Genomic-based BLUP (GBLUP)

HaipengYu

Overview

Predict genomic estimated breeding values (GEBV) with GBLUP using phenotypes and genomic data.

Mice Data

```
# install.packages("BGLR")
library(BGLR)
data(mice)
`?`(mice)
`?`(mice.X) # genotype matrix
`?`(mice.pheno) # phenotypes
```

Step One : Fit ordinary least square (OLS) to estimate fixed effects (b_hat)

```
dat<-mice.pheno[,c("GENDER","Litter","Obesity.BMI")] # dataset
head(dat)
fit<- lm(Obesity.BMI ~ -1 + Litter + factor(GENDER) , data= dat)

fit2<- lm(Obesity.BMI ~ Litter + factor(GENDER) , data= dat)

fit2$coefficients
fit$coefficients

# fit OLS
summary(fit)
b_hat<- fit$coefficients
```

Step Two : Predict genomic estimated breed value (u_hatG)

Construct G matrix VanRaden (2008) with R-function of computeG. Two arguments need to be given within computeG function: 1) genotype matrix of W; 2) minor allele frequency threshold (0.05 is used in this example).

```
computeG <- function(W, maf) {
  p <- colMeans(W)/2
  maf2 <- pmin(p, 1 - p)
  index <- which(maf2 < maf)
  W2 <- W[, -index]
  p2 <- p[-index]
  Wc <- scale(W2, center = TRUE, scale = FALSE)
  G <- tcrossprod(Wc)/(2 * sum(p2 * (1 - p2)))
  return(G)
}
```

```r
G <- computeG(W=mice.X , maf = 0.05)
dim(G) # 1814 x 1814
```

Phenotype Matrix y

```r
y<-dat$Obesity.BMI
```

Incidence Matrix X

```r
X<-model.matrix(~ -1 + dat$Litter + dat$GENDER)
dim(X)
head(X)
```

Incidence Matrix Z

```r
Z <- diag(nrow(G))
dim(Z)
diag(Z)
```

Incidence Matrix I

```r
I <- diag(nrow(G))
dim(I)
diag(I)
```

Assign values for two variance components (In practice, they need to be estimated from data)

```r
sigma2G<- 0.4
sigma2e<- 0.6
```

Inverse of V_y matrix

```r
V <- Z %*% G %*% t(Z) * sigma2G + I * sigma2e
dim(V)
Vinv <- solve(V)
dim(Vinv)
```

Compute GEBV

```r
u_hatG <- sigma2G * G %*% t(Z) %*% Vinv %*% (matrix(y) - matrix(X %*% fit$coefficients))
head(u_hatG)
dim(u_hatG)
```

Mixed Model Equation (MME)

We can estimate fixed effects and predict GEBV simultaneously by using MME. The function of computeMME requires 6 arguments of phenotype matrix (y), incidence matrix for fixed effects (X), incidence matrix for random effects (Z), genomic relationship matrix (G), variance components sigma2G and sigma2e.

```
computeMME <- function(y, X, Z, G, sigma2G, sigma2e) {

  X<-model.matrix(~ -1 + dat$Litter + factor(dat$GENDER))
  Z <- diag(nrow(G))
  I <- diag(nrow(G))
  lambdaG <- sigma2e/sigma2G

  XtX <- crossprod(X)
  XtZ <- crossprod(X, Z)
  ZtX <- crossprod(Z, X)
  ZtZG <- crossprod(Z) + solve(G) * lambdaG
  Xty <- crossprod(X, y)
  Zty <- crossprod(Z, y)
  LHS1 <- cbind(XtX, XtZ)
  LHS2 <- cbind(ZtX, ZtZG)
  LHSG <- rbind(LHS1, LHS2)
  RHS <- rbind(Xty, Zty)
  sol.G <- solve(LHSG) %*% RHS
  return(sol.G)

}

sol.mme <- computeMME(y = y, X = X, Z = Z, G = G, sigma2G = 0.4, sigma2e = 0.6)
head(sol.mme)
```