STAT 602 Final Project

cps

4/20/2020

Kinematic Features Final Project (STAT 602 2020)

Instructions

This project is intended to give you practice of doing "skunk work" i.e. finding out if something is possible and reasonable to pursue.

You will work by yourself to prepare a 15-minute presentation on the problem background; model building process and your recommendations. Provide a short white paper of your model building process, the selected model, and the model's properties. Include annotated code that replicates your analysis. **MAKE SURE YOU PROPERLY REFERENCE ANY MATERIAL YOU USE IN THIS EXAM-this is an easy way to fail the final and course!!**

Due to internet upload speeds and connectivity issues we will not have an in person presentation.

Due Date- We will use the in person final exam time as the due date for the **Wednesday 11:30 a.m. - 1:30 p.m. May 6, 2020 Central Standard Time**. It is not required but if you would like to present your work to me at this time I will be available in zoom.

> If you need additional time or any other concerns please let me know sooner than later so that we can make alternative arrangements.

Background

Recently I have been working with a team of researchers (Lead by Prof. Michael Caliguri) exploring the relationship between kinematics of handwritten text and the ability of forensic document examiners to determine if two documents were written by the same writer.

A side debate has started among the team concerning whether or not the the statisticians and machine learning experts could build a reasonably accurate classifier that predicts three things at once 1. The writer of a short note (40 writers) 2. Which of 6 phrase was written 3. What type/style of writing was used (Print or Script) Based on the kinematic features recorded by MovAlyzeR.

Data structure

The column Labels of the data are as follows (This is the basic output order from MovAlyzeR.) Please see the provide powerpoints and papers for details.

1. Group- The type of handwriting- either Cursive or print.
2. Subject is the writer of the handwriting sample.
3. Condition- Is the Phrase that is being written.
4. Trial is one handwriting sample
5. Segment is one part of the writing sample a 'stroke'.
6. Direction- if the stroke is up or down.

7. StartTime- When the segment was started to be written.
8. Duration- How long it has taken to complete the segment.
9. VerticalSize
10. PeakVerticalVelocity
11. PeakVerticalAcceleration
12. HorizontalSize
13. StraightnessError
14. Slant
15. LoopSurface
16. RelativeInitialSlant
17. RelativeTimeToPeakVerticalVelocity
18. RelativePenDownDuration
19. RelativeDurationofPrimary
20. RelativeSizeofPrimary
21. AbsoluteSize
22. AverageAbsoluteVelocity
23. Roadlength
24. AbsoluteyJerk
25. NormalizedyJerk
26. AverageNormalizedyJerkPerTrial
27. AbsoluteJerk
28. NormalizedJerk
29. AverageNormalizedJerkPerTrial
30. NumberOfPeakAccelerationPoints
31. AveragePenPressure

Data Sets

You will be provided with

1. labeled.csv
2. example of unlabeled data to design your predictions on my unlabeled data set on my computer. (There are two documents to classify.)

```
labeled.dat=read.csv("/Users/saunders/labeled.csv", stringsAsFactors=F)[,-1]
str(labeled.dat)
```

```
## 'data.frame':    103311 obs. of  30 variables:
##  $ Group                              : chr  "CUR" "CUR" "CUR" "CUR" ...
##  $ Subject                            : chr  "0" "0" "0" "0" ...
##  $ Condition                          : chr  "L1" "L1" "L1" "L1" ...
##  $ Trial                              : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Segment                            : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Direction                          : int  1 -1 1 -1 1 -1 1 -1 1 -1 ...
##  $ Duration                           : num  0.0894 0.2604 0.2688 0.6222 0.2233 ...
##  $ VerticalSize                       : num  0.0826 -0.5291 0.4285 -0.395 0.2724 ...
##  $ PeakVerticalVelocity               : num  1.67 -3.33 2.52 -1.58 2.12 ...
##  $ PeakVerticalAcceleration           : num  53.7 -40.2 56.8 -23.3 28.8 ...
##  $ HorizontalSize                     : num  -0.0946 -0.0362 0.1273 0.1741 0.2388 ...
##  $ StraightnessError                  : num  0.0604 0.1225 0.1451 0.2099 0.0533 ...
##  $ Slant                              : num  2.346 -1.69 1.331 -1.504 0.886 ...
##  $ LoopSurface                        : num  0e+00 0e+00 0e+00 1e-03 0e+00 4e-04 0e+00 0e+00 0e+00 0e
##  $ RelativeInitialSlant               : num  -0.0996 -0.9133 -0.7585 1.3047 -0.3695 ...
##  $ RelativeTimeToPeakVerticalVelocity: num  0.494 0.496 0.235 0.378 0.517 ...
##  $ RelativePenDownDuration            : num  1 1 1 0.29 1 ...
##  $ RelativeDurationofPrimary          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RelativeSizeofPrimary              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ AbsoluteSize                       : num  0.126 0.53 0.447 0.432 0.362 ...
##  $ AverageAbsoluteVelocity            : num  1.54 2.95 2.48 1.26 1.74 ...
##  $ Roadlength                         : num  0.138 0.767 0.666 0.785 0.388 ...
##  $ AbsoluteyJerk                      : num  580 237 370 240 215 ...
##  $ NormalizedyJerk                    : num  7.81 10.93 21.94 119.81 9.88 ...
##  $ AverageNormalizedyJerkPerTrial     : num  174 174 174 174 174 ...
##  $ AbsoluteJerk                       : num  617 339 472 751 301 ...
##  $ NormalizedJerk                     : num  8.31 15.61 27.95 375.76 13.85 ...
##  $ AverageNormalizedJerkPerTrial      : num  293 293 293 293 293 ...
##  $ NumberOfPeakAccelerationPoints     : int  1 1 4 5 3 0 2 4 8 2 ...
##  $ AveragePenPressure                 : num  338.2 489.3 629 99.7 300 ...
```

```
unlab.example.trial=
  read.csv("/Users/saunders/unlab.example.trial.csv", stringsAsFactors=F)[,-1]
str(labeled.dat)
```

```
## 'data.frame':    103311 obs. of  30 variables:
##  $ Group                              : chr  "CUR" "CUR" "CUR" "CUR" ...
##  $ Subject                            : chr  "0" "0" "0" "0" ...
##  $ Condition                          : chr  "L1" "L1" "L1" "L1" ...
##  $ Trial                              : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Segment                            : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Direction                          : int  1 -1 1 -1 1 -1 1 -1 1 -1 ...
##  $ Duration                           : num  0.0894 0.2604 0.2688 0.6222 0.2233 ...
##  $ VerticalSize                       : num  0.0826 -0.5291 0.4285 -0.395 0.2724 ...
##  $ PeakVerticalVelocity               : num  1.67 -3.33 2.52 -1.58 2.12 ...
##  $ PeakVerticalAcceleration           : num  53.7 -40.2 56.8 -23.3 28.8 ...
##  $ HorizontalSize                     : num  -0.0946 -0.0362 0.1273 0.1741 0.2388 ...
##  $ StraightnessError                  : num  0.0604 0.1225 0.1451 0.2099 0.0533 ...
##  $ Slant                              : num  2.346 -1.69 1.331 -1.504 0.886 ...
##  $ LoopSurface                        : num  0e+00 0e+00 0e+00 1e-03 0e+00 4e-04 0e+00 0e+00 0e+00 0e
##  $ RelativeInitialSlant               : num  -0.0996 -0.9133 -0.7585 1.3047 -0.3695 ...
```

```
##  $ RelativeTimeToPeakVerticalVelocity: num  0.494 0.496 0.235 0.378 0.517 ...
##  $ RelativePenDownDuration           : num  1 1 1 0.29 1 ...
##  $ RelativeDurationofPrimary         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RelativeSizeofPrimary             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ AbsoluteSize                      : num  0.126 0.53 0.447 0.432 0.362 ...
##  $ AverageAbsoluteVelocity           : num  1.54 2.95 2.48 1.26 1.74 ...
##  $ Roadlength                        : num  0.138 0.767 0.666 0.785 0.388 ...
##  $ AbsoluteyJerk                     : num  580 237 370 240 215 ...
##  $ NormalizedyJerk                   : num  7.81 10.93 21.94 119.81 9.88 ...
##  $ AverageNormalizedyJerkPerTrial    : num  174 174 174 174 174 ...
##  $ AbsoluteJerk                      : num  617 339 472 751 301 ...
##  $ NormalizedJerk                    : num  8.31 15.61 27.95 375.76 13.85 ...
##  $ AverageNormalizedJerkPerTrial     : num  293 293 293 293 293 ...
##  $ NumberOfPeakAccelerationPoints    : int  1 1 4 5 3 0 2 4 8 2 ...
##  $ AveragePenPressure                : num  338.2 489.3 629 99.7 300 ...
```

```
xtabs(~Trial+Condition+Group+Subject, labeled.dat)
```

```
## , , Group = CUR, Subject = 0
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 67 75 75 75 75 75
##     2  8 75 13 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 0
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 68 75 62
##     2 75 75 75 72 75 62
##     3 75 75 75 68 70 60
##
## , , Group = CUR, Subject = 00A
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 73 72 73 65
##     2 75 75 75 71 75 65
##     3 75 75 75 70 75 72
##
## , , Group = PRI, Subject = 00A
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 70
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 72
##
## , , Group = CUR, Subject = 00B
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 62 71 61
```

```
##     2 75 74 75 64 75 57
##     3 75 75 75 62 73 57
##
## , , Group = PRI, Subject = 00B
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 74
##     2 74 75 75 68 75 70
##     3 63 75 75 64 75 66
##
## , , Group = CUR, Subject = 00C
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 72
##
## , , Group = PRI, Subject = 00C
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 72 75 75 65 63 59
##     2 74 75 75 65 67 60
##     3 72 75 75 65 67 63
##
## , , Group = CUR, Subject = 00D
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 00D
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 70
##     2 73 75 75 75 75 68
##     3 75 75 75 75 75 64
##
## , , Group = CUR, Subject = 00F
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 55 75 70 73 75 59
##     2 75 75 75 75 74 71
##     3 75 75 75 75 75 67
##
## , , Group = PRI, Subject = 00F
##
##      Condition
```

```
## Trial L1 L2 L3 L4 L5 L6
##     1 39 75 75 75 75 66
##     2 65 75 75 68 75 63
##     3 75 75 75 74 75 28
##
## , , Group = CUR, Subject = 00H
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75  9 75 75 75  8
##     2 75 75 38 75 75 61
##     3 72 43 75 65 72 67
##
## , , Group = PRI, Subject = 00H
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 61 70 54
##     2 72 75 75 62 68 50
##     3 72 75 75 70 75 56
##
## , , Group = CUR, Subject = 00I
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 73 74 64
##     2 75 75 75 71 75 63
##     3 75 75 75 74 75 71
##
## , , Group = PRI, Subject = 00I
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 66
##     2 75 75 75 75 75 70
##     3 75 75 75 75 75 67
##
## , , Group = CUR, Subject = 00J
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 00J
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 74 73
##     2 75 75 75 75 74 71
##     3 75 75 75 75 75 69
##
## , , Group = CUR, Subject = 00K
```

```
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 64 75 74 75 72
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 00K
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 72 70 60
##     2 75 75 75 70 70 61
##     3 75 75 75 72 72 64
##
## , , Group = CUR, Subject = 00L
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 69
##     2 75 75 75 75 75 69
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 00L
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 67 75 75 69 73 61
##     2 75 75 75 73 68 68
##     3 75 70 75 75 69 61
##
## , , Group = CUR, Subject = 00N
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 74
##     2 75 75 75 75 75 68
##     3 75 75 75 75 75 66
##
## , , Group = PRI, Subject = 00N
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 64
##     2 75 75 75 73 75 66
##     3 75 75 75 72 73 64
##
## , , Group = CUR, Subject = 00O
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 70 75 75 75 75 75
##     2 75 75 75 74 75 65
##     3 73 75 75 72 75 67
```

```
##
## , , Group = PRI, Subject = 00O
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 68
##
## , , Group = CUR, Subject = 00P
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 74 75 74
##     2 75 75 75 74 73 69
##     3 75 75 75 75 75 69
##
## , , Group = PRI, Subject = 00P
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 65
##     2 75 75 75 75 75 67
##     3 75 75 75 73 75 68
##
## , , Group = CUR, Subject = 00Q
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 71
##     2 75 75 75 75 75 73
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 00Q
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 72 75 62
##     2 75 75 75 75 75 65
##     3 74 75 75 75 75 63
##
## , , Group = CUR, Subject = 00R
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 71 75 65
##     2 74 75 75 75 75 75
##     3 75 75 75 75 74 71
##
## , , Group = PRI, Subject = 00R
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 66 74 61
```

```
##     2 75 75 75 68 71 60
##     3 75 75 75 74 73 65
##
## , , Group = CUR, Subject = 00T
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 68 72
##     2 75 75 75 75 75 71
##     3 75 75 75 75 75 71
##
## , , Group = PRI, Subject = 00T
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 58 75 75 15 75 70
##     2 75 75 75 75 75 65
##     3 75 75 75 68 75 67
##
## , , Group = CUR, Subject = 00U
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 70
##     2 75 75 75 75 75 70
##     3 75 75 75 75 75 66
##
## , , Group = PRI, Subject = 00U
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 51
##     2 75 75 75 32 75 15
##     3 75 75 75 75 75 17
##
## , , Group = CUR, Subject = 00V
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 29 75 75 75
##     3 75 75 23 75 75 75
##
## , , Group = PRI, Subject = 00V
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 69 67 60
##     2 71 75 75 74 69 60
##     3 75 75 75 75 63 68
##
## , , Group = CUR, Subject = 00W
##
##      Condition
```

```
## Trial L1 L2 L3 L4 L5 L6
##     1 57 75 51 75 75 71
##     2 75 75 75 75 75 71
##     3 75 75 75 75 75 71
##
## , , Group = PRI, Subject = 00W
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 68
##     2 75 75 75 75 75 66
##     3 75 75 75 75 75 66
##
## , , Group = CUR, Subject = 00X
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 49 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 38 75 75 75 75 75
##
## , , Group = PRI, Subject = 00X
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 73 75 75 70 75 68
##     2 75 75 75 70 75 68
##     3 75 75 75 66 75 70
##
## , , Group = CUR, Subject = 00Z
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 73
##     2 75 75 75 75 75 70
##     3 75 75 75 75 75 72
##
## , , Group = PRI, Subject = 00Z
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 72
##     2 75 75 75 75 75 72
##     3 75 75 75 75 75 75
##
## , , Group = CUR, Subject = 01A
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 16 75 65
##     2 75 75 24 51 75 75
##     3 75 75 32 75 75 74
##
## , , Group = PRI, Subject = 01A
```

```
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 65 75 75 75 69
##     2 75 75 75 75 75 74
##     3 75 75 75 75 75 72
##
## , , Group = CUR, Subject = 01B
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 74 66
##     2 75 75 75 70 75 60
##     3 75 75 75 74 71 64
##
## , , Group = PRI, Subject = 01B
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 72
##     2 75 75 75 75 75 74
##     3 75 75 75 75 75 66
##
## , , Group = CUR, Subject = 1
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 1
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 68
##     2 75 75 75 75 75 72
##     3 75 75 75 75 75 74
##
## , , Group = CUR, Subject = 10
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 72 75 75 69 69 60
##     2 73 75 75 72 68 58
##     3 74 75 75 71 70 60
##
## , , Group = PRI, Subject = 10
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 70 72 64
##     2 75 75 75 66 69 60
##     3 75 75 75 66 67 60
```

```
##
## , , Group = CUR, Subject = 11
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 60 75
##     2 74 75 75 75 75 73
##     3 75 75 75 75 75 71
##
## , , Group = PRI, Subject = 11
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 69 75 75 70 75 62
##     2 73 75 75 75 75 64
##     3 75 75 75 72 75 64
##
## , , Group = CUR, Subject = 12
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 67 75 66 73 58
##     2 75 75 75 66 75 55
##     3 74 75 75 74 74 51
##
## , , Group = PRI, Subject = 12
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1  9 75 75 64 75 68
##     2 73 75 75 68 69 72
##     3 69 75 75 66 75 59
##
## , , Group = CUR, Subject = 13
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 61 75 75 70 73 71
##     2 73 70 75 60 69 62
##     3 63 75 75 63 71 58
##
## , , Group = PRI, Subject = 13
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 74
##     2 75 75 75 75 75 68
##     3 75 75 75 75 75 67
##
## , , Group = CUR, Subject = 16
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 59 75 68 53
```

```
##     2 75 75 75 70 74 59
##     3 75 75 75 68 70 43
##
## , , Group = PRI, Subject = 16
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 64 65 56
##     2 74 75 75 62 66 56
##     3 75 75 75 58 63 56
##
## , , Group = CUR, Subject = 18
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 73 75 75 73 74 65
##     2 72 75 75 75 69 64
##     3 72 75 75 71 69 59
##
## , , Group = PRI, Subject = 18
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 64
##     2 75 75 75 75 75 70
##     3 75 75 75 75 75 63
##
## , , Group = CUR, Subject = 19
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 73 75 52 67 72 75
##     2 75 75 37 73 75 71
##     3 75 75 43 65 75 65
##
## , , Group = PRI, Subject = 19
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 72 75 69
##     2 75 75 75 74 75 75
##     3 75 75 75 75 75 69
##
## , , Group = CUR, Subject = 2
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 67
##     2 75 75 10 75 75 67
##     3 75 75 75 75 75 69
##
## , , Group = PRI, Subject = 2
##
##      Condition
```

```
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 74 75 75
##     3 75 75 75 75 75 70
##
## , , Group = CUR, Subject = 3
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 72
##
## , , Group = PRI, Subject = 3
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 64 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = CUR, Subject = 4
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 75
##     2 75 75 75 75 75 75
##     3 75 75 75 75 75 75
##
## , , Group = PRI, Subject = 4
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 73  9 66
##     2 75 75 75 75 67 64
##     3 75 75 75 75 75 64
##
## , , Group = CUR, Subject = 5
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 71 75 69
##     2 75 75 75 69 75 68
##     3 75 75 75 71 70 69
##
## , , Group = PRI, Subject = 5
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 70 68 64
##     2 75 75 75 71 64 59
##     3 71 75 75 69 62 66
##
## , , Group = CUR, Subject = 6
```

```
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 75 75 71
##     2 17 75 75 75 75 67
##     3 75 75 75 73 75 67
##
## , , Group = PRI, Subject = 6
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 70 75 57
##     2 71 75 75 73 75 63
##     3 75 75 75 70 75 68
##
## , , Group = CUR, Subject = 7
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 65 75 75
##     2 75 75 75 53 75 75
##     3 75 75 75 60 75 75
##
## , , Group = PRI, Subject = 7
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 63 75 75 53 68 54
##     2 75 75 68 51 73 53
##     3 67 75 70 64 68 52
##
## , , Group = CUR, Subject = 8
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 75 75 75 67 65 55
##     2 74 75 75 74 66 58
##     3 74 75 75 74 70 58
##
## , , Group = PRI, Subject = 8
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 73 75 75 68 67 62
##     2 75 75 75 68 72 60
##     3 75 75 75 68 72 60
##
## , , Group = CUR, Subject = 9
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 63 71 75 13 71 63
##     2 59 75 75 69 66 62
##     3 70 75 75 61 61 52
```

```
##
## , , Group = PRI, Subject = 9
##
##      Condition
## Trial L1 L2 L3 L4 L5 L6
##     1 71 75 75 75 75 70
##     2 73 75 75 75 75 65
##     3 75 75 75 67 75 68
```

Grading

Your grade (either pass or fail) and will be determined by three components:

1. The write-up of the methods you have used in building your classifiers.

   ```
   > i- You should use multiple classification methods and document how you constructed them.

   > ii- YOU NEED TO PROVIDE AN ESTIMATE OF THE ACCURACY OF THE CLASSIFIER IN YOUR WRITE UP!
   ```

2. You will need to provide an estimate of the accuracy of the classifier you build when applied to the third dataset of that I will use your code on. (Please make sure that you explain how to get your predictions.)
3. Your knowledge of the material in Stat 601, Stat 602, and the methods used in constructing your classifiers as demonstrated by the white paper component of the exam.

You will need to turn in four components (**Parts 1, 2, 3, and 4 are Due at Wednesday 11:30 a.m. - 1:30 p.m. May 6, 2020 Central Standard Time**):

- An annotated R-script that build your classifier off of labeled.csv.
- A write-up on the construction on your classifier. (Whitepaper approx. 2 pages written text max. Tables and plots do not count.)
- A powerpoint presentation discussing the work you have done on the project.
- A second annotated R-script that predicts the class for a set on unlabeled observations that I can run on my computer. I will provide an example of how the this data set will be formatted.

  Do ask me (or the GTAs) for help if you are having trouble. Beware this is a rather open ended question.

  Start early get something that works then come back and build up a better solution.

```
Good Luck

-cps
```