u v w x y z \]

1. Two possible features are \( x_1 \) “contains curved segments” and \( x_2 \) “contains straight segments”. What is the minimum number of such binary features that would be needed to successfully classify all 26 letters?

2. Choose five additional features and indicate the subset of the letters having the features (if binary) or the values of the feature for each letter if the features are not binary.

3. What is the largest subset of the lower case letters that can be uniquely classified using your 7 features?

4. What is the largest set of indistinguishable letters?

5. Are any of the features redundant?

6. Add another feature that improves your ability to classify. Explain how you are defining improvement.

7. Try to come up with additional features so that you can uniquely identify all of the letters. (Try to use as few features as possible.)

8. Now consider the font below and discuss how well your classifier works on this font. (In particular, consider all of the issues discussed above.)

\[ a b c d e f g h i j k l m n o p q r s t u v w x y z \]