Genome-wide Selection Tutorial

Requiring package and loading data

```
require(BGLR)
```

```
## Loading required package: BGLR
```

```
data(wheat)
```

Defining phenotypic and genotypic matrices

```
Y<- wheat.Y[,4]    # Phenotypes
X<- wheat.X        # Genotypes
```

Dimensions of the matrices

```
dim(X)
```

```
## [1]  599 1279
```

```
length(Y)
```

```
## [1] 599
```

Scale and center the genotypic matrix

```
X<-scale(wheat.X, center = TRUE, scale = TRUE) # Molecular data
```

GS analysis - Full model

Setting the linear predictor (factors to be used in the analysis)

```
ETA<-list(X=list(X=X,model="BRR")) # To run BayesB -> model="BayesB"
```

Running the full model

```
fm<-BGLR(y=Y,ETA=ETA,nIter=10000,burnIn=2000, thin = 5, verbose = FALSE)
```

Computing accuracy - Correlation between the breeding values (BV) and the phenotypic values

```
(accur1<-cor(fm$yHat, Y))
```

```
## [1] 0.7816485
```

GS validation modelm - FiveFold Cross-Validation (manually)

Defining the five groups

```
tst_5fold<-sample(1:5,size=599,replace=TRUE)
```

Setting the linear predictor and creating a new Y

```
ETA<-list(X=list(X=X,model="BRR"))
Y2<-Y
```

Removing the phenotyic values for each one of the groups and Running the analysis for each set

```
# Group 1
Y2.1<-Y2
Y2.1[tst_5fold==1]<-NA
vm1.1<-BGLR(y=Y2.1,ETA=ETA,nIter=10000,burnIn=2000, thin = 5,verbose = FALSE)

# Group 2
Y2.2<-Y2
Y2.2[tst_5fold==2]<-NA
vm1.2<-BGLR(y=Y2.2,ETA=ETA,nIter=10000,burnIn=2000, thin = 5, verbose=FALSE)

# Group 3
Y2.3<-Y2
Y2.3[tst_5fold==3]<-NA
vm1.3<-BGLR(y=Y2.3,ETA=ETA,nIter=10000,burnIn=2000, thin = 5, verbose = FALSE)

# Group 4
Y2.4<-Y2
Y2.4[tst_5fold==4]<-NA
vm1.4<-BGLR(y=Y2.4,ETA=ETA,nIter=10000,burnIn=2000, thin = 5, verbose = FALSE)

# Group 5
Y2.5<-Y2
Y2.5[tst_5fold==5]<-NA
vm1.5<-BGLR(y=Y2.5,ETA=ETA,nIter=10000,burnIn=2000, thin = 5, verbose = FALSE)
```

Function to get the breeding values from the Validation model

```
data_5fold<-data.frame(Y=Y, i= 1:599, sub=tst_5fold)

result1<-data.frame(yHat1=vm1.1$yHat, yHat2=vm1.2$yHat, yHat3=vm1.3$yHat,yHat4=vm1.4$yHat, yHat5=vm1.5$y

yHat_5fold1 <- data.frame(yHat=apply(data_5fold[,c(2,3)], 1, function(x) result1[x[1], x[2]]))
```

Computing accuracy - Correlation between the breeding values from 5fold cross-validation and the phenotypic values

```
(accur2<-cor(yHat_5fold1$yHat,Y))
```

```
## [1] 0.4685991
```

GS validation modelm - FiveFold Cross-Validation (loop)

Defining loop

```
fivefold_model <- function(i){
  trn.y <- Y
  trn.y[tst_5fold==i] <- NA
  BGLR(y=trn.y, ETA=ETA, nIter = 10000, burnIn = 2000,thin = 5,  verbose=FALSE,saveAt = '5fold_')$yHat
}
```

Running the analysis - Jacknife Fivefold

```
result2 <- data.frame(lapply(1:5, function(i) fivefold_model(i)))
```

Function to get the breeding values from the validation model

```
data_5fold<-data.frame(Y=Y, i= 1:599, sub=tst_5fold)
yHat_5fold2 <- data.frame(yHat=apply(data_5fold[,c(2,3)], 1, function(x) result2[x[1], x[2]]))
```

Correlation between the breeding values from 5fold cross-validation and the phenotypic values

```
(accur3<-cor(yHat_5fold2$yHat,Y))
```

```
## [1] 0.4699082
```

Shrinkage effect

Comparing the range of the predictions and the phenotypic value

```
range(Y)                    # Phenotypes
```

```
## [1] -3.575634  2.792202
```

```
range(fm$yHat)              # Full model
```

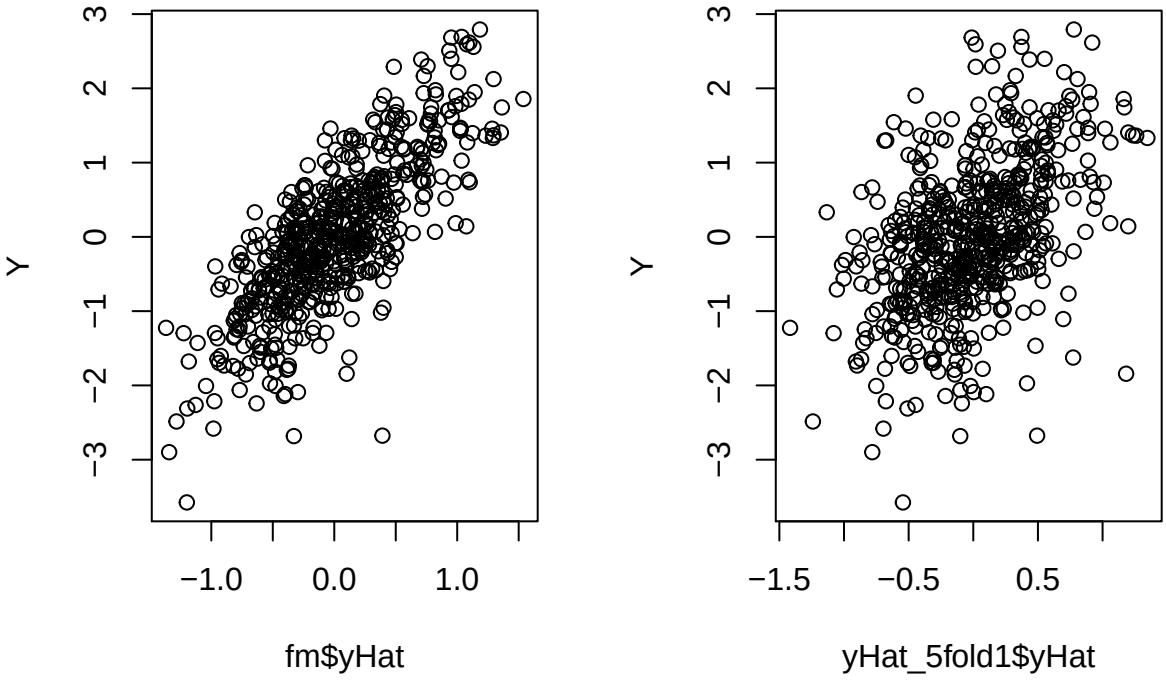
```
## [1] -1.368086  1.538333
```

```
range(yHat_5fold1$yHat)     # Cross-Validation 1
```

```
## [1] -1.417214  1.346772
```

Visualizing the shrinkage effect

```
par(mfrow=c(1,2))
plot(fm$yHat, Y)              # BV from Full model vs Phenotypes
plot(yHat_5fold1$yHat, Y)     # BV 5fold model vs Phenotypes
```



Histograms

```
par(mfrow=c(1,3))
hist(Y,xlim=c(-4,3), ylim=c(0,130),15)       # Phenotypes
```

```r
hist(fm$yHat,xlim=c(-4,3),ylim=c(0,130),15)  # BV from Full model
hist(yHat_5fold1$yHat,xlim=c(-4,3),ylim=c(0,130),15) # BV 5fold model
```