**Logistic Regression:**

#reading data

eye.dia<-read.csv(file.choose())

#splitting data

smp\_size\_log <- floor(0.60 \* nrow(eye.dia))

## set the seed to make your partition reproductible

set.seed(123)

train\_ind\_log <- sample(seq\_len(nrow(eye.dia)), size = smp\_size)

train.log <- eye.dia[train\_ind\_log, ]

test.log <- eye.dia[-train\_ind\_log, ]

#Creating model

eye.model <- glm(Class~., data = train.log, family = binomial)

#Testing with test data

predict.model <- predict(eye.model, newdata = test.log, type = "response")

#Calculating error

mean(predict.model != test.log$Class)

Model has 1 percent error:



#Typing new data to predict

new.data.2 <- data.frame(1, 22, 22 , 22, 19 , 18, 14,49.89, 17.77, 5.27, 0.77, 0.0186, 0.0068, 0.0039, 0.0039, 0.4869, 0.100, 1)

column.names <- c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "R")

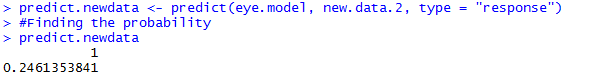
names(new.data.2) <- column.names

#Predicting values

predict.newdata <- predict(eye.model, new.data.2, type = "response")

#Finding the probability

predict.newdata



**Neural Net:**

#reading data

eye.dia<-read.csv(file.choose())

#scale the data

eye.scale <- scale(eye.dia)

#splitting data

smp\_size\_nn <- floor(0.60 \* nrow(eye.scale))

## set the seed to make your partition reproductible

set.seed(123)

train\_ind\_nn <- sample(seq\_len(nrow(eye.scale)), size = smp\_size)

train.nn <- eye.dia[train\_ind\_nn, ]

test.nn <- eye.dia[-train\_ind\_nn, ]

## build the neural network (NN)

#Train the neural network

#Going to have 7 hidden layers

#Threshold is a numeric value specifying the threshold for the partial

#derivatives of the error function as stopping criteria.

install.packages("neuralnet")

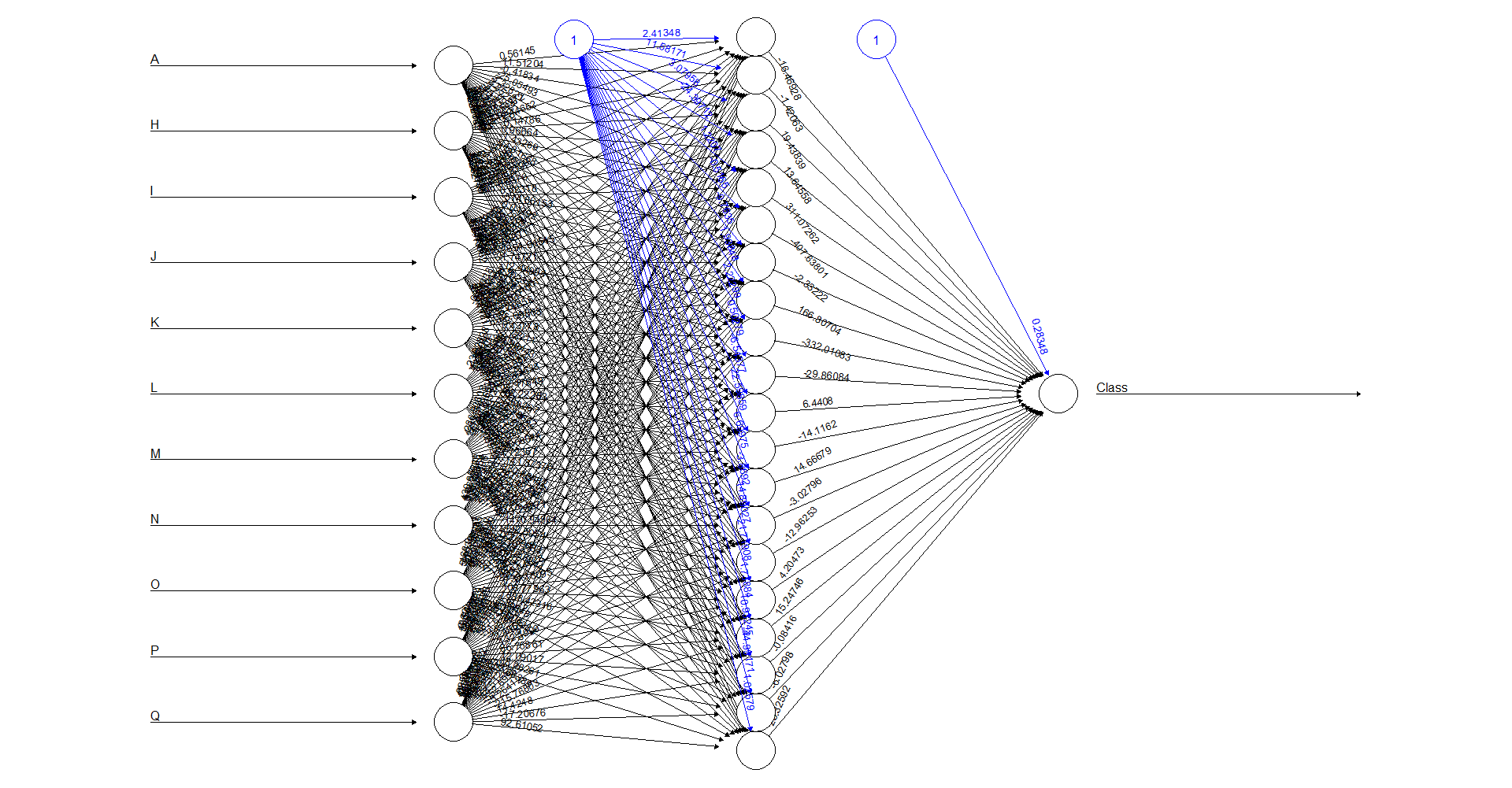
library(neuralnet)

eye.net <- neuralnet(Class~A+B+C+D+E+F+G+H+I+J+K+L+M+N+O+P+Q+R, data = train.nn, hidden = 20,lifesign = "minimal",linear.output = T, threshold=0.1)



#Plot the neural network

plot(eye.net, rep = "best")



## test the resulting output

temp\_test <- subset(test.nn, select = c("A", "B","C","D","E","F","G","H","I","J","K","L","M","N","O","P","Q","R"))

eye.net.results <- compute(eye.net, temp\_test)

head(temp\_test)

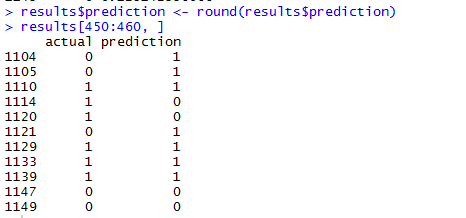
head(test)

results <- data.frame(actual = test.nn[,19], prediction =eye.net.results$net.result)

results[450:460, ]

results$prediction <- round(results$prediction)

results[450:460, ]



Tree Classification:

##Tree Classification

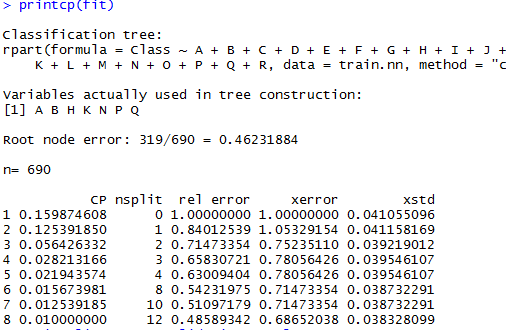
install.packages("rpart")

library(rpart)

fit <- rpart(Class~A+B+C+D+E+F+G+H+I+J+K+L+M+N+O+P+Q+R, method="class", data=train.nn)

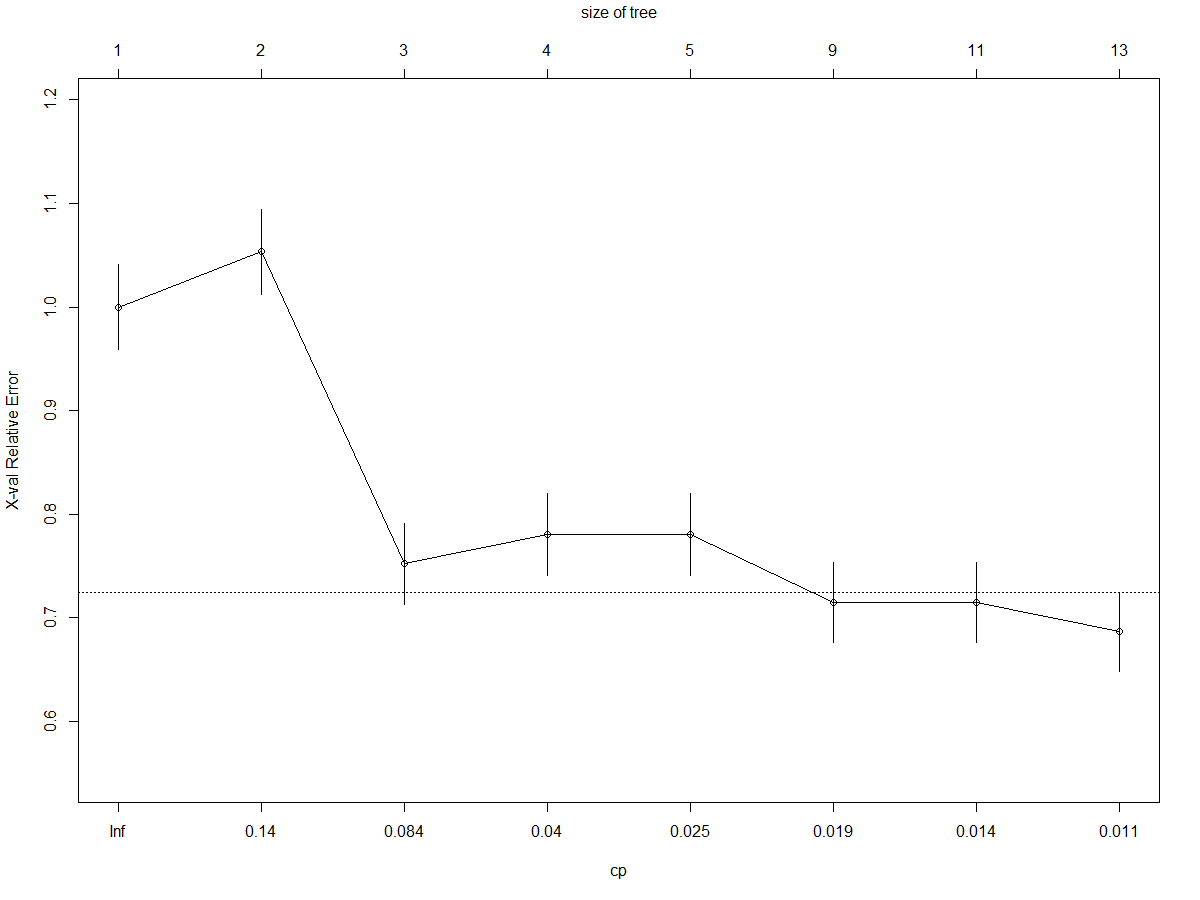
# display the results

printcp(fit)



# visualize cross-validation results

plotcp(fit)



# detailed summary of splits

summary(fit)

Call:

rpart(formula = Class ~ A + B + C + D + E + F + G + H + I + J +

K + L + M + N + O + P + Q + R, data = train.nn, method = "class")

n= 690

CP nsplit rel error xerror xstd

1 0.15987460815 0 1.0000000000 1.0000000000 0.04105509561

2 0.12539184953 1 0.8401253918 1.0532915361 0.04115816854

3 0.05642633229 2 0.7147335423 0.7523510972 0.03921901205

4 0.02821316614 3 0.6583072100 0.7805642633 0.03954610725

5 0.02194357367 4 0.6300940439 0.7805642633 0.03954610725

6 0.01567398119 8 0.5423197492 0.7147335423 0.03873229071

7 0.01253918495 10 0.5109717868 0.7147335423 0.03873229071

8 0.01000000000 12 0.4858934169 0.6865203762 0.03832809926

Variable importance

B C D E F G N O K H M J Q L I P A

13 12 12 11 9 7 6 4 4 4 4 3 3 3 2 2 2

Node number 1: 690 observations, complexity param=0.1598746082

predicted class=1 expected loss=0.4623188406 P(node) =1

class counts: 319 371

probabilities: 0.462 0.538

left son=2 (535 obs) right son=3 (155 obs)

Primary splits:

B < 57.5 to the left, improve=34.69400148, (0 missing)

C < 56.5 to the left, improve=31.07468751, (0 missing)

N < 0.007392 to the left, improve=30.28471598, (0 missing)

D < 55.5 to the left, improve=28.49746932, (0 missing)

O < 0.005771 to the left, improve=28.18713269, (0 missing)

Surrogate splits:

C < 56.5 to the left, agree=0.988, adj=0.948, (0 split)

D < 55.5 to the left, agree=0.981, adj=0.916, (0 split)

E < 52.5 to the left, agree=0.958, adj=0.813, (0 split)

F < 48.5 to the left, agree=0.925, adj=0.665, (0 split)

G < 36.5 to the left, agree=0.897, adj=0.542, (0 split)

Node number 2: 535 observations, complexity param=0.1253918495

predicted class=0 expected loss=0.4523364486 P(node) =0.7753623188

class counts: 293 242

probabilities: 0.548 0.452

left son=4 (483 obs) right son=5 (52 obs)

Primary splits:

N < 0.06238 to the left, improve=21.526238030, (0 missing)

O < 0.0244575 to the left, improve=15.743560940, (0 missing)

M < 0.086986 to the left, improve=14.226544500, (0 missing)

H < 127.8097455 to the left, improve=13.674797140, (0 missing)

L < 0.118709 to the left, improve= 8.295625829, (0 missing)

Surrogate splits:

O < 0.0322075 to the left, agree=0.974, adj=0.731, (0 split)

M < 0.1667775 to the left, agree=0.959, adj=0.577, (0 split)

L < 1.4629525 to the left, agree=0.920, adj=0.173, (0 split)

K < 15.2998025 to the left, agree=0.908, adj=0.058, (0 split)

Node number 3: 155 observations

predicted class=1 expected loss=0.1677419355 P(node) =0.2246376812

class counts: 26 129

probabilities: 0.168 0.832

Node number 4: 483 observations, complexity param=0.05642633229

predicted class=0 expected loss=0.4057971014 P(node) =0.7

class counts: 287 196

probabilities: 0.594 0.406

left son=8 (417 obs) right son=9 (66 obs)

Primary splits:

H < 127.8097455 to the left, improve=8.127885044, (0 missing)

K < 0.014747 to the right, improve=5.259302167, (0 missing)

A < 0.5 to the right, improve=5.131197559, (0 missing)

Q < 0.0908065 to the right, improve=4.381817258, (0 missing)

I < 5.3231155 to the right, improve=3.223346396, (0 missing)

Surrogate splits:

I < 61.295939 to the left, agree=0.909, adj=0.333, (0 split)

J < 24.757428 to the left, agree=0.907, adj=0.318, (0 split)

K < 4.729295 to the left, agree=0.884, adj=0.152, (0 split)

Node number 5: 52 observations

predicted class=1 expected loss=0.1153846154 P(node) =0.07536231884

class counts: 6 46

probabilities: 0.115 0.885

Node number 8: 417 observations, complexity param=0.02821316614

predicted class=0 expected loss=0.3693045564 P(node) =0.6043478261

class counts: 263 154

probabilities: 0.631 0.369

left son=16 (310 obs) right son=17 (107 obs)

Primary splits:

K < 0.034637 to the right, improve=8.590765832, (0 missing)

J < 2.56611 to the right, improve=7.505168145, (0 missing)

H < 30.0974785 to the right, improve=5.767222991, (0 missing)

I < 5.3231155 to the right, improve=5.382534349, (0 missing)

Q < 0.0908065 to the right, improve=5.121636053, (0 missing)

Surrogate splits:

J < 0.9011755 to the right, agree=0.890, adj=0.570, (0 split)

H < 17.0617135 to the right, agree=0.861, adj=0.458, (0 split)

I < 4.535492 to the right, agree=0.839, adj=0.374, (0 split)

L < 0.0009895 to the right, agree=0.823, adj=0.308, (0 split)

E < 2.5 to the right, agree=0.746, adj=0.009, (0 split)

Node number 9: 66 observations, complexity param=0.01253918495

predicted class=1 expected loss=0.3636363636 P(node) =0.09565217391

class counts: 24 42

probabilities: 0.364 0.636

left son=18 (40 obs) right son=19 (26 obs)

Primary splits:

B < 21.5 to the left, improve=3.776223776, (0 missing)

L < 0.006141 to the left, improve=3.678707847, (0 missing)

D < 18.5 to the left, improve=3.382320905, (0 missing)

C < 21.5 to the left, improve=3.337649667, (0 missing)

P < 0.5331905 to the left, improve=3.156565657, (0 missing)

Surrogate splits:

C < 21.5 to the left, agree=0.985, adj=0.962, (0 split)

D < 19.5 to the left, agree=0.970, adj=0.923, (0 split)

E < 17.5 to the left, agree=0.924, adj=0.808, (0 split)

F < 13.5 to the left, agree=0.924, adj=0.808, (0 split)

G < 10.5 to the left, agree=0.848, adj=0.615, (0 split)

Node number 16: 310 observations, complexity param=0.02194357367

predicted class=0 expected loss=0.3096774194 P(node) =0.4492753623

class counts: 214 96

probabilities: 0.690 0.310

left son=32 (267 obs) right son=33 (43 obs)

Primary splits:

Q < 0.0908065 to the right, improve=6.164093832, (0 missing)

A < 0.5 to the right, improve=5.058064516, (0 missing)

P < 0.554281 to the left, improve=2.510454002, (0 missing)

K < 2.5534135 to the right, improve=2.434021815, (0 missing)

H < 11.3429745 to the right, improve=2.344858633, (0 missing)

Surrogate splits:

G < 43.5 to the left, agree=0.871, adj=0.070, (0 split)

F < 50.5 to the left, agree=0.865, adj=0.023, (0 split)

Node number 17: 107 observations, complexity param=0.02194357367

predicted class=1 expected loss=0.4579439252 P(node) =0.1550724638

class counts: 49 58

probabilities: 0.458 0.542

left son=34 (87 obs) right son=35 (20 obs)

Primary splits:

B < 44.5 to the left, improve=3.273219465, (0 missing)

C < 44.5 to the left, improve=3.273219465, (0 missing)

J < 0.071816 to the left, improve=2.776040782, (0 missing)

Q < 0.1199965 to the right, improve=2.606477012, (0 missing)

D < 43 to the left, improve=2.525703522, (0 missing)

Surrogate splits:

C < 44.5 to the left, agree=1.000, adj=1.00, (0 split)

D < 43 to the left, agree=0.991, adj=0.95, (0 split)

E < 40.5 to the left, agree=0.972, adj=0.85, (0 split)

F < 38.5 to the left, agree=0.953, adj=0.75, (0 split)

G < 28.5 to the left, agree=0.879, adj=0.35, (0 split)

Node number 18: 40 observations, complexity param=0.01253918495

predicted class=0 expected loss=0.5 P(node) =0.05797101449

class counts: 20 20

probabilities: 0.500 0.500

left son=36 (26 obs) right son=37 (14 obs)

Primary splits:

P < 0.5305125 to the left, improve=3.516483516, (0 missing)

Q < 0.1030085 to the right, improve=2.506393862, (0 missing)

K < 1.2529585 to the right, improve=2.400000000, (0 missing)

N < 0.004036 to the right, improve=2.400000000, (0 missing)

H < 144.5870175 to the left, improve=2.164502165, (0 missing)

Surrogate splits:

K < 1.6983945 to the right, agree=0.725, adj=0.214, (0 split)

H < 158.610309 to the right, agree=0.675, adj=0.071, (0 split)

Q < 0.097139 to the right, agree=0.675, adj=0.071, (0 split)

Node number 19: 26 observations

predicted class=1 expected loss=0.1538461538 P(node) =0.03768115942

class counts: 4 22

probabilities: 0.154 0.846

Node number 32: 267 observations, complexity param=0.01567398119

predicted class=0 expected loss=0.2696629213 P(node) =0.3869565217

class counts: 195 72

probabilities: 0.730 0.270

left son=64 (237 obs) right son=65 (30 obs)

Primary splits:

A < 0.5 to the right, improve=5.962632153, (0 missing)

K < 2.5534135 to the right, improve=2.620871983, (0 missing)

N < 0.0019425 to the right, improve=2.200254979, (0 missing)

O < 0.0004815 to the right, improve=1.959571232, (0 missing)

Q < 0.093008 to the left, improve=1.827362855, (0 missing)

Surrogate splits:

H < 11.3429745 to the right, agree=0.899, adj=0.1, (0 split)

Node number 33: 43 observations, complexity param=0.02194357367

predicted class=1 expected loss=0.4418604651 P(node) =0.06231884058

class counts: 19 24

probabilities: 0.442 0.558

left son=66 (13 obs) right son=67 (30 obs)

Primary splits:

B < 15.5 to the left, improve=6.091353608, (0 missing)

C < 13.5 to the left, improve=4.186575053, (0 missing)

D < 12.5 to the left, improve=4.186575053, (0 missing)

E < 12.5 to the left, improve=3.081223508, (0 missing)

K < 0.404308 to the right, improve=1.875968992, (0 missing)

Surrogate splits:

C < 13.5 to the left, agree=0.953, adj=0.846, (0 split)

D < 12.5 to the left, agree=0.953, adj=0.846, (0 split)

E < 12.5 to the left, agree=0.930, adj=0.769, (0 split)

F < 9.5 to the left, agree=0.860, adj=0.538, (0 split)

G < 6.5 to the left, agree=0.837, adj=0.462, (0 split)

Node number 34: 87 observations, complexity param=0.02194357367

predicted class=0 expected loss=0.4827586207 P(node) =0.1260869565

class counts: 45 42

probabilities: 0.517 0.483

left son=68 (52 obs) right son=69 (35 obs)

Primary splits:

Q < 0.110513 to the right, improve=3.561462675, (0 missing)

E < 38.5 to the right, improve=2.255237887, (0 missing)

F < 36.5 to the right, improve=2.255237887, (0 missing)

H < 3.546562 to the left, improve=1.758990148, (0 missing)

B < 11.5 to the right, improve=1.452753474, (0 missing)

Surrogate splits:

O < 0.000957 to the left, agree=0.667, adj=0.171, (0 split)

L < 0.000959 to the left, agree=0.655, adj=0.143, (0 split)

N < 0.0017305 to the left, agree=0.655, adj=0.143, (0 split)

P < 0.5335545 to the left, agree=0.655, adj=0.143, (0 split)

M < 0.000957 to the left, agree=0.644, adj=0.114, (0 split)

Node number 35: 20 observations

predicted class=1 expected loss=0.2 P(node) =0.02898550725

class counts: 4 16

probabilities: 0.200 0.800

Node number 36: 26 observations

predicted class=0 expected loss=0.3461538462 P(node) =0.03768115942

class counts: 17 9

probabilities: 0.654 0.346

Node number 37: 14 observations

predicted class=1 expected loss=0.2142857143 P(node) =0.02028985507

class counts: 3 11

probabilities: 0.214 0.786

Node number 64: 237 observations

predicted class=0 expected loss=0.2320675105 P(node) =0.3434782609

class counts: 182 55

probabilities: 0.768 0.232

Node number 65: 30 observations, complexity param=0.01567398119

predicted class=1 expected loss=0.4333333333 P(node) =0.04347826087

class counts: 13 17

probabilities: 0.433 0.567

left son=130 (12 obs) right son=131 (18 obs)

Primary splits:

K < 0.1885995 to the left, improve=4.011111111, (0 missing)

P < 0.544232 to the right, improve=3.050793651, (0 missing)

J < 1.4423495 to the left, improve=2.187878788, (0 missing)

Q < 0.1060835 to the right, improve=2.074242424, (0 missing)

I < 14.936477 to the left, improve=1.666666667, (0 missing)

Surrogate splits:

I < 14.936477 to the left, agree=0.900, adj=0.750, (0 split)

J < 2.124169 to the left, agree=0.900, adj=0.750, (0 split)

L < 0.0109 to the left, agree=0.867, adj=0.667, (0 split)

P < 0.541562 to the right, agree=0.867, adj=0.667, (0 split)

H < 24.0237435 to the left, agree=0.700, adj=0.250, (0 split)

Node number 66: 13 observations

predicted class=0 expected loss=0.1538461538 P(node) =0.01884057971

class counts: 11 2

probabilities: 0.846 0.154

Node number 67: 30 observations

predicted class=1 expected loss=0.2666666667 P(node) =0.04347826087

class counts: 8 22

probabilities: 0.267 0.733

Node number 68: 52 observations

predicted class=0 expected loss=0.3653846154 P(node) =0.07536231884

class counts: 33 19

probabilities: 0.635 0.365

Node number 69: 35 observations

predicted class=1 expected loss=0.3428571429 P(node) =0.05072463768

class counts: 12 23

probabilities: 0.343 0.657

Node number 130: 12 observations

predicted class=0 expected loss=0.25 P(node) =0.01739130435

class counts: 9 3

probabilities: 0.750 0.250

Node number 131: 18 observations

predicted class=1 expected loss=0.2222222222 P(node) =0.02608695652

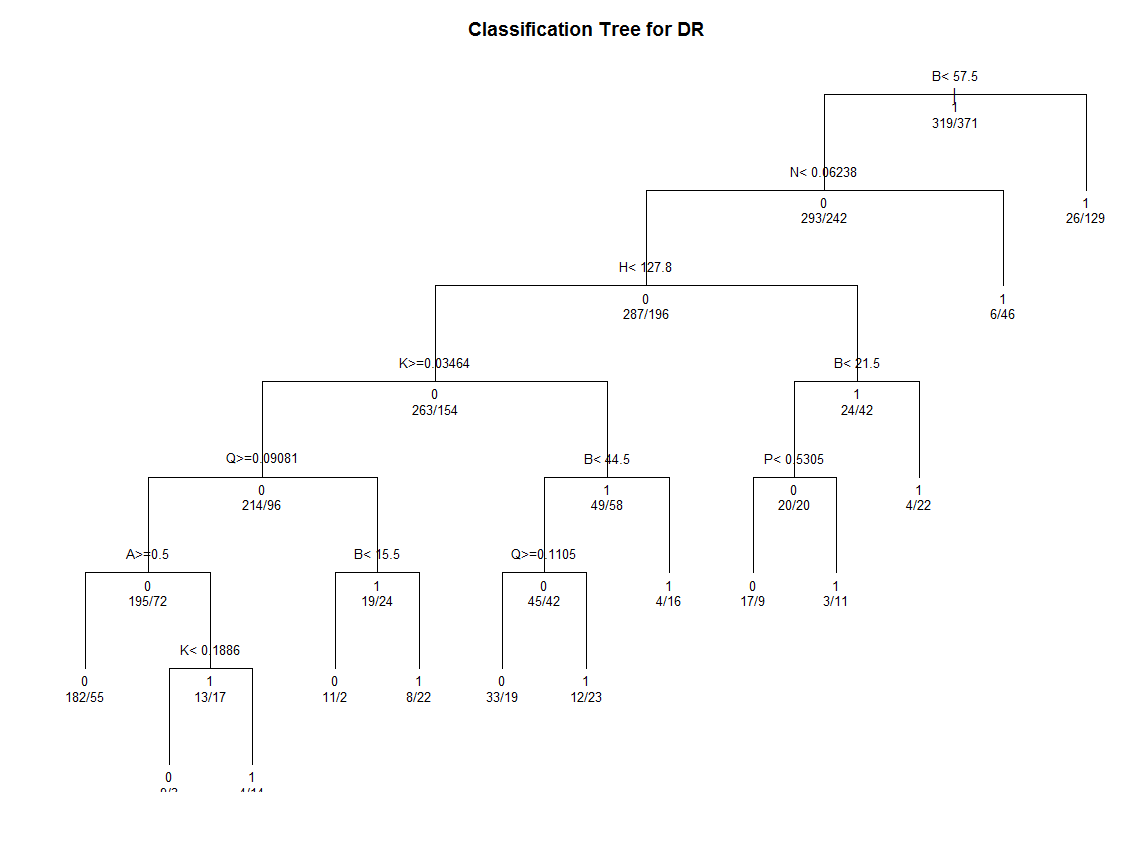
class counts: 4 14

probabilities: 0.222 0.778

# plot tree

plot(fit, uniform=TRUE, main="Classification Tree for DR")

text(fit, use.n=TRUE, all=TRUE, cex=.8)



**KNN:**

#reading data

eye.dia<-read.csv(file.choose())

#splitting data

smp\_size <- floor(0.60 \* nrow(eye.dia))

## set the seed to make your partition reproductible

set.seed(123)

train\_ind\_knn <- sample(seq\_len(nrow(eye.dia)), size = smp\_size)

train.knn <- eye.dia[train\_ind\_knn, ]

test.knn <- eye.dia[-train\_ind\_knn, ]

install.packages("class")

library(class)

knn\_predict <- knn(train.knn,test.knn,train.knn[,19], k = 17, use.all = TRUE)

knn\_predict[1:3]

table(knn\_predict,test.knn[,19] )

mean(knn\_predict == test.knn[,19])

> mean(knn\_predict == test.knn[,19])

[1] 0.6919739696