**Default**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), max\_dep+th=6, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.468874

[2] train-rmse:0.369729

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 16.091, df = 1, p-value = 6.038e-05

FALSE TRUE

0 9574 85

1 56 4

**Max\_Depth (Default = 6) [0,** **∞)**

**Max Depth = 1**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), max\_depth=1, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.494256

[2] train-rmse:0.423755

>

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 19.384, df = 1, p-value = 1.069e-05

chi\_t

FALSE TRUE

0 9658 1

1 59 1

**Max Depth = 30**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), max\_depth=30, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.450427

[2] train-rmse:0.332809

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 5.7099, df = 1, p-value = 0.01687

chi\_t

FALSE TRUE

0 9411 248

1 55 5

**Max Depth = 1000 produces same result as Max Depth = 30.**

**Min\_Child\_Weight (Default = 1) [0,** **∞)**

**Min\_Child\_Weight = 0 produces same result as default**

**Min\_Child\_Weight = 10**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), min\_child\_weight = 10, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.472867

[2] train-rmse:0.371482

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 19.52, df = 1, p-value = 9.957e-06

chi\_t

FALSE TRUE

0 9586 73

1 56 4

**Min\_Child\_Weight = 1000 produces same chi2 as 10.**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), min\_child\_weight = 1000, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.496955

[2] train-rmse:0.429520

**Gamma (Default = 0) [0,** **∞)**

**Gamma = 1**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), gamma = 1, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.472976

[2] train-rmse:0.382588

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 37.78, df = 1, p-value = 7.918e-10

chi\_t

FALSE TRUE

0 9618 41

1 56 4

**Gamma = 100 produces same result as Gamma = 1.**

**Subsample (Default = 1) (0,** **1]**

**Subsample = 0.5**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), subsample = .5, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.479158

[2] train-rmse:0.395314

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.021298, df = 1, p-value = 0.884

chi\_t

FALSE TRUE

0 9594 65

1 59 1

**Subsample = 0.75**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), subsample=0.75, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.472098

[2] train-rmse:0.377165

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 29.262, df = 1, p-value = 6.323e-08

chi\_t

FALSE TRUE

0 9576 83

1 55 5

**Colsample\_bytree (Default = 1) (0,** **1]**

**Colsample\_bytree = 0.5**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), colsample\_bytree = .5, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.470430

[2] train-rmse:0.371125

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 6.1722e-29, df = 1, p-value = 1

chi\_t

FALSE TRUE

0 9565 94

1 59 1

**ETA (Default = 0.3) [0,** **1]**

**eta = 0**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), eta = 0 , nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.611851

[2] train-rmse:0.611851

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 17.969, df = 1, p-value = 2.245e-05

chi\_t

FALSE TRUE

0 9617 42

1 57 3

**eta = 1**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), eta = 1 , nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.257150

[2] train-rmse:0.192755

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 15.609, df = 1, p-value = 7.789e-05

chi\_t

FALSE TRUE

0 9572 87

1 56 4

**eta = 0.01 Nrounds = 2**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), eta = 0.01, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.606922

[2] train-rmse:0.602049

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 17.017, df = 1, p-value = 3.704e-05

chi\_t

FALSE TRUE

0 9615 44

1 57 3

**eta = 0.01 Nrounds = 20**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), eta = 0.01, nrounds = 20, objective = "reg:squarederror")

[1] train-rmse:0.606922

[2] train-rmse:0.602049

[3] train-rmse:0.597233

[4] train-rmse:0.592470

[5] train-rmse:0.587717

[6] train-rmse:0.583018

[7] train-rmse:0.578364

[8] train-rmse:0.573763

[9] train-rmse:0.569238

[10] train-rmse:0.564823

[11] train-rmse:0.560358

[12] train-rmse:0.555981

[13] train-rmse:0.551642

[14] train-rmse:0.547358

[15] train-rmse:0.543115

[16] train-rmse:0.538911

[17] train-rmse:0.534708

[18] train-rmse:0.530656

[19] train-rmse:0.526596

[20] train-rmse:0.522606

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 6.3243, df = 1, p-value = 0.01191

chi\_t

FALSE TRUE

0 9500 159

1 56 4

**Nrounds (No Default) [1,** **∞)**

**1 Round produces same 1st round RMSE as default**

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 17.969, df = 1, p-value = 2.245e-05

> chi\_t

FALSE TRUE

0 9617 42

1 57 3

**100 Rounds -> Descending RMSE until**

[95] train-rmse:0.000414

[96] train-rmse:0.000394

[97] train-rmse:0.000394

[98] train-rmse:0.000394

[99] train-rmse:0.000394

[100] train-rmse:0.000394

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 1.0404, df = 1, p-value = 0.3077

FALSE TRUE

0 7384 2275

1 42 18

**Check combos of parameters that work well individually**

**Max\_depth, nrounds, eta**

my\_xgbst = xgboost(label =as.matrix(residframe), data = as.matrix(Chr22geno), eta = 1, max\_depth = 6, nrounds = 11 , objective = "reg:squarederror")

[1] train-rmse:0.257150

[2] train-rmse:0.192755

[3] train-rmse:0.172290

[4] train-rmse:0.134465

[5] train-rmse:0.106627

[6] train-rmse:0.092599

[7] train-rmse:0.075376

[8] train-rmse:0.057891

[9] train-rmse:0.044186

[10] train-rmse:0.038172

[11] train-rmse:0.032175

chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 6.5023, df = 1, p-value = 0.01077

chi\_t

FALSE TRUE

0 9255 404

1 53 7

**Write cerca abstract**

**Nrounds = 1**

**Max\_depth = 6**

**Min\_child\_weight = 100**

**Gamma = 1**

**Subsample = 1**

**Eta = doesn’t matter, only one round**

importance\_matrix

Feature Gain Cover Frequency

1: rs9926138 0.1979540 0.21302578 0.1428571

2: rs2280141 0.1951901 0.30122117 0.1428571

3: rs1549114 0.1422911 0.16987788 0.1428571

4: rs870445 0.1275825 0.09525102 0.1428571

5: rs2886616 0.1231954 0.07462687 0.1428571

6: rs2588548 0.1137074 0.05780190 0.1428571

7: rs2248799 0.1000794 0.08819539 0.1428571

Rs2280141 : <https://www.ncbi.nlm.nih.gov/snp/rs2280141#publications>

<https://diseases.jensenlab.org/Entity?order=textmining,knowledge,experiments&textmining=10&knowledge=10&experiments=10&type1=9606&type2=-26&id1=ENSP00000357986>

Rs9926138 : <https://diseases.jensenlab.org/Entity?order=textmining,knowledge,experiments&textmining=10&knowledge=10&experiments=10&type1=9606&type2=-26&id1=ENSP00000309117>

Rs2248799

<https://www.ncbi.nlm.nih.gov/snp/rs2248799#publications>

XGBoost on Breast Cancer Data

Make a note in paper draft of where data is

>

> # Get the number of snp in Chr22\_snp

> numberofSNPs<-dim(Chr22\_snp)[1]

>

> # Loop through each SNP and store the results in res.list(Not working, ERROR)

> for (i in 1:totalCol){

+

+ ## Create a formula that combining all the rsIDs

+ form=as.formula(paste("as.factor(cbind(Chr22\_sim$ordinal))~",combos$rs[i]));

+

+ ## Error using cloglog "why"

+ ## Error details: Error in optim(s0, fmin, gmin, method = "BFGS", ...) : initial value in 'vmmin' is not finite

+ #res.list[[i]] <- polr(form, data=Chr22\_sim, Hess=TRUE,na.action=na.omit,model = TRUE,method = c("cloglog")) - got error when using cloglog

+

+ ## Without specifying any method

+ res.list[[i]] <- polr(form, data=Chr22\_sim, Hess=TRUE,na.action=na.omit,model = TRUE)

+

+ ## store summary of each regressed SNP

+ table <- coef(summary(res.list[[i]]))

+

+ ## calculate and store p values

+ p <- pnorm(abs(table[, "t value"]), lower.tail = FALSE) \* 2

+

+ ## debug

+ #print(i)

+ #print(p[1:1])

+

+ ## if p < 0.05, then store it to the significantSNPs

+ if(p[1:1] < 0.05){

+ significantSNPs = append(significantSNPs, i)

+ }else{

+ notSignificantSNPs = append(notSignificantSNPs, i)

+ }

+

+ }

Warning message:

glm.fit: fitted probabilities numerically 0 or 1 occurred

>

> #save(significantSNPs,file="significantSNPs.RData")

> #save(notSignificantSNPs,file="notSignificantSNPs.RData")

>

> ## Convert ordinal to factor

> #Chr22\_sim$ordinal = as.factor(Chr22\_sim$ordinal)

>

> ## The significantSNPs contains the column index, which you can grab the

> #Chr22\_sim[,significantSNPs[1]]

>

>

> # Loop regression through the SNPs with small p-values (we only need the one that has 'main effect')

> #significantSNPs\_maineffect<-Chr22\_snp%>%

> #filter(type=="main effect", #we only want Casual SNPs

> # pvals<0.05/numberofSNPs)

>

> significantSNPs\_maineffect = which(Chr22\_snp$type=="main effect",Chr22\_snp$pvals<0.05/numberofSNPs)

> significantSNPs\_intersect = intersect(significantSNPs,significantSNPs\_maineffect)

>

> #length(significantSNPs\_maineffect) - 50

> #length(significantSNPs\_intersect) - 31

> #length(significantSNPs)- 4020

>

> Chr22\_sim$ordinal = factor(Chr22\_sim$ordinal)

> ordinalCol=which(names(Chr22\_sim)=="ordinal")

> #length(Chr22\_sim$ordinal) - 1023

>

> # Create new data frame (only contains phenotype col and SNP that we want to use)

> forReg=Chr22\_sim[,c(significantSNPs\_intersect,ordinalCol)]

> # length(forReg) - 32

>

> # Period = use every other stuff as predictors

> # USE CLOGLOG HEREEE

> sigSNP.plr <- polr(ordinal~., data=forReg, Hess=TRUE,na.action=na.omit,model = TRUE, method=c("cloglog"))

> # length(sigSNP.plr) -18

>

>

> # check for missing values

> # length(which(is.na(sigSNP.resid))) - 0

>

> # Extract residuals from polr model

> sigSNP.resid <- presid(sigSNP.plr)

>

>

> # Take all the rows, and take all cols excluding the intersect SNP

> forForest=Chr22\_sim[,-c(significantSNPs\_intersect)]

>

> Chr22geno <- read.csv("C:/Users/ericm/OneDrive/Desktop/Chr22\_sim.csv", header = TRUE) #sim

> Chr22geno = as.data.frame(Chr22geno)

> #Chr22geno = Chr22geno%>%

> #select(-c(pheno1,binary,ordinal,all\_of(significantSNPs\_intersect)))

> forXGB = merge(Chr22geno, forForest)

>

> Chr22SNPs\_nosig <- Chr22\_snp[Chr22\_snp[,1] %in% names(forXGB),]

>

> forXGB = forXGB[names(forXGB) %in% Chr22SNPs\_nosig$rsID]

>

> set.seed(1);

>

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.468144

[2] train-rmse:0.332562

[3] train-rmse:0.240020

[4] train-rmse:0.179917

[5] train-rmse:0.141609

[6] train-rmse:0.119998

[7] train-rmse:0.109236

[8] train-rmse:0.103029

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 15.973, df = 1, p-value = 6.424e-05

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9285 384

1 51 9

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 6, nrounds = 1, objective = "reg:squarederror")

[1] train-rmse:0.540690

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 5.9165e-31, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9633 36

1 60 0

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 3, nrounds = 1, objective = "reg:squarederror")

[1] train-rmse:0.563165

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4043e-25, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9662 7

1 60 0

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 1, nrounds = 1, objective = "reg:squarederror")

[1] train-rmse:0.567506

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 1.066e-22, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9668 1

1 60 0

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 1, nrounds = 5, objective = "reg:squarederror")

[1] train-rmse:0.567506

[2] train-rmse:0.504964

[3] train-rmse:0.470461

[4] train-rmse:0.451956

[5] train-rmse:0.441927

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 5.3098e-24, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9664 5

1 60 0

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 1, nrounds = 50, objective = "reg:squarederror")

[1] train-rmse:0.567506

[2] train-rmse:0.504964

[3] train-rmse:0.470461

[4] train-rmse:0.451956

[5] train-rmse:0.441927

[6] train-rmse:0.436277

[7] train-rmse:0.432840

[8] train-rmse:0.430613

[9] train-rmse:0.428930

[10] train-rmse:0.427526

[11] train-rmse:0.426225

[12] train-rmse:0.424969

[13] train-rmse:0.423790

[14] train-rmse:0.422654

[15] train-rmse:0.421535

[16] train-rmse:0.420473

[17] train-rmse:0.419454

[18] train-rmse:0.418416

[19] train-rmse:0.417428

[20] train-rmse:0.416406

[21] train-rmse:0.415396

[22] train-rmse:0.414440

[23] train-rmse:0.413473

[24] train-rmse:0.412534

[25] train-rmse:0.411589

[26] train-rmse:0.410695

[27] train-rmse:0.409777

[28] train-rmse:0.408870

[29] train-rmse:0.407988

[30] train-rmse:0.407143

[31] train-rmse:0.406293

[32] train-rmse:0.405421

[33] train-rmse:0.404559

[34] train-rmse:0.403737

[35] train-rmse:0.402870

[36] train-rmse:0.402009

[37] train-rmse:0.401195

[38] train-rmse:0.400378

[39] train-rmse:0.399562

[40] train-rmse:0.398731

[41] train-rmse:0.397906

[42] train-rmse:0.397119

[43] train-rmse:0.396336

[44] train-rmse:0.395540

[45] train-rmse:0.394781

[46] train-rmse:0.394034

[47] train-rmse:0.393270

[48] train-rmse:0.392529

[49] train-rmse:0.391760

[50] train-rmse:0.390992

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 1.7825e-29, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9620 49

1 60 0

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 10, nrounds = 50, objective = "reg:squarederror")

[1] train-rmse:0.512316

[2] train-rmse:0.392802

[3] train-rmse:0.303295

[4] train-rmse:0.232390

[5] train-rmse:0.185770

[6] train-rmse:0.145969

[7] train-rmse:0.111528

[8] train-rmse:0.087591

[9] train-rmse:0.071577

[10] train-rmse:0.056805

[11] train-rmse:0.047380

[12] train-rmse:0.037485

[13] train-rmse:0.031192

[14] train-rmse:0.026415

[15] train-rmse:0.021267

[16] train-rmse:0.018207

[17] train-rmse:0.015206

[18] train-rmse:0.012811

[19] train-rmse:0.010555

[20] train-rmse:0.008858

[21] train-rmse:0.007381

[22] train-rmse:0.006382

[23] train-rmse:0.005593

[24] train-rmse:0.004542

[25] train-rmse:0.003681

[26] train-rmse:0.003348

[27] train-rmse:0.002703

[28] train-rmse:0.002352

[29] train-rmse:0.002101

[30] train-rmse:0.001789

[31] train-rmse:0.001580

[32] train-rmse:0.001394

[33] train-rmse:0.001240

[34] train-rmse:0.001027

[35] train-rmse:0.000902

[36] train-rmse:0.000856

[37] train-rmse:0.000709

[38] train-rmse:0.000618

[39] train-rmse:0.000553

[40] train-rmse:0.000482

[41] train-rmse:0.000430

[42] train-rmse:0.000370

[43] train-rmse:0.000331

[44] train-rmse:0.000331

[45] train-rmse:0.000331

[46] train-rmse:0.000331

[47] train-rmse:0.000331

[48] train-rmse:0.000331

[49] train-rmse:0.000331

[50] train-rmse:0.000331

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 2.4104, df = 1, p-value = 0.1205

> chi\_t

FALSE TRUE

0 6921 2748

1 37 23

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 10, nrounds = 40, objective = "reg:squarederror")

[1] train-rmse:0.512316

[2] train-rmse:0.392802

[3] train-rmse:0.303295

[4] train-rmse:0.232390

[5] train-rmse:0.185770

[6] train-rmse:0.145969

[7] train-rmse:0.111528

[8] train-rmse:0.087591

[9] train-rmse:0.071577

[10] train-rmse:0.056805

[11] train-rmse:0.047380

[12] train-rmse:0.037485

[13] train-rmse:0.031192

[14] train-rmse:0.026415

[15] train-rmse:0.021267

[16] train-rmse:0.018207

[17] train-rmse:0.015206

[18] train-rmse:0.012811

[19] train-rmse:0.010555

[20] train-rmse:0.008858

[21] train-rmse:0.007381

[22] train-rmse:0.006382

[23] train-rmse:0.005593

[24] train-rmse:0.004542

[25] train-rmse:0.003681

[26] train-rmse:0.003348

[27] train-rmse:0.002703

[28] train-rmse:0.002352

[29] train-rmse:0.002101

[30] train-rmse:0.001789

[31] train-rmse:0.001580

[32] train-rmse:0.001394

[33] train-rmse:0.001240

[34] train-rmse:0.001027

[35] train-rmse:0.000902

[36] train-rmse:0.000856

[37] train-rmse:0.000709

[38] train-rmse:0.000618

[39] train-rmse:0.000553

[40] train-rmse:0.000482

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 2.0214, df = 1, p-value = 0.1551

> chi\_t

FALSE TRUE

0 7001 2668

1 38 22

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 10, nrounds = 10, objective = "reg:squarederror")

[1] train-rmse:0.512316

[2] train-rmse:0.392802

[3] train-rmse:0.303295

[4] train-rmse:0.232390

[5] train-rmse:0.185770

[6] train-rmse:0.145969

[7] train-rmse:0.111528

[8] train-rmse:0.087591

[9] train-rmse:0.071577

[10] train-rmse:0.056805

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.62378, df = 1, p-value = 0.4296

> chi\_t

FALSE TRUE

0 8609 1060

1 51 9

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.468144

[2] train-rmse:0.332562

[3] train-rmse:0.240020

[4] train-rmse:0.179917

[5] train-rmse:0.141609

[6] train-rmse:0.119998

[7] train-rmse:0.109236

[8] train-rmse:0.103029

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 15.973, df = 1, p-value = 6.424e-05

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9285 384

1 51 9

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.468144

[2] train-rmse:0.332562

[3] train-rmse:0.240020

[4] train-rmse:0.179917

[5] train-rmse:0.141609

[6] train-rmse:0.119998

[7] train-rmse:0.109236

[8] train-rmse:0.103029

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.498790

[2] train-rmse:0.369158

[3] train-rmse:0.274347

[4] train-rmse:0.204865

[5] train-rmse:0.153001

[6] train-rmse:0.114469

[7] train-rmse:0.085880

[8] train-rmse:0.064906

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.075883, df = 1, p-value = 0.783

> chi\_t

FALSE TRUE

0 8509 1160

1 54 6

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.498790

[2] train-rmse:0.369158

[3] train-rmse:0.274347

[4] train-rmse:0.204865

[5] train-rmse:0.153001

[6] train-rmse:0.114469

[7] train-rmse:0.085880

[8] train-rmse:0.064906

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.075883, df = 1, p-value = 0.783

> chi\_t

FALSE TRUE

0 8509 1160

1 54 6

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 1, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.519841

[2] train-rmse:0.414932

[3] train-rmse:0.346845

[4] train-rmse:0.300505

[5] train-rmse:0.274299

[6] train-rmse:0.260397

[7] train-rmse:0.250831

[8] train-rmse:0.244229

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 6.3274e-30, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9488 181

1 59 1

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 10, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.568831

[2] train-rmse:0.507815

[3] train-rmse:0.475037

[4] train-rmse:0.458106

[5] train-rmse:0.449570

[6] train-rmse:0.445324

[7] train-rmse:0.443226

[8] train-rmse:0.442194

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

Error in xgb.model.dt.tree(feature\_names = feature\_names, text = model\_text\_dump, :

Non-tree model detected! This function can only be used with tree models.

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 6.3274e-30, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9488 181

1 59 1

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.5, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.507869

[2] train-rmse:0.390067

[3] train-rmse:0.309166

[4] train-rmse:0.257906

[5] train-rmse:0.225433

[6] train-rmse:0.205534

[7] train-rmse:0.192551

[8] train-rmse:0.181716

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.075528, df = 1, p-value = 0.7835

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9367 302

1 59 1

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.25, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.503751

[2] train-rmse:0.380395

[3] train-rmse:0.291452

[4] train-rmse:0.230066

[5] train-rmse:0.189121

[6] train-rmse:0.164026

[7] train-rmse:0.146867

[8] train-rmse:0.137205

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.51779, df = 1, p-value = 0.4718

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9242 427

1 59 1

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.4, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.505745

[2] train-rmse:0.386609

[3] train-rmse:0.303895

[4] train-rmse:0.247679

[5] train-rmse:0.209312

[6] train-rmse:0.184633

[7] train-rmse:0.173841

[8] train-rmse:0.167130

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 2.127e-30, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9338 331

1 58 2

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.15, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.501441

[2] train-rmse:0.375136

[3] train-rmse:0.284134

[4] train-rmse:0.219592

[5] train-rmse:0.174552

[6] train-rmse:0.144155

[7] train-rmse:0.124800

[8] train-rmse:0.113126

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 3.2884e-29, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9128 541

1 57 3

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.1, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.500353

[2] train-rmse:0.372711

[3] train-rmse:0.279725

[4] train-rmse:0.214068

[5] train-rmse:0.167752

[6] train-rmse:0.135321

[7] train-rmse:0.114688

[8] train-rmse:0.100041

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.74864, df = 1, p-value = 0.3869

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9049 620

1 54 6

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.05, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.499462

[2] train-rmse:0.371167

[3] train-rmse:0.276301

[4] train-rmse:0.207944

[5] train-rmse:0.158655

[6] train-rmse:0.123866

[7] train-rmse:0.099511

[8] train-rmse:0.082724

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.3619, df = 1, p-value = 0.5475

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 8901 768

1 57 3

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), gamma = 0.08, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

[1] train-rmse:0.499974

[2] train-rmse:0.372254

[3] train-rmse:0.278655

[4] train-rmse:0.211244

[5] train-rmse:0.163365

[6] train-rmse:0.130741

[7] train-rmse:0.107594

[8] train-rmse:0.092921

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.034197, df = 1, p-value = 0.8533

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9003 666

1 55 5

> chi\_t

FALSE TRUE

0 9543 126

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.13, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.606958

[2] train-rmse:0.545661

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.3749, df = 1, p-value = 0.0022

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9544 125

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.12, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.612240

[2] train-rmse:0.554870

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.2613, df = 1, p-value = 0.00234

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9543 126

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.145, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.599056

[2] train-rmse:0.531956

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.3749, df = 1, p-value = 0.0022

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9544 125

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.15, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.596428

[2] train-rmse:0.527413

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.2613, df = 1, p-value = 0.00234

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9543 126

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.1425, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.600372

[2] train-rmse:0.534237

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.3749, df = 1, p-value = 0.0022

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9544 125

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.135, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.604321

[2] train-rmse:0.541059

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 5, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.597209

[2] train-rmse:0.528165

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 2.4891e-26, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9483 186

1 59 1

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 7, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.599470

[2] train-rmse:0.531550

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.30757, df = 1, p-value = 0.5792

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9515 154

1 58 2

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 8, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.600890

[2] train-rmse:0.534313

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 8.7506e-30, df = 1, p-value = 1

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9528 141

1 59 1

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 10, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.602456

[2] train-rmse:0.537728

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 4.0132, df = 1, p-value = 0.04515

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9548 121

1 57 3

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 8.5, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 8.25, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 8.1, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 8.001, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 50, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601205

[2] train-rmse:0.535931

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.0397, df = 1, p-value = 0.002642

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9541 128

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 11, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.602239

[2] train-rmse:0.537040

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 4.0132, df = 1, p-value = 0.04515

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9548 121

1 57 3

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 10, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.603159

[2] train-rmse:0.538478

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 0.85532, df = 1, p-value = 0.3551

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9554 115

1 58 2

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 3, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

[3] train-rmse:0.479958

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 4.4599, df = 1, p-value = 0.0347

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9477 192

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = .19, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 5, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.623806

[2] train-rmse:0.581840

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

Error in xgb.model.dt.tree(feature\_names = feature\_names, text = model\_text\_dump, :

Non-tree model detected! This function can only be used with tree models.

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 100, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.623806

[2] train-rmse:0.581840

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

Error in xgb.model.dt.tree(feature\_names = feature\_names, text = model\_text\_dump, :

Non-tree model detected! This function can only be used with tree models.

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0.15, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536373

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.2613, df = 1, p-value = 0.00234

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9543 126

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0.1, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601643

[2] train-rmse:0.536255

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 8.9315, df = 1, p-value = 0.002803

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9540 129

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601567

[2] train-rmse:0.536134

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 8.3168, df = 1, p-value = 0.003928

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9534 135

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0.18, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0.17, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0.16, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536411

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.3749, df = 1, p-value = 0.0022

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9544 125

1 56 4

> my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB),eta = 0.14, min\_child\_weight = 9, gamma = 0.17, max\_depth = 12, nrounds = 2, objective = "reg:squarederror")

[1] train-rmse:0.601687

[2] train-rmse:0.536459

> #my\_xgbst = xgboost(label =as.matrix(sigSNP.resid), data = as.matrix(forXGB), eta = .4, gamma = 0.15, min\_child\_weight = 9, max\_depth = 20, nrounds = 8, objective = "reg:squarederror")

>

> names = dimnames(data.matrix(names(forXGB)))

> snp\_names = names(forXGB);

> snp\_names = as.data.frame(snp\_names);

> importance\_matrix = xgb.importance(names, model = my\_xgbst)

>

> snp\_names[,2] = importance\_matrix[match(snp\_names[,1],importance\_matrix$Feature),2]

> snp\_names[is.na(snp\_names)] = 0;

>

> snp\_names[,3] = Chr22SNPs\_nosig[,3];

> snp\_names[,4] = Chr22SNPs\_nosig[,5];

> snp\_names[,5] = snp\_names[,3] + snp\_names[,4];

> colnames(snp\_names) <- c("SNP", "VariableImportance", "IsCausal", "InCluster", "CausesDisease")

>

> chi\_t = table(snp\_names[,5],snp\_names[,2]>0)

> chisq.test(chi\_t)

Pearson's Chi-squared test with Yates' continuity correction

data: chi\_t

X-squared = 9.4903, df = 1, p-value = 0.002066

Warning message:

In chisq.test(chi\_t) : Chi-squared approximation may be incorrect

> chi\_t

FALSE TRUE

0 9545 124

1 56 4

gamma = 0 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.01 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.02 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.03 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.04 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.05 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.06 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.07 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.08 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.09 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.1 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.11 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.12 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.13 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.14 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.15 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.16 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.17 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.18 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.19 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.2 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.21 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.22 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.23 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.24 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.25 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.26 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.27 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.28 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.29 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.3 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.31 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.32 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.33 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.34 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.35 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.36 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.37 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.38 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.39 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.4 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0.41 min\_child = 47 max\_depth = 5 pval = 0.1576059

gamma = 0 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.01 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.01 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.01 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.02 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.02 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.02 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.03 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.03 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.03 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.04 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.04 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.04 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.05 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.05 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.05 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.06 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.06 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.06 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.07 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.07 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.07 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.08 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.08 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.08 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.09 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.09 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.09 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.1 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.1 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.1 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.11 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.11 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.11 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.12 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.12 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.12 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.13 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.13 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.13 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.14 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.14 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.14 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.15 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.15 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.15 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.16 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.16 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.16 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.17 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.17 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.17 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.18 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.18 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.18 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.19 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.19 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.19 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.2 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.2 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.2 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.21 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.21 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.21 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.22 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.22 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.22 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.23 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.23 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.23 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.24 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.24 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.24 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.25 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.25 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.25 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.26 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.26 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.26 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.27 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.27 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.27 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.28 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.28 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.28 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.29 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.29 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.29 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.3 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.3 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.3 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.31 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.31 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.31 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.32 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.32 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.32 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.33 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.33 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.33 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.34 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.34 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.34 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.35 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.35 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.35 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.36 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.36 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.36 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.37 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.37 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.37 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.38 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.38 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.38 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.39 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.39 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.39 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.4 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.4 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.4 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.41 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.41 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.41 min\_child = 47 max\_depth = 6 pval = 0.1996269

gamma = 0.42 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.42 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.43 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.43 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.44 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.44 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.45 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.45 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.46 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.46 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.47 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.47 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.48 min\_child = 19 max\_depth = 6 pval = 0.4619108

gamma = 0.48 min\_child = 20 max\_depth = 6 pval = 0.4287897

gamma = 0.49 min\_child = 19 max\_depth = 6 pval = 0.4455275

gamma = 0.49 min\_child = 20 max\_depth = 6 pval = 0.4116896

gamma = 0.5 min\_child = 19 max\_depth = 6 pval = 0.4455275

gamma = 0.5 min\_child = 20 max\_depth = 6 pval = 0.4116896

gamma = 0.51 min\_child = 19 max\_depth = 6 pval = 0.4455275

gamma = 0.51 min\_child = 20 max\_depth = 6 pval = 0.4116896

gamma = 0.52 min\_child = 19 max\_depth = 6 pval = 0.4455275

gamma = 0.52 min\_child = 20 max\_depth = 6 pval = 0.4116896

gamma = 0.53 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.53 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.54 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.54 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.55 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.55 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.56 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.56 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.57 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.57 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.58 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.58 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.59 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.59 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.6 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.6 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.61 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.61 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.62 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.62 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.63 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.63 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.64 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.64 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.65 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.65 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.66 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.66 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.67 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.67 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.68 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.68 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.69 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.69 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.7 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.7 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.71 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.71 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.72 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.72 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.73 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.73 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.74 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.74 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.75 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.75 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.76 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.76 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.77 min\_child = 19 max\_depth = 6 pval = 0.4287897

gamma = 0.77 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.78 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.78 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.79 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.79 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.8 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.8 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.81 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.81 min\_child = 20 max\_depth = 6 pval = 0.3942204

gamma = 0.82 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.82 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.83 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.83 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.84 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.84 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.85 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.85 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.86 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.86 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.87 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.87 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.88 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.88 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.89 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.89 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.9 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.9 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.91 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.91 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.92 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.92 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.93 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.93 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.94 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.94 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.95 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.95 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.96 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.96 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.97 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.97 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.98 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.98 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 0.99 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 0.99 min\_child = 20 max\_depth = 6 pval = 0.3763769

gamma = 1 min\_child = 19 max\_depth = 6 pval = 0.4116896

gamma = 1 min\_child = 20 max\_depth = 6 pval = 0.3763769

|  |
| --- |
| importance\_matrix  Feature Gain Cover Frequency  1: rs4917199 0.08812564 0.03025065 0.05263158  2: rs2072468 0.06883303 0.09579372 0.05263158 LMBR1 Cancer  3: rs2728944 0.06091750 0.10011524 0.05263158 SRGAP3 Cancer  4: rs2247746 0.06078531 0.10702967 0.05263158 RBBP9 Cancer  5: rs5973794 0.05919076 0.11970614 0.05263158  6: rs16938969 0.05877589 0.07029674 0.05263158 PI15 Cancer  7: rs2878682 0.05812120 0.03759723 0.05263158 CFAP47 Cancer  8: rs4978718 0.05334542 0.05517142 0.05263158  9: rs1548936 0.05286775 0.06986459 0.05263158 LRP2 Cancer  10: rs17168486 0.05284556 0.01959090 0.05263158 DGKB Cancer  11: rs10051918 0.05218798 0.04465572 0.05263158 TENM2 Cancer  12: rs231515 0.05073868 0.11970614 0.05263158 MPP2 Cancer  13: rs6775178 0.04920184 0.01512532 0.05263158  14: rs17438524 0.04766272 0.02520887 0.05263158  15: rs1667354 0.04724909 0.02549697 0.05263158 ZNF568 Cancer  16: rs300657 0.04170577 0.02203976 0.05263158  17: rs9887131 0.03680950 0.01411697 0.05263158  18: rs11713084 0.03456481 0.01454912 0.05263158 TMPRSS7 Cancer  19: rs10253917 0.02607155 0.01368482 0.05263158 TAX1BP1 Cancer |
|  |
| |  | | --- | | > | |