Tennis Group Project

Samuel Stockman, Angie Gray, Jack Simons

10/12/2020

Let's first load in the relevant data (see `tennis.R`)

```
file_list = list.files()
```

Now we want to compile the data into a better format. We need the input `variables` which should be equal to something of the form: `variables = c('surface', 'winner_id', 'loser_id', 'loser_age', 'winner_age')` - alternatively if it left blank then we will load all the variables

```
load_data <- function(file_list, variables=FALSE) {
  no_files <- length(file_list)
  for (file_index in 1:no_files) {
    temp_file <- read.csv(file_list[[file_index]])
    if (class(variables)=="character") {
      temp_file <- subset(temp_file, select = variables)
    }
    if (file_index==1) {
      melted_files <- temp_file
    } else {
      melted_files <- rbind(melted_files, temp_file)
    }
  }
  return(melted_files)
}
```

Let's now use this function

```
# variables <- c('surface', 'winner_id', 'loser_id', 'winner_age', 'loser_age')
my_data <- load_data(file_list)
```

Let's see what the data looks like

```
head(my_data, 30)
```

```
##    tourney_id              tourney_name surface draw_size tourney_level
## 1   2015-D001  Fed Cup WG F: CZE vs RUS    Hard         4             D
## 2   2015-D001  Fed Cup WG F: CZE vs RUS    Hard         4             D
## 3   2015-D001  Fed Cup WG F: CZE vs RUS    Hard         4             D
## 4   2015-D001  Fed Cup WG F: CZE vs RUS    Hard         4             D
## 5   2015-D002 Fed Cup WG R1: CAN vs CZE    Hard         4             D
## 6   2015-D002 Fed Cup WG R1: CAN vs CZE    Hard         4             D
## 7   2015-D002 Fed Cup WG R1: CAN vs CZE    Hard         4             D
## 8   2015-D003 Fed Cup WG R1: ITA vs FRA    Clay         4             D
## 9   2015-D003 Fed Cup WG R1: ITA vs FRA    Clay         4             D
## 10  2015-D003 Fed Cup WG R1: ITA vs FRA    Clay         4             D
## 11  2015-D003 Fed Cup WG R1: ITA vs FRA    Clay         4             D
```

```
## 12  2015-D004 Fed Cup WG R1: POL vs RUS    Hard         4             D
## 13  2015-D004 Fed Cup WG R1: POL vs RUS    Hard         4             D
## 14  2015-D004 Fed Cup WG R1: POL vs RUS    Hard         4             D
## 15  2015-D005 Fed Cup WG R1: GER vs AUS    Hard         4             D
## 16  2015-D005 Fed Cup WG R1: GER vs AUS    Hard         4             D
## 17  2015-D005 Fed Cup WG R1: GER vs AUS    Hard         4             D
## 18  2015-D005 Fed Cup WG R1: GER vs AUS    Hard         4             D
## 19  2015-D006 Fed Cup WG SF: CZE vs FRA    Hard         4             D
## 20  2015-D006 Fed Cup WG SF: CZE vs FRA    Hard         4             D
## 21  2015-D006 Fed Cup WG SF: CZE vs FRA    Hard         4             D
## 22  2015-D007 Fed Cup WG SF: RUS vs GER    Clay         4             D
## 23  2015-D007 Fed Cup WG SF: RUS vs GER    Clay         4             D
## 24  2015-D007 Fed Cup WG SF: RUS vs GER    Clay         4             D
## 25  2015-D007 Fed Cup WG SF: RUS vs GER    Clay         4             D
## 26  2015-D008 Fed Cup WG R1: NED vs SVK    Clay         4             D
## 27  2015-D008 Fed Cup WG R1: NED vs SVK    Clay         4             D
## 28  2015-D008 Fed Cup WG R1: NED vs SVK    Clay         4             D
## 29  2015-D008 Fed Cup WG R1: NED vs SVK    Clay         4             D
## 30  2015-D009 Fed Cup WG R1: ROU vs ESP    Hard         4             D
##    tourney_date match_num winner_id winner_seed winner_entry
## 1      20151114         1    201520        <NA>
## 2      20151114         2    201345        <NA>
## 3      20151114         3    201345        <NA>
## 4      20151114         4    201662        <NA>
## 5      20150207         1    201662        <NA>
## 6      20150207         2    202702        <NA>
## 7      20150207         3    201662        <NA>
## 8      20150207         1    201506        <NA>
## 9      20150207         2    202429        <NA>
## 10     20150207         3    201540        <NA>
## 11     20150207         4    201614        <NA>
## 12     20150207         1    201320        <NA>
## 13     20150207         2    201345        <NA>
## 14     20150207         3    201345        <NA>
## 15     20150207         1    201457        <NA>
## 16     20150207         2    201492        <NA>
## 17     20150207         3    201493        <NA>
## 18     20150207         4    201492        <NA>
## 19     20150418         1    201425        <NA>
## 20     20150418         2    201520        <NA>
## 21     20150418         3    201520        <NA>
## 22     20150418         1    201320        <NA>
## 23     20150418         2    201499        <NA>
## 24     20150418         3    201492        <NA>
## 25     20150418         4    201493        <NA>
## 26     20150207         1    203533        <NA>
## 27     20150207         2    201551        <NA>
## 28     20150207         3    202428        <NA>
## 29     20150207         4    201551        <NA>
## 30     20150207         1    201594        <NA>
##                  winner_name winner_hand winner_ht winner_ioc winner_age
## 1              Petra Kvitova           L       183        CZE   25.68652
## 2            Maria Sharapova           R        NA        RUS   28.57221
## 3            Maria Sharapova           R        NA        RUS   28.57221
```

```
## 4          Karolina Pliskova           R       184        CZE   23.64956
## 5          Karolina Pliskova           R       184        CZE   22.88296
## 6            Tereza Smitkova           R        NA        CZE   20.32854
## 7          Karolina Pliskova           R       184        CZE   22.88296
## 8                Sara Errani           R       164        ITA   27.77823
## 9              Camila Giorgi           R       168        ITA   23.10746
## 10       Kristina Mladenovic           R       184        FRA   21.73580
## 11           Caroline Garcia           R        NA        FRA   21.31143
## 12       Svetlana Kuznetsova           R       174        RUS   29.61533
## 13           Maria Sharapova           R        NA        RUS   27.80561
## 14           Maria Sharapova           R        NA        RUS   27.80561
## 15         Jarmila Gajdosova           R       174        AUS   27.78645
## 16           Andrea Petkovic           R       180        GER   27.41410
## 17          Angelique Kerber           L       173        GER   27.05544
## 18           Andrea Petkovic           R       180        GER   27.41410
## 19            Lucie Safarova           L       177        CZE   28.19986
## 20             Petra Kvitova           L       183        CZE   25.11157
## 21             Petra Kvitova           L       183        CZE   25.11157
## 22       Svetlana Kuznetsova           R       174        RUS   29.80698
## 23  Anastasia Pavlyuchenkova           R       177        RUS   23.79192
## 24           Andrea Petkovic           R       180        GER   27.60575
## 25          Angelique Kerber           L       173        GER   27.24709
## 26 Anna Karolina Schmiedlova           R        NA        SVK   20.40246
## 27               Arantxa Rus           L       180        NED   24.15332
## 28              Kiki Bertens           R       182        NED   23.16222
## 29               Arantxa Rus           L       180        NED   24.15332
## 30              Simona Halep           R       168        ROU   23.36482
##    loser_id loser_seed loser_entry               loser_name loser_hand
## 1    201499       <NA>              Anastasia Pavlyuchenkova          R
## 2    201662       <NA>                     Karolina Pliskova          R
## 3    201520       <NA>                         Petra Kvitova          L
## 4    201499       <NA>              Anastasia Pavlyuchenkova          R
## 5    211796       <NA>                      Francoise Abanda          R
## 6    202681       <NA>                    Gabriela Dabrowski          R
## 7    202681       <NA>                    Gabriela Dabrowski          R
## 8    201614       <NA>                       Caroline Garcia          R
## 9    201427       <NA>                          Alize Cornet          R
## 10   201506       <NA>                           Sara Errani          R
## 11   202429       <NA>                         Camila Giorgi          R
## 12   201474       <NA>                   Agnieszka Radwanska          R
## 13   201524       <NA>                     Urszula Radwanska          R
## 14   201474       <NA>                   Agnieszka Radwanska          R
## 15   201493       <NA>                      Angelique Kerber          L
## 16   201325       <NA>                        Samantha Stosur          R
## 17   201325       <NA>                        Samantha Stosur          R
## 18   201457       <NA>                      Jarmila Gajdosova          R
## 19   201614       <NA>                       Caroline Garcia          R
## 20   201540       <NA>                    Kristina Mladenovic          R
## 21   201614       <NA>                       Caroline Garcia          R
## 22   201504       <NA>                         Julia Goerges          R
## 23   201513       <NA>                        Sabine Lisicki          R
## 24   201320       <NA>                    Svetlana Kuznetsova          R
## 25   201499       <NA>              Anastasia Pavlyuchenkova          R
## 26   202428       <NA>                         Kiki Bertens          R
```

```
## 27   201517       <NA>                 Magdalena Rybarikova          R
## 28   201517       <NA>                 Magdalena Rybarikova          R
## 29   203533       <NA>             Anna Karolina Schmiedlova         R
## 30   201562       <NA>                 Silvia Soler Espinosa         R
##    loser_ht loser_ioc loser_age          score best_of round minutes w_ace w_df
## 1       177       RUS  24.36687    2-6 6-1 6-1       3    RR      NA    NA   NA
## 2       184       CZE  23.64956        6-3 6-4       3    RR      NA    NA   NA
## 3       183       CZE  25.68652    3-6 6-4 6-2       3    RR      NA    NA   NA
## 4       177       RUS  24.36687        6-3 6-4       3    RR      NA    NA   NA
## 5        NA       CAN  18.00411        6-2 6-4       3    RR      NA    NA   NA
## 6        NA       CAN  22.85284        6-1 6-2       3    RR      NA    NA   NA
## 7        NA       CAN  22.85284        6-4 6-2       3    RR      NA    NA   NA
## 8        NA       FRA  21.31143     7-6(2) 7-5       3    RR      NA    NA   NA
## 9       173       FRA  25.04312        6-4 6-2       3    RR      NA    NA   NA
## 10      164       ITA  27.77823        6-4 6-3       3    RR      NA    NA   NA
## 11      168       ITA  23.10746    4-6 6-0 6-2       3    RR      NA    NA   NA
## 12      170       POL  25.92471    6-4 2-6 6-2       3    RR      NA    NA   NA
## 13      177       POL  24.16975        6-0 6-3       3    RR      NA    NA   NA
## 14      170       POL  25.92471        6-1 7-5       3    RR      NA    NA   NA
## 15      173       GER  27.05544    4-6 6-2 6-4       3    RR      NA    NA   NA
## 16      172       AUS  30.85832  6-4 3-6 12-10       3    RR      NA    NA   NA
## 17      172       AUS  30.85832        6-2 6-4       3    RR      NA    NA   NA
## 18      174       AUS  27.78645    6-3 3-6 8-6       3    RR      NA    NA   NA
## 19       NA       FRA  21.50308 4-6 7-6(1) 6-1       3    RR      NA    NA   NA
## 20      184       FRA  21.92745        6-3 6-4       3    RR      NA    NA   NA
## 21       NA       FRA  21.50308        6-4 6-4       3    RR      NA    NA   NA
## 22      180       GER  26.45585        6-4 6-4       3    RR      NA    NA   NA
## 23      178       GER  25.56879 4-6 7-6(4) 6-3       3    RR      NA    NA   NA
## 24      174       RUS  29.80698        6-2 6-1       3    RR      NA    NA   NA
## 25      177       RUS  23.79192        6-1 6-0       3    RR      NA    NA   NA
## 26      182       NED  23.16222        6-2 7-5       3    RR      NA    NA   NA
## 27      180       SVK  26.34360        6-3 6-4       3    RR      NA    NA   NA
## 28      180       SVK  26.34360    6-1 2-6 6-1       3    RR      NA    NA   NA
## 29       NA       SVK  20.40246    6-3 2-6 6-4       3    RR      NA    NA   NA
## 30       NA       ESP  27.21971        6-2 6-1       3    RR      NA    NA   NA
##    w_svpt w_1stIn w_1stWon w_2ndWon w_SvGms w_bpSaved w_bpFaced l_ace l_df
## 1      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 2      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 3      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 4      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 5      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 6      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 7      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 8      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 9      NA      NA       NA       NA      NA        NA        NA    NA   NA
## 10     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 11     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 12     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 13     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 14     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 15     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 16     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 17     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 18     NA      NA       NA       NA      NA        NA        NA    NA   NA
```

```
## 19     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 20     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 21     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 22     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 23     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 24     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 25     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 26     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 27     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 28     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 29     NA      NA       NA       NA      NA        NA        NA    NA   NA
## 30     NA      NA       NA       NA      NA        NA        NA    NA   NA
##    l_svpt l_1stIn l_1stWon l_2ndWon l_SvGms l_bpSaved l_bpFaced winner_rank
## 1      NA      NA       NA       NA      NA        NA        NA           6
## 2      NA      NA       NA       NA      NA        NA        NA           4
## 3      NA      NA       NA       NA      NA        NA        NA           4
## 4      NA      NA       NA       NA      NA        NA        NA          11
## 5      NA      NA       NA       NA      NA        NA        NA          22
## 6      NA      NA       NA       NA      NA        NA        NA          62
## 7      NA      NA       NA       NA      NA        NA        NA          22
## 8      NA      NA       NA       NA      NA        NA        NA          13
## 9      NA      NA       NA       NA      NA        NA        NA          31
## 10     NA      NA       NA       NA      NA        NA        NA          74
## 11     NA      NA       NA       NA      NA        NA        NA          30
## 12     NA      NA       NA       NA      NA        NA        NA          27
## 13     NA      NA       NA       NA      NA        NA        NA           2
## 14     NA      NA       NA       NA      NA        NA        NA           2
## 15     NA      NA       NA       NA      NA        NA        NA          54
## 16     NA      NA       NA       NA      NA        NA        NA          12
## 17     NA      NA       NA       NA      NA        NA        NA          10
## 18     NA      NA       NA       NA      NA        NA        NA          12
## 19     NA      NA       NA       NA      NA        NA        NA          13
## 20     NA      NA       NA       NA      NA        NA        NA           4
## 21     NA      NA       NA       NA      NA        NA        NA           4
## 22     NA      NA       NA       NA      NA        NA        NA          24
## 23     NA      NA       NA       NA      NA        NA        NA          38
## 24     NA      NA       NA       NA      NA        NA        NA          11
## 25     NA      NA       NA       NA      NA        NA        NA          14
## 26     NA      NA       NA       NA      NA        NA        NA          75
## 27     NA      NA       NA       NA      NA        NA        NA         218
## 28     NA      NA       NA       NA      NA        NA        NA          70
## 29     NA      NA       NA       NA      NA        NA        NA         218
## 30     NA      NA       NA       NA      NA        NA        NA           3
##    winner_rank_points loser_rank loser_rank_points
## 1                4220         28              1840
## 2                5011         11              3285
## 3                5011          6              4220
## 4                3285         28              1840
## 5                2015        230               195
## 6                 863        185               285
## 7                2015        185               285
## 8                2551         30              1461
## 9                1455         19              2125
## 10                780         13              2551
```

```
## 11              1461          31              1455
## 12              1701           8              4270
## 13              8210         135               407
## 14              8210           8              4270
## 15               925          10              3130
## 16              2735          25              1835
## 17              3130          25              1835
## 18              2735          54               925
## 19              2870          29              1560
## 20              6060          58               875
## 21              6060          29              1560
## 22              1870          63               840
## 23              1256          19              2127
## 24              3260          24              1870
## 25              2865          38              1256
## 26               727          70               811
## 27               207          47              1046
## 28               811          47              1046
## 29               207          75               727
## 30              6571          67               848
```

Now, for the logistic regression. We need to create our $D = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$

We need an activation function

```
activ_func <- function(t) {
  return(1/(1+exp(t)))
}
```