Predicting Run Direction with K-NN Algorithm

Sean Bradford

```r
rm(list=ls())

library(FNN)
```

```
## Warning: package 'FNN' was built under R version 4.2.1
```

```r
library(ISLR)
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.2.1
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data=read.csv("C:\\Users\\Sean\\OneDrive\\Desktop\\Grad School\\Machine Learning\\Final\\train.csv")

# create data frame for Colts possessing ball
colts.data=data%>%filter(data$PossessionTeam=="IND")

# create data frame for Colts running backs
colts.run=colts.data%>%filter(colts.data$Position=="RB")

head(colts.run)
```

```
##        GameId       PlayId Team      X      Y    S    A  Dis Orientation    Dir
## 1 2017091009 2.017091e+13 away  29.91  28.48 4.22 3.20 0.50       39.34 104.05
## 2 2017091009 2.017091e+13 away  31.70  28.83 3.09 2.32 0.42       14.97 104.35
## 3 2017091009 2.017091e+13 away  35.04  25.60 3.60 3.13 0.44       32.38  93.66
## 4 2017091009 2.017091e+13 away  80.45  20.56 4.68 1.51 0.59       35.14 128.92
## 5 2017091009 2.017091e+13 away 104.69  22.64 3.39 2.02 0.42       20.92 104.80
## 6 2017091009 2.017091e+13 away 104.82  24.13 3.64 2.46 0.41      350.65  49.76
##     NflId DisplayName JerseyNumber Season YardLine Quarter GameClock
## 1 2506404  Frank Gore           23   2017       25       1  12:43:00
## 2 2506404  Frank Gore           23   2017       27       1  12:05:00
## 3 2506404  Frank Gore           23   2017       30       1  11:04:00
## 4 2506404  Frank Gore           23   2017       25       1  09:14:00
## 5 2558123 Marlon Mack           25   2017        1       1  07:46:00
## 6 2558123 Marlon Mack           25   2017        2       1  07:06:00
##   PossessionTeam Down Distance FieldPosition HomeScoreBeforePlay
## 1            IND    1       10           IND                   3
## 2            IND    2        8           IND                   3
## 3            IND    1       10           IND                  10
## 4            IND    1       10            LA                  10
## 5            IND    1        1            LA                  10
## 6            IND    2        2            LA                  10
##   VisitorScoreBeforePlay NflIdRusher OffenseFormation OffensePersonnel
## 1                      0     2506404       SINGLEBACK  1 RB, 2 TE, 2 WR
## 2                      0     2506404           I_FORM  1 RB, 2 TE, 2 WR
## 3                      0     2506404       SINGLEBACK  1 RB, 1 TE, 3 WR
## 4                      0     2506404           I_FORM  1 RB, 2 TE, 2 WR
## 5                      0     2558123       SINGLEBACK  1 RB, 2 TE, 2 WR
## 6                      0     2558123          SHOTGUN  1 RB, 1 TE, 3 WR
##   DefendersInTheBox DefensePersonnel PlayDirection              TimeHandoff
## 1                 7   4 DL, 3 LB, 4 DB         right 2017-09-10T20:13:21.000Z
## 2                 7   4 DL, 3 LB, 4 DB         right 2017-09-10T20:14:00.000Z
## 3                 7   4 DL, 3 LB, 4 DB         right 2017-09-10T20:20:42.000Z
## 4                 8   4 DL, 3 LB, 4 DB         right 2017-09-10T20:22:41.000Z
## 5                 8   4 DL, 3 LB, 4 DB         right 2017-09-10T20:24:32.000Z
## 6                 6   4 DL, 3 LB, 4 DB         right 2017-09-10T20:25:12.000Z
##                  TimeSnap Yards PlayerHeight PlayerWeight PlayerBirthDate
## 1 2017-09-10T20:13:20.000Z     2          5-9          212      05/14/1983
## 2 2017-09-10T20:13:58.000Z    -2          5-9          212      05/14/1983
## 3 2017-09-10T20:20:40.000Z     4          5-9          212      05/14/1983
## 4 2017-09-10T20:22:39.000Z     8          5-9          212      05/14/1983
## 5 2017-09-10T20:24:31.000Z    -1          6-0          210      03/07/1996
## 6 2017-09-10T20:25:11.000Z     0          6-0          210      03/07/1996
##   PlayerCollegeName Position HomeTeamAbbr VisitorTeamAbbr Week
## 1             Miami       RB           LA             IND    1
## 2             Miami       RB           LA             IND    1
## 3             Miami       RB           LA             IND    1
## 4             Miami       RB           LA             IND    1
## 5     South Florida       RB           LA             IND    1
## 6     South Florida       RB           LA             IND    1
##                          Stadium           Location StadiumType  Turf
## 1 Los Angeles Memorial Coliseum Los Angeles, Calif.     Outdoor Grass
## 2 Los Angeles Memorial Coliseum Los Angeles, Calif.     Outdoor Grass
## 3 Los Angeles Memorial Coliseum Los Angeles, Calif.     Outdoor Grass
## 4 Los Angeles Memorial Coliseum Los Angeles, Calif.     Outdoor Grass
```

```
## 5 Los Angeles Memorial Coliseum Los Angeles, Calif.    Outdoor Grass
## 6 Los Angeles Memorial Coliseum Los Angeles, Calif.    Outdoor Grass
##   GameWeather Temperature Humidity WindSpeed WindDirection
## 1      Sunny          90       38         2            SW
## 2      Sunny          90       38         2            SW
## 3      Sunny          90       38         2            SW
## 4      Sunny          90       38         2            SW
## 5      Sunny          90       38         2            SW
## 6      Sunny          90       38         2            SW
```

summary(colts.run)

```
##      GameId              PlayId             Team                  X
##  Min.   :2.017e+09   Min.   :2.017e+13   Length:1055        Min.   :  8.77
##  1st Qu.:2.017e+09   1st Qu.:2.017e+13   Class :character   1st Qu.: 40.40
##  Median :2.018e+09   Median :2.018e+13   Mode  :character   Median : 61.81
##  Mean   :2.018e+09   Mean   :2.018e+13                      Mean   : 61.00
##  3rd Qu.:2.019e+09   3rd Qu.:2.019e+13                      3rd Qu.: 82.98
##  Max.   :2.019e+09   Max.   :2.019e+13                      Max.   :113.66
##        Y               S               A               Dis
##  Min.   :16.82   Min.   :0.230   Min.   :0.160   Min.   :0.0100
##  1st Qu.:23.75   1st Qu.:3.370   1st Qu.:2.040   1st Qu.:0.3500
##  Median :26.90   Median :3.970   Median :2.640   Median :0.4200
##  Mean   :26.80   Mean   :4.008   Mean   :2.653   Mean   :0.4107
##  3rd Qu.:29.71   3rd Qu.:4.610   3rd Qu.:3.150   3rd Qu.:0.4700
##  Max.   :36.44   Max.   :8.550   Max.   :8.850   Max.   :0.8600
##   Orientation         Dir             NflId          DisplayName
##  Min.   :  0.30   Min.   :  0.35   Min.   :2506404   Length:1055
##  1st Qu.: 87.84   1st Qu.: 84.10   1st Qu.:2533460   Class :character
##  Median :187.78   Median :174.31   Median :2558123   Mode  :character
##  Mean   :180.93   Mean   :177.34   Mean   :2545458
##  3rd Qu.:273.54   3rd Qu.:269.69   3rd Qu.:2558123
##  Max.   :358.67   Max.   :359.92   Max.   :2560933
##   JerseyNumber       Season        YardLine         Quarter
##  Min.   :20.00   Min.   :2017   Min.   : 1.00   Min.   :1.000
##  1st Qu.:23.00   1st Qu.:2017   1st Qu.:20.00   1st Qu.:2.000
##  Median :25.00   Median :2018   Median :29.00   Median :3.000
##  Mean   :24.17   Mean   :2018   Mean   :28.85   Mean   :2.534
##  3rd Qu.:25.00   3rd Qu.:2019   3rd Qu.:39.00   3rd Qu.:4.000
##  Max.   :39.00   Max.   :2019   Max.   :50.00   Max.   :5.000
##   GameClock         PossessionTeam         Down          Distance
##  Length:1055        Length:1055        Min.   :1.000   Min.   : 1.00
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.: 5.00
##  Mode  :character   Mode  :character   Median :1.000   Median :10.00
##                                        Mean   :1.575   Mean   : 7.95
##                                        3rd Qu.:2.000   3rd Qu.:10.00
##                                        Max.   :4.000   Max.   :20.00
##  FieldPosition      HomeScoreBeforePlay VisitorScoreBeforePlay
##  Length:1055        Min.   : 0.00       Min.   : 0.000
##  Class :character   1st Qu.: 3.00       1st Qu.: 0.000
##  Mode  :character   Median :10.00       Median : 7.000
##                     Mean   :11.18       Mean   : 8.387
##                     3rd Qu.:17.00       3rd Qu.:13.000
##                     Max.   :46.00       Max.   :42.000
```

```
##    NflIdRusher     OffenseFormation  OffensePersonnel  DefendersInTheBox
## Min.   :2506404    Length:1055       Length:1055       Min.   : 4.000
## 1st Qu.:2533460    Class :character  Class :character  1st Qu.: 6.000
## Median :2558123    Mode  :character  Mode  :character  Median : 7.000
## Mean   :2545422                                        Mean   : 6.825
## 3rd Qu.:2558123                                        3rd Qu.: 7.000
## Max.   :2562495                                        Max.   :11.000
## DefensePersonnel   PlayDirection     TimeHandoff         TimeSnap
## Length:1055        Length:1055       Length:1055       Length:1055
## Class :character   Class :character  Class :character  Class :character
## Mode  :character   Mode  :character  Mode  :character  Mode  :character
##
##
##
##      Yards          PlayerHeight       PlayerWeight  PlayerBirthDate
## Min.   :-10.000   Length:1055        Min.   :198    Length:1055
## 1st Qu.:  1.000   Class :character   1st Qu.:210    Class :character
## Median :  3.000   Mode  :character   Median :210    Mode  :character
## Mean   :  4.299                      Mean   :210
## 3rd Qu.:  6.000                      3rd Qu.:212
## Max.   : 63.000                      Max.   :239
## PlayerCollegeName    Position          HomeTeamAbbr      VisitorTeamAbbr
## Length:1055        Length:1055        Length:1055        Length:1055
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       Week           Stadium           Location          StadiumType
## Min.   : 1.000   Length:1055        Length:1055        Length:1055
## 1st Qu.: 4.000   Class :character   Class :character   Class :character
## Median : 8.000   Mode  :character   Mode  :character   Mode  :character
## Mean   : 8.468
## 3rd Qu.:12.000
## Max.   :17.000
##     Turf           GameWeather         Temperature        Humidity
## Length:1055        Length:1055        Min.   : 9.00    Min.   : 0.00
## Class :character   Class :character   1st Qu.:44.00    1st Qu.:54.00
## Mode  :character   Mode  :character   Median :59.00    Median :66.00
##                                       Mean   :58.66    Mean   :64.52
##                                       3rd Qu.:72.00    3rd Qu.:75.00
##                                       Max.   :93.00    Max.   :96.00
##   WindSpeed         WindDirection
## Length:1055        Length:1055
## Class :character   Class :character
## Mode  :character   Mode  :character
##
##
##
```

```
# isolate predictor variables
colts.run2=colts.run[,c("X","Y","Orientation","PlayDirection")]
```

```
# normalize data using z-score
norm1=preProcess(colts.run2,method=c('center','scale'))
colts.norm=predict(norm1,colts.run2)
```

```
# partition the data 80/20
set.seed(111)
train.index=createDataPartition(colts.norm$PlayDirection,p=.8,list=FALSE)

train=colts.norm[train.index,]
test=colts.norm[-train.index,]

train.predictors=train[,-4]
test.predictors=test[,-4]

train.labels=train[,4]
test.labels=test[,4]
```

```
# test model using k=4
pred.test.label=knn(train.predictors,test.predictors,cl=train.labels,k=4)
head(pred.test.label)
```

```
## [1] left  right left  right right right
## Levels: left right
```

```
# determine optimal k value
search.grid=expand.grid(k=c(1:20))
model=train(PlayDirection~X+Y+Orientation,data=colts.norm,method="knn",tuneGrid=search.grid)

model
```

```
## k-Nearest Neighbors
##
## 1055 samples
##    3 predictor
##    2 classes: 'left', 'right'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 1055, 1055, 1055, 1055, 1055, 1055, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.8486459  0.6970366
##    2  0.8467785  0.6932780
##    3  0.8492533  0.6983345
##    4  0.8534288  0.7065403
##    5  0.8629152  0.7255847
##    6  0.8681220  0.7360443
##    7  0.8734799  0.7467450
##    8  0.8751057  0.7499621
##    9  0.8776293  0.7550176
##   10  0.8775656  0.7548936
```

```
##    11  0.8794870   0.7588092
##    12  0.8799889   0.7597913
##    13  0.8802718   0.7603687
##    14  0.8785436   0.7569387
##    15  0.8781946   0.7562461
##    16  0.8778164   0.7554482
##    17  0.8791457   0.7581484
##    18  0.8793554   0.7585924
##    19  0.8799342   0.7597784
##    20  0.8768020   0.7535522
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 13.
```

```
# evaluate confusion matrix & determine model's fitness

library('gmodels')
```

```
## Warning: package 'gmodels' was built under R version 4.2.1
```

```
pred.test.label1=knn(train.predictors,test.predictors,cl=train.labels,k=13)
xyz=CrossTable(x=test.labels,y=pred.test.label1,prop.chisq=FALSE)
```

```
##
##
##     Cell Contents
## |-------------------------|
## |                       N |
## |            N / Row Total |
## |            N / Col Total |
## |          N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:   211
##
##
##                | pred.test.label1
##   test.labels |      left |     right | Row Total |
## --------------|-----------|-----------|-----------|
##          left |        94 |        10 |       104 |
##               |     0.904 |     0.096 |     0.493 |
##               |     0.895 |     0.094 |           |
##               |     0.445 |     0.047 |           |
## --------------|-----------|-----------|-----------|
##         right |        11 |        96 |       107 |
##               |     0.103 |     0.897 |     0.507 |
##               |     0.105 |     0.906 |           |
##               |     0.052 |     0.455 |           |
## --------------|-----------|-----------|-----------|
##  Column Total |       105 |       106 |       211 |
##               |     0.498 |     0.502 |           |
## --------------|-----------|-----------|-----------|
```

```
##
##
```

```
xyz
```

```
## $t
##        y
## x       left right
##    left   94    10
##    right  11    96
##
## $prop.row
##        y
## x             left       right
##    left  0.90384615 0.09615385
##    right 0.10280374 0.89719626
##
## $prop.col
##        y
## x             left       right
##    left  0.89523810 0.09433962
##    right 0.10476190 0.90566038
##
## $prop.tbl
##        y
## x             left       right
##    left  0.44549763 0.04739336
##    right 0.05213270 0.45497630
```

```
# create variables for true positives, false positives, true negatives, false negatives, total rows
TP=96
FP=10
TN=94
FN=11
N=211

# calculate model accuracy
accuracy=(TP+TN)/N
accuracy
```

```
## [1] 0.9004739
```

```
# calculate model recall
recall=TP/(TP+FN)
recall
```

```
## [1] 0.8971963
```

```
# calculate model precision
precision=TP/(TP+FP)
precision
```

```
## [1] 0.9056604
```

```
# calculate model specificity
specificity=TN/(TN+FP)
specificity
```

```
## [1] 0.9038462
```

```
# predict run direction during game
game.scenario=data.frame(X=105,Y=25,Orientation=21)

game.prediction=knn(train.predictors,game.scenario,cl=train.labels,k=13)
head(game.prediction)
```

```
## [1] left
## Levels: left
```