Random Forest

Library(randomForest)

Aire <- read.delim("Aire comparisons.txt")

> View(Aire)

> set.seed(71)

> ID <- as.factor(Aire$ID)

Aire.rf <- randomForest (Aire[,3:1277],Aire$ID, ntree=100000, mtry = 500, do.trace = 500, replace=TRUE, proximity = TRUE, importance = TRUE)

print(Aire.rf)

MDSplot(Aire.rf, Aire$Condition, k=2, palette=NULL, pch=20)

varImpPlot(Aire.rf)

Aire.rf.i <- importance(Aire.rf)

write.table(Aire.rf.i, "Aire Analyses Importance.txt", sep="\t")

Aire.impvar <- rownames(Aire.rf.i)[order(Aire.rf.i[, 1], decreasing=TRUE)]

write.table(Aire.impvar, "Aire Important Variables.txt")

Bayes Factor

Library(BayesFactor)

AireBFstack <- read.delim("AireBFstack.txt", header=TRUE)

ID <- as.factor(AireBFstack$ID)

str(AireBFstack)

AireBFstacktest <- ttestBF(x = AireBFstack$Data[AireBFstack$ID==1], y = AireBFstack$Data[AireBFstack$ID==2], paired=FALSE)

print(AireBFstacktest)

AireBFstackchains = posterior(AireBFstacktest, iterations = 1000)

summary(AireBFstackchains) ###cop to text

bfInterval = ttestBF(x = AireBFstack$Data[AireBFstack$ID==1], nullInterval=c(-Inf,0))

print(bfInterval)

write.table(AireBFstacktest, "AireBFstacktest.txt", sep="\t")

write.table(AireBFstackchains, "AireBFstackchains.txt", sep="\t") ###Copy and paste Summary into txt

write.table(bfInterval, "Airebfinterval.txt", sep="\t")

Save figures from posterior (aka chains) and interval tests

plot(AireBFstackchains[,1:4]) (or eg 1:2, 1:3 etc, I plot 1:2 and 3:4)

###Can also use bf = ttestBF(formula = Data ~ ID, data = AireBFstack) for main analysis###

BESTmcmc (for comparison of two group means)

Library(BEST)

AireBESTmcmc <- read.delim("AireBESTmcmc.txt", header=TRUE)

BESTmcmc <- BESTmcmc(AireBESTmcmc$V1, AireBESTmcmc$V2)

BESTmcmc ###Displays the model info, must be copied to txt)

summary(BESTmcmc) ###Displays summary of results, must be copied to txt, for more info use summary call below)

plot(BESTmcmc) ###Plots difference of means etc, must be copied by bitmap to paint

plot(BESTmcmc, “sd”) ###Plots difference of SDs

plotAll(BESTmcmc, credMass=0.80, ROPEm=c(-0.00002,0.00002), ROPEeff=c(-0.2,0.2), compValm=0.5) ###Plots everything needed, must be done on big screen, can also use credMass=0.95. Note the ROPE values must match the mean values used, which in the case of frequencies are very low (10^-5)

summary(BESTmcmc, credMass=0.8, ROPEm=c(-0.00002,0.00002), ROPEsd=c(-0.00002,0.00002), ROPEeff=c(-0.2,0.2)) ###Can also use credMass=0.95, must save in txt file. Note the ROPE values must match the mean values used, which in the case of frequencies are very low (10^-5)

pairs(BESTmcmc) ###gives an XY matrix plot of mu 1 v sd 1, sd 2, and so on

head(BESTmcmc$mu1) ###mu1, mu2, and so on, gives the values only for mu1, can be printed or saved into file form

hdi(BESTmcmc, credMass = 0.80) ###highest density interval

plotAreaInROPE(BESTmcmc, credMass = 0.8, compVal = 0, maxROPEradius = 3) ### how much area in the rope

MCMCpack (for MCMC logistic regression)

library(MCMCpack)

AireMCMCpack <- read.delim("AireBESTmcmc.txt", header=TRUE)

library(plyr) ###for recoding variable)

AireMCMCpack$V1 <- mapvalues(AireMCMCpack$V1, from = c("1", "2"), to = c("0", "1")) ###to recode into binary 0 and 1, the response variable must be 0 and 1 for MCMClogit to run

Aire.posterior <- MCMClogit(V1~V2, data=AireMCMCpack)

plot(Aire.posterior)

summary(Aire.posterior)

Chyi’s Deseq2 (not used with TCR data but could, requires counts not frequencies)

dds <- DESeqDataSetFromMatrix(countData = table, colData = colData, design = ~condition)

The countData = table should be a table of only sequence counts. rownames = TCR

colData should be a data frame of column names.  Can have multiple columns (i.e. location (thymus vs spleen),

design = ~ columnname1

(or columnname1 + name2) for multifactorial design.

Data Structure

randomForest

Condition ID TCR1 TCR2 …TCRz

ATGhet 1 A Afreq1 Afreq2 …Afreqz

ATGhet2 A “” “” “”

ATGKO1 B Bfreq1 Bfreq2 …Bfreqz

ATGKO2 B “” “” “”

Bayes Factor

ID Data (mean frequencies)

0 TCR1

0 TCR2

0 TCR3

1 TCR1

1 TCR2

1 TCR3

BESTmcmc

TCRname Airehet (mean freq) AireKO (mean freq)

TCR1 mean freq1 mean freq1

TCR2 mean freq2 mean freq2

TCR3 mean freq3 mean freq3

MCMCpack (for function MCMClogit)

Use also randomForest, but with ID as binary (0 and 1) and additional coded variables (for determination of fixed and random effects).