

local_imp_functions

Kelvyn Bladen

2024-11-11

Contents

Data Regression	2
Data Classification	15

Data Regression

```
s = Sys.time()
n <- 500 # increase higher 1000
d = 6
set.seed(123)
sig = diag(d)
df = MASS::mvrnorm(n = n, mu = rep(0, d), Sigma = sig)
y = df %*% 5:0

dfc <- data.frame(df, y)
```

```
nreps = 10
per = mat.or.vec(nreps, d)
gri = mat.or.vec(nreps, d)
pl = vector("list", nreps)
gl = vector("list", nreps)

cl <- makePSOCKcluster(5)
registerDoParallel(cl)
for(q in 1:nreps){
  pg = global_perm_imp_cm(formula = y ~ ., data = dfc,
                           method = "rf", tuneGrid = NULL,
                           nsim = 16, folds = 5, seed = q)
  xg = global_grid_imp_cm(formula = y ~ ., data = dfc,
                           method = "rf", tuneGrid = NULL,
                           nsim = 16, folds = 5, quantile_grid = TRUE, seed = q)

  indp = (((pg$global_imp > 0) - .5) * 2)
  indg = (((xg$global_imp > 0) - .5) * 2)

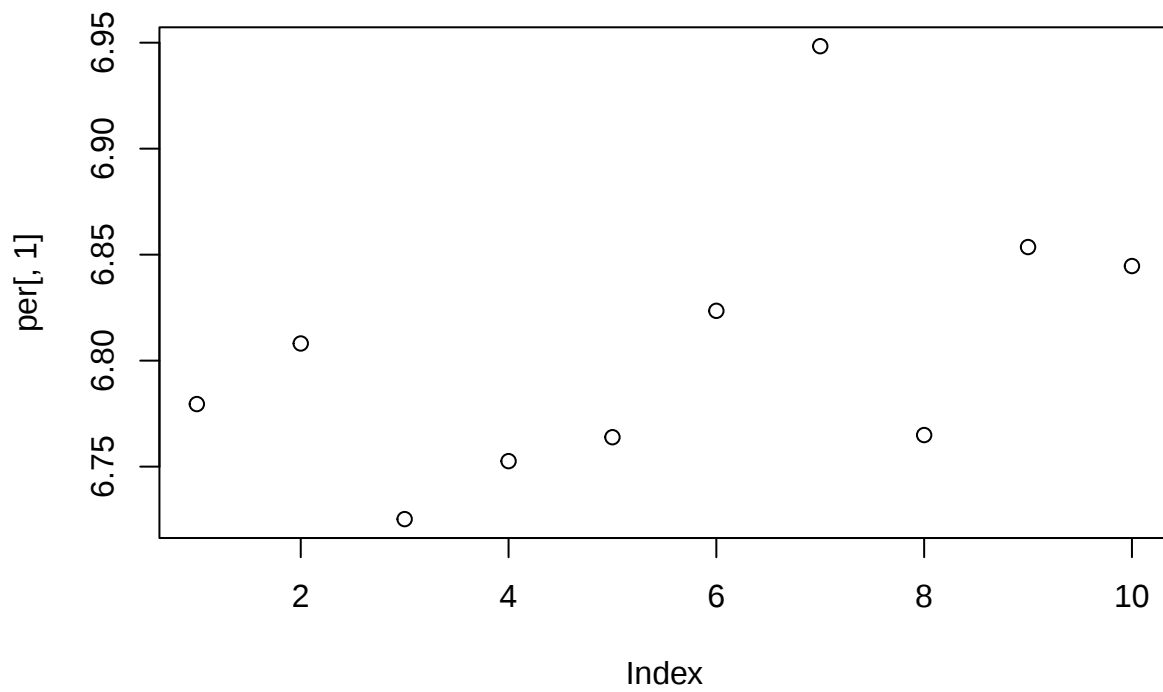
  pl[[q]] = indp * sqrt(indp * pg$global_imp)
  gl[[q]] = indg * sqrt(indg * xg$global_imp)
  per[q,] = colMeans(indp * sqrt(indp * pg$global_imp))
  gri[q,] = colMeans(indg * sqrt(indg * xg$global_imp))
  #colMeans(sqrt(pmax(pg$global_imp, 0)))
}
stopCluster(cl)
unregister_dopar()

summary(per)
```

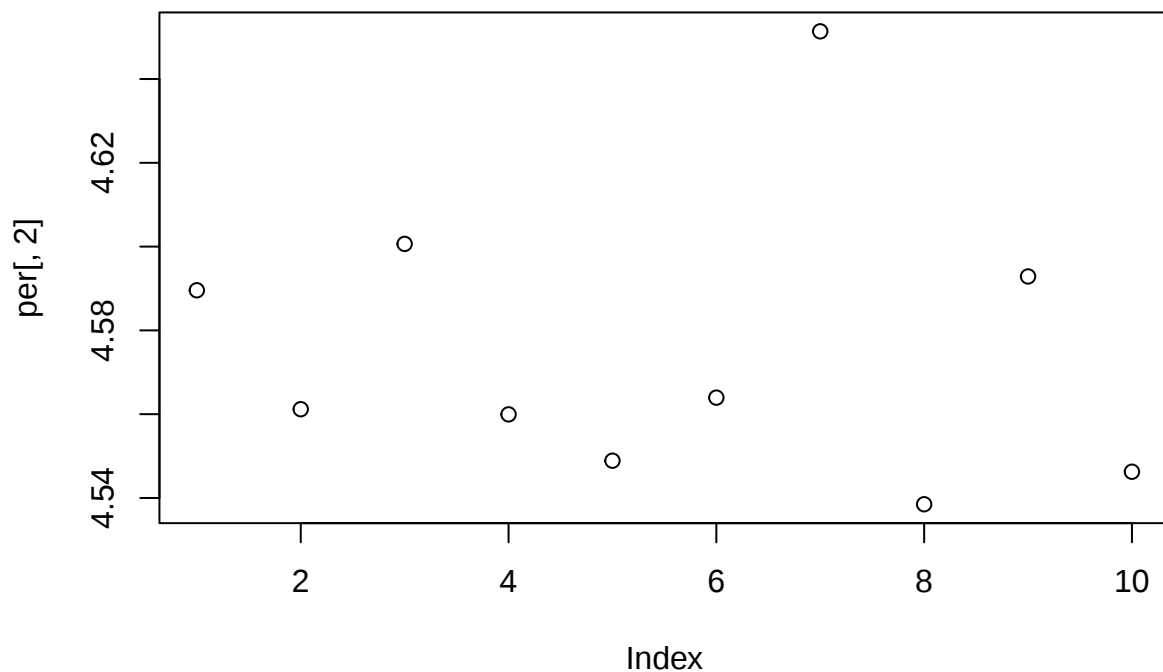
##	V1	V2	V3	V4
##	Min. :6.725	Min. :4.539	Min. :3.459	Min. :1.823
##	1st Qu.:6.764	1st Qu.:4.552	1st Qu.:3.525	1st Qu.:1.875
##	Median :6.794	Median :4.563	Median :3.530	Median :1.901
##	Mean :6.806	Mean :4.575	Mean :3.538	Mean :1.895
##	3rd Qu.:6.839	3rd Qu.:4.592	3rd Qu.:3.575	3rd Qu.:1.918
##	Max. :6.948	Max. :4.651	Max. :3.587	Max. :1.937
##	V5	V6		
##	Min. :0.7772	Min. :-0.18278		
##	1st Qu.:0.8839	1st Qu.: 0.02729		

```
## Median :0.8922   Median : 0.21040  
## Mean   :0.8812   Mean   : 0.15282  
## 3rd Qu.:0.9043   3rd Qu.: 0.30947  
## Max.   :0.9101   Max.    : 0.36461
```

```
plot(per[,1])
```



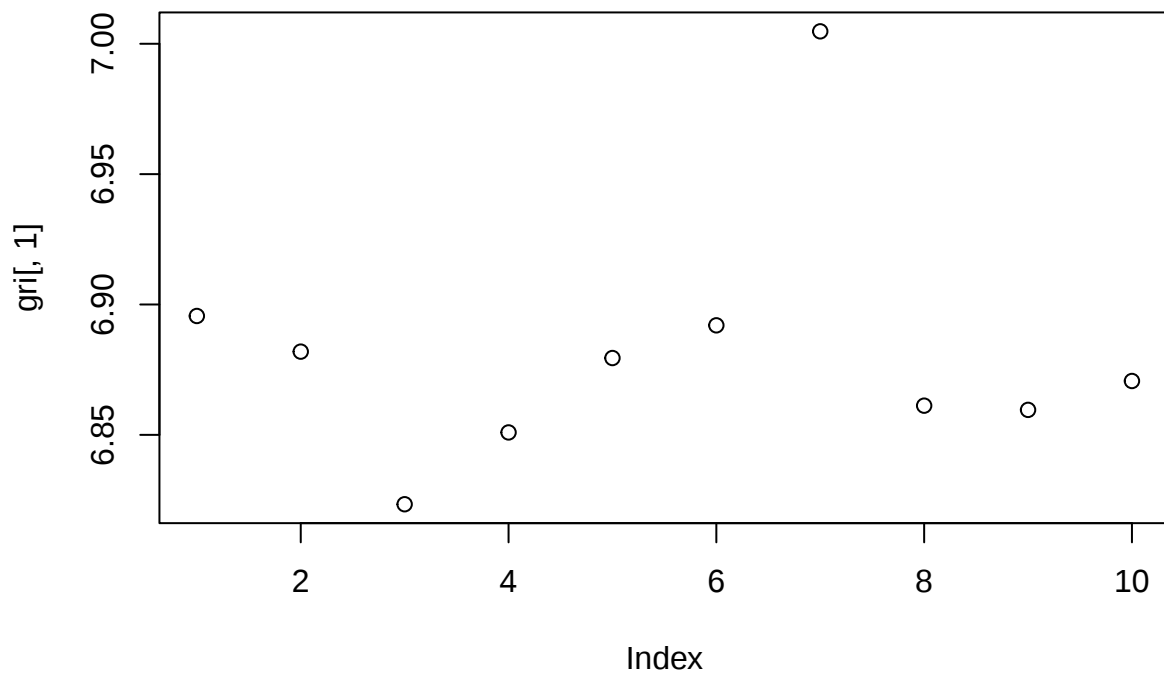
```
plot(per[,2])
```



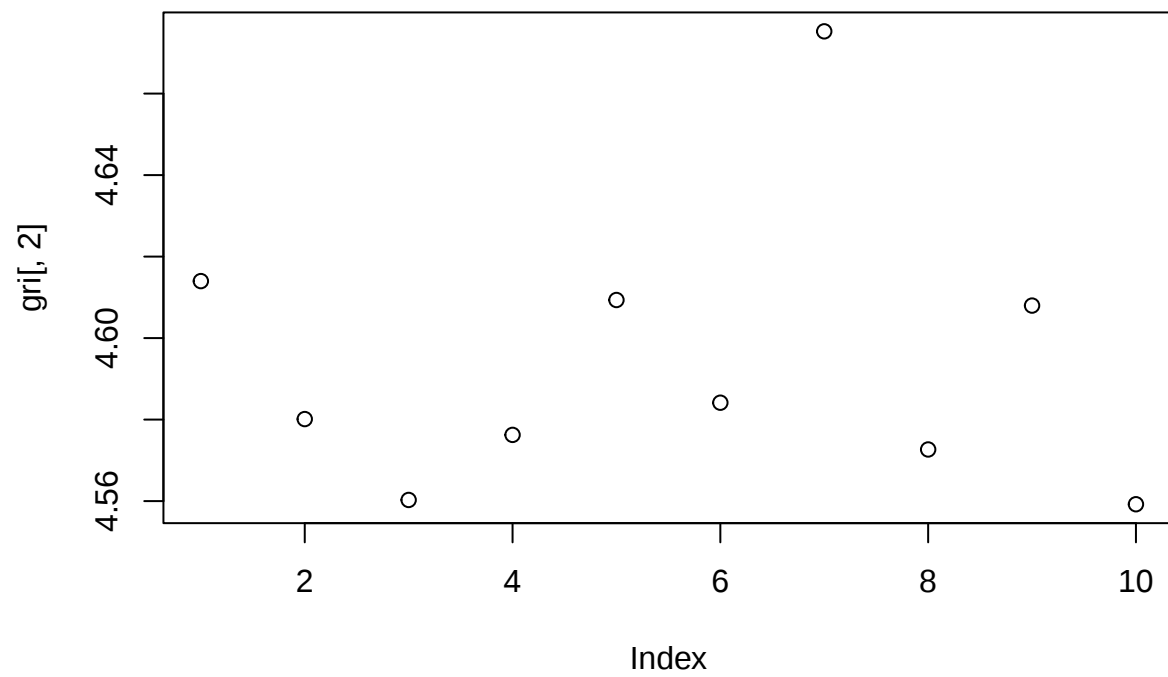
```
summary(gri)
```

```
##           V1           V2           V3           V4
##  Min.    :6.823  Min.    :4.559  Min.    :3.506  Min.    :1.853
## 1st Qu.:6.860  1st Qu.:4.574  1st Qu.:3.539  1st Qu.:1.905
## Median :6.875  Median :4.582  Median :3.551  Median :1.923
## Mean   :6.882  Mean   :4.594  Mean   :3.553  Mean   :1.916
## 3rd Qu.:6.889  3rd Qu.:4.609  3rd Qu.:3.569  3rd Qu.:1.936
## Max.    :7.005  Max.    :4.675  Max.    :3.596  Max.    :1.955
##           V5           V6
##  Min.    :0.7997  Min.    : -0.11829
## 1st Qu.:0.9004  1st Qu.: 0.01773
## Median :0.9088  Median : 0.20126
## Mean   :0.8990  Mean   : 0.14734
## 3rd Qu.:0.9186  3rd Qu.: 0.28985
## Max.    :0.9283  Max.    : 0.33451
```

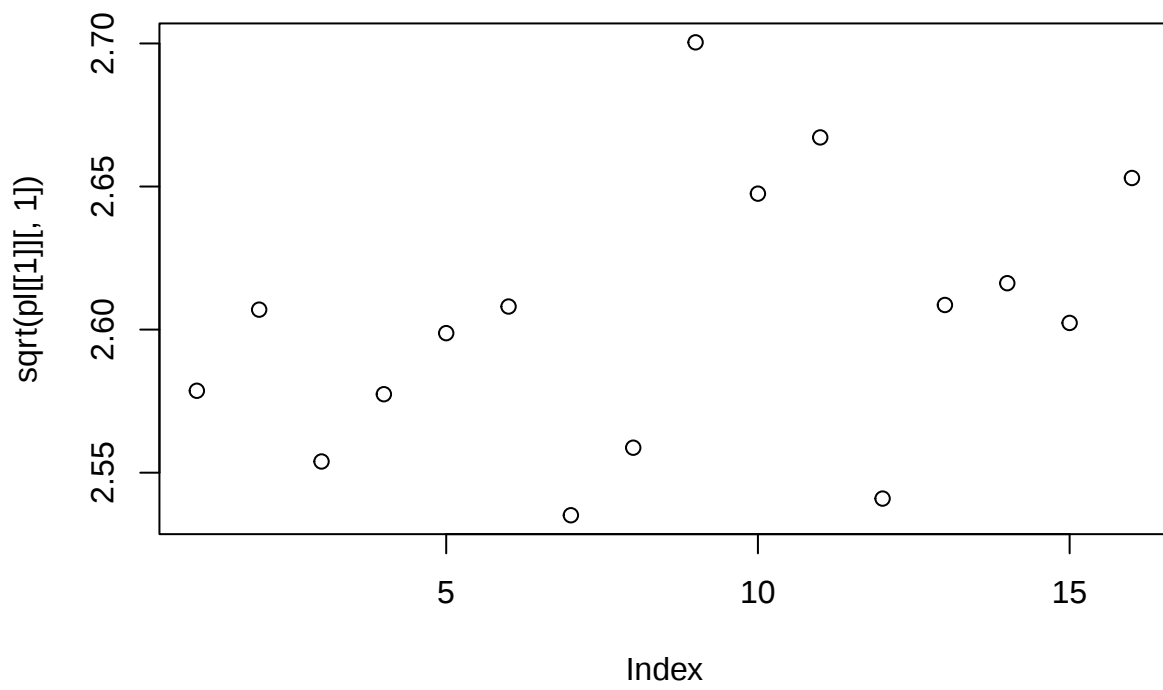
```
plot(gri[,1])
```



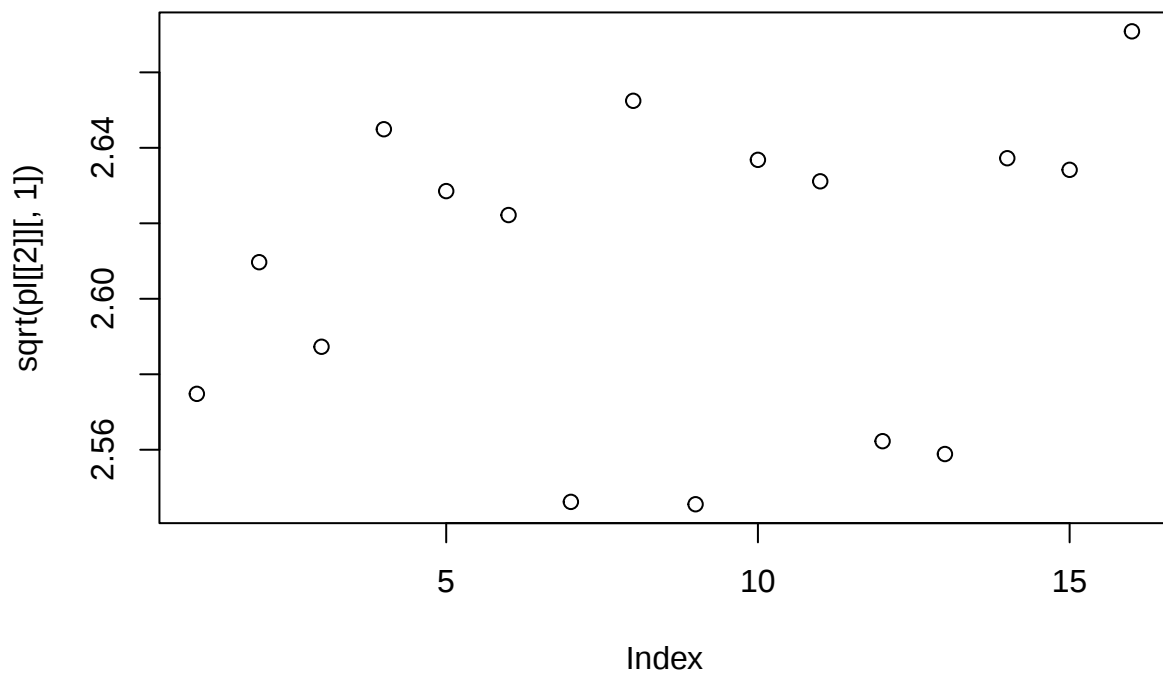
```
plot(gri[,2])
```



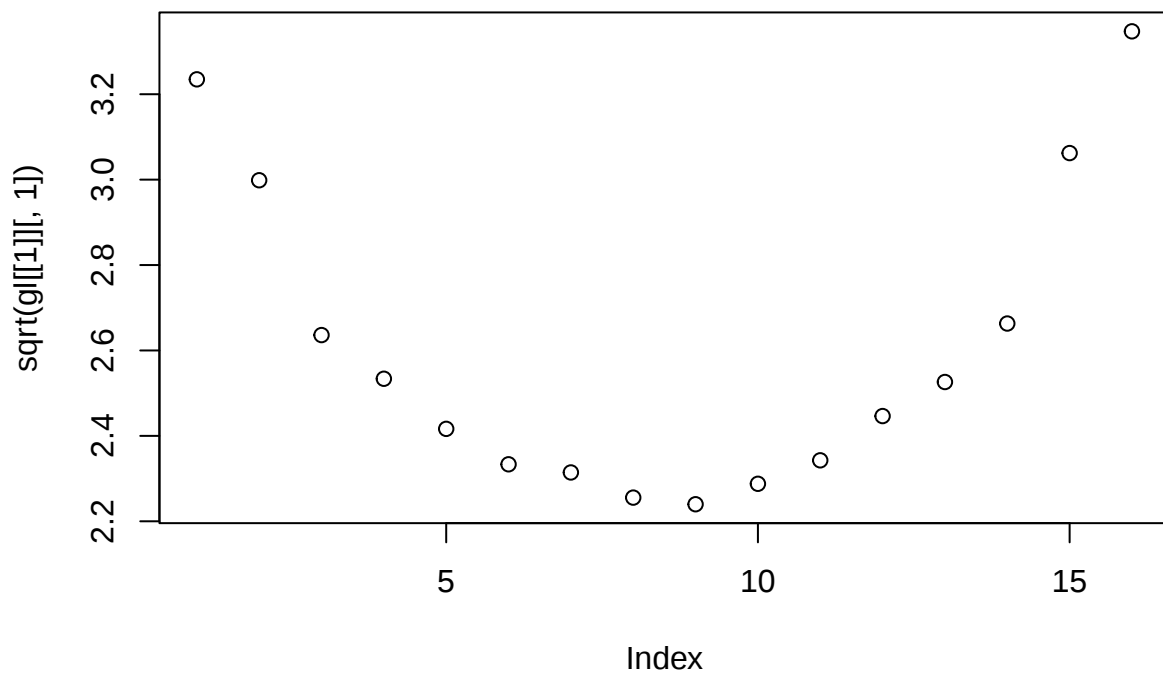
```
plot(sqrt(p1[[1]][,1]))
```



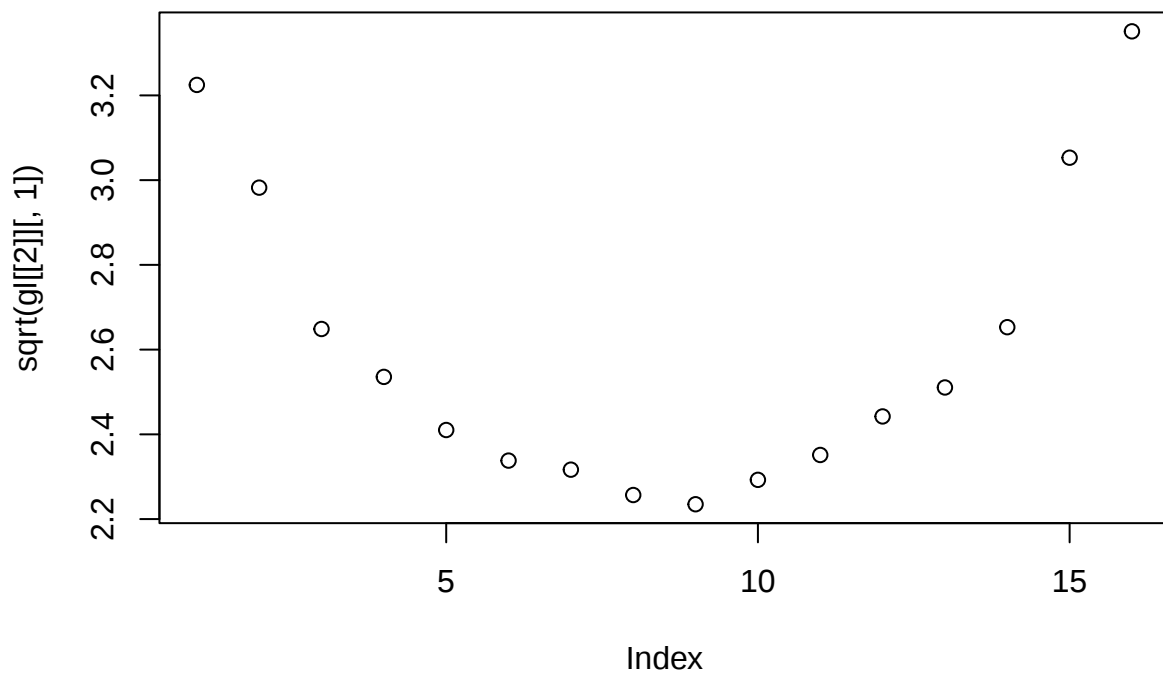
```
plot(sqrt(p1[[2]][, 1]))
```



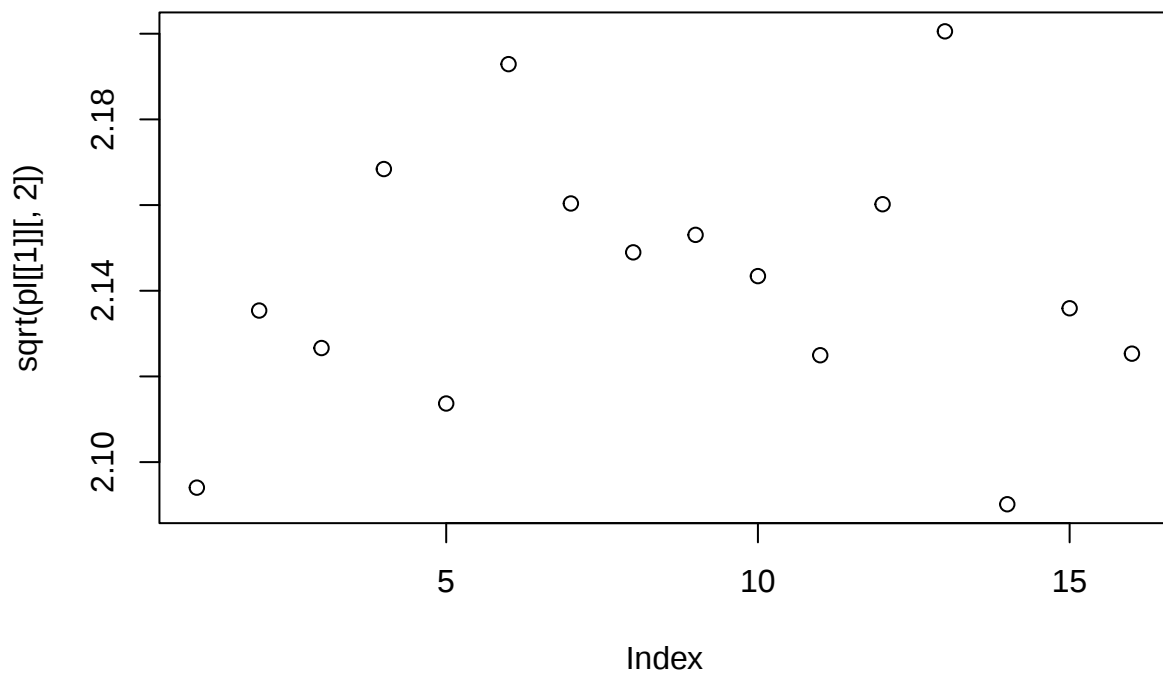
```
plot(sqrt(g1[[1]][,1]))
```



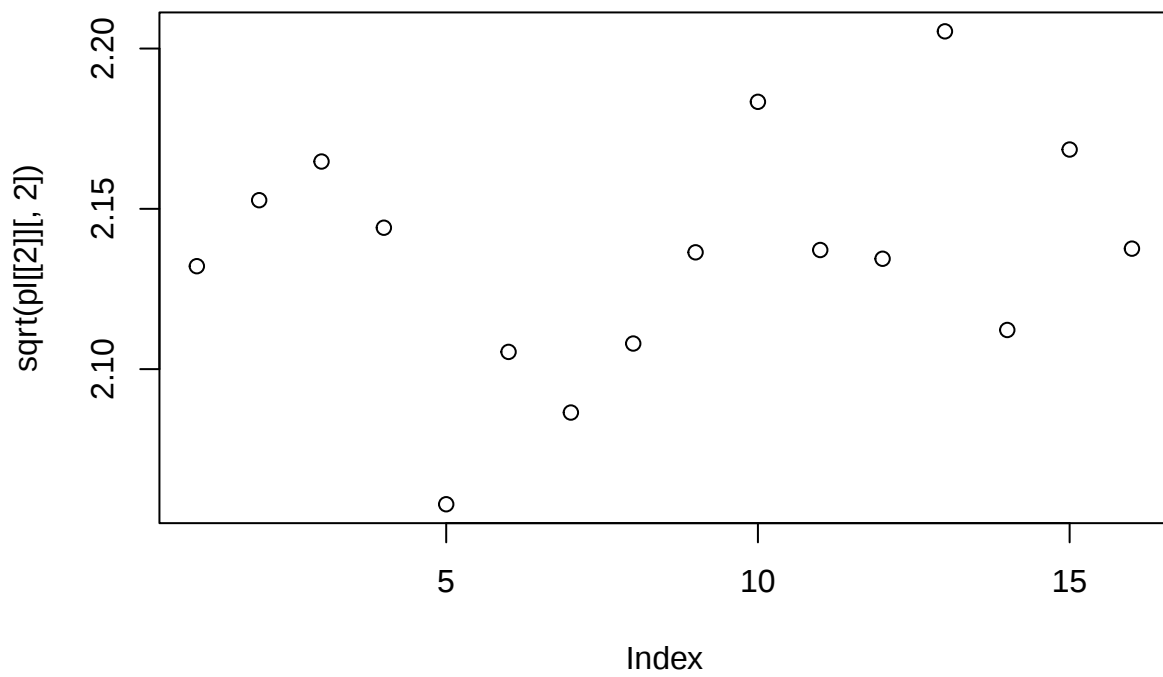
```
plot(sqrt(g1[[2]][, 1]))
```



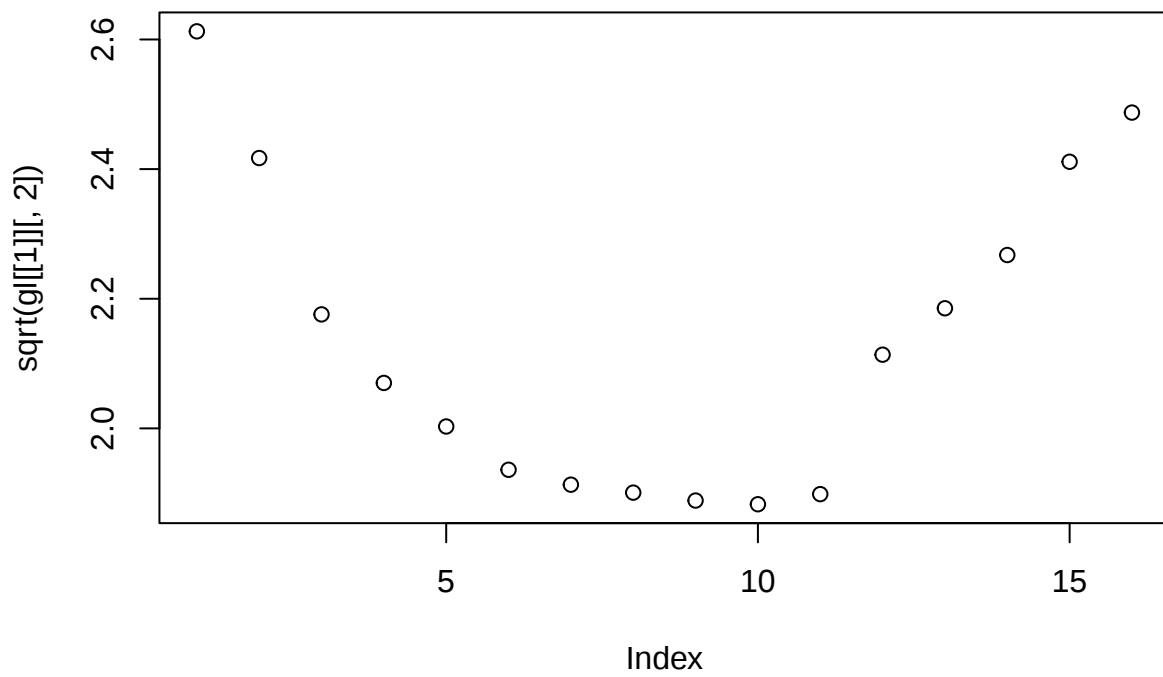
```
plot(sqrt(pl[[1]][,2]))
```



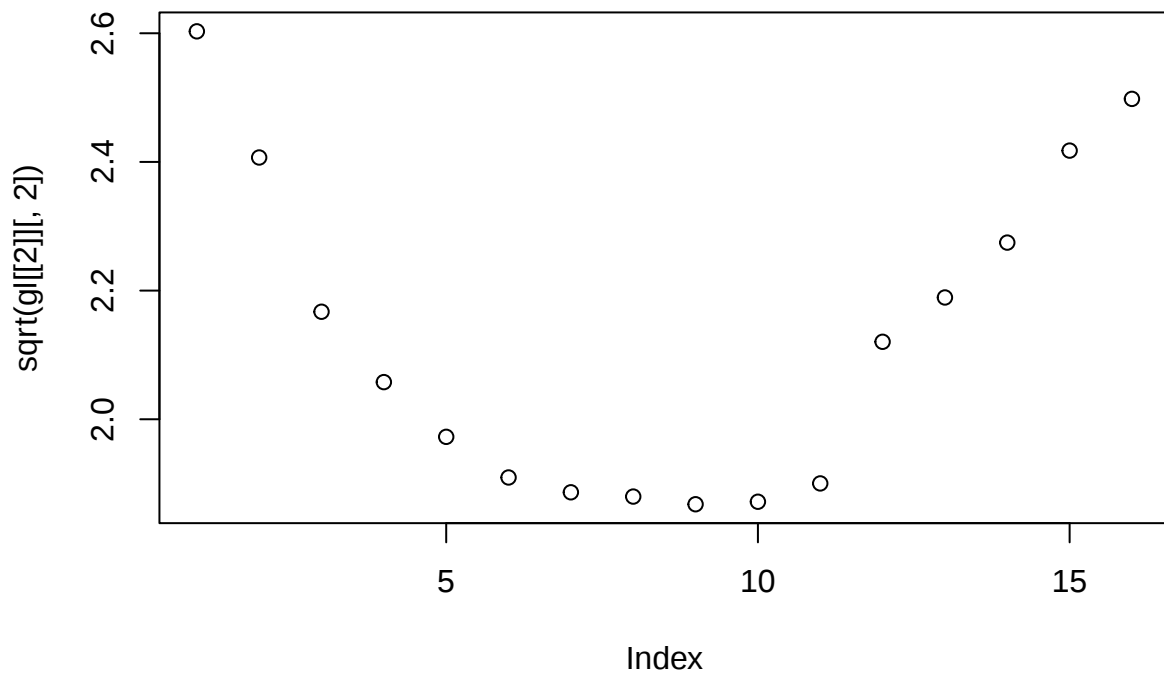
```
plot(sqrt(p1[[2]][, 2]))
```



```
plot(sqrt(g1[[1]][,2]))
```



```
plot(sqrt(g1[[2]][, 2]))
```



Data Classification

```
n <- 500 # increase higher 1000
set.seed(12345)

v1 = runif(n, -1, 1)
v2 = runif(n, -1, 1)
v3 = runif(n, -1, 1)
v4 = runif(n, -1, 1)

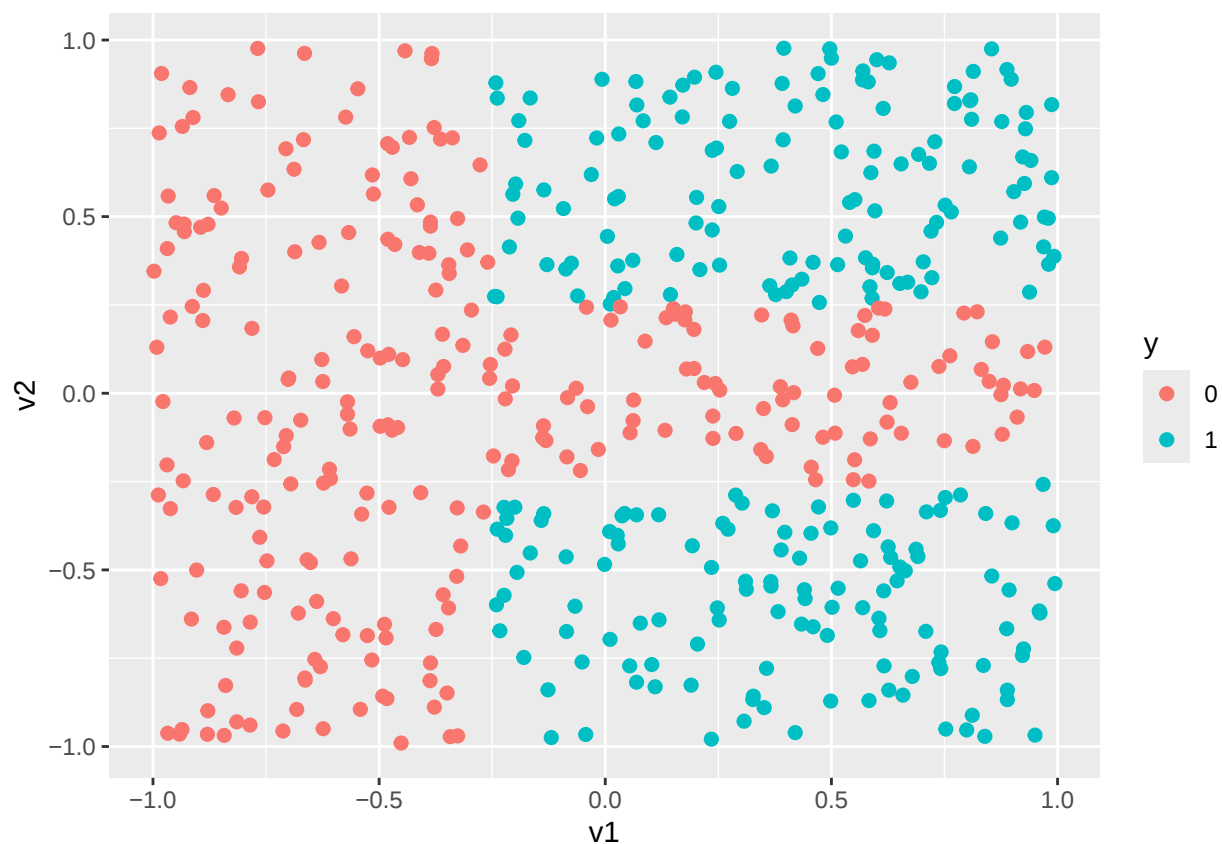
y = ifelse(v1 + rnorm(n, sd = 0.00) < -1/4, 0,
           ifelse(abs(v2) + rnorm(n, sd = 0.00) > 1/4, 1, 0))
y = as.factor(y)

dfc <- data.frame(v1, v2, v3, v4, y)

table(y)
```

```
## y
##  0  1
## 251 249
```

```
ggplot(dfc, aes(v1, v2, colour = y)) + geom_point(size = 2)
```



```

nreps = 10
per = mat.or.vec(nreps, 4)
gri = mat.or.vec(nreps, 4)
pl = vector("list", nreps)
gl = vector("list", nreps)

cl <- makePSOCKcluster(5)
registerDoParallel(cl)
for(q in 1:nreps){
  pg = global_perm_imp_cm(formula = y ~ ., data = dfc,
                           method = "rf", tuneGrid = NULL,
                           nsim = 16, folds = 5, seed = q)
  xg = global_grid_imp_cm(formula = y ~ ., data = dfc,
                           method = "rf", tuneGrid = NULL,
                           nsim = 16, folds = 5, quantile_grid = TRUE, seed = q)

  indp = (((pg$global_imp > 0) - .5) * 2)
  indg = (((xg$global_imp > 0) - .5) * 2)

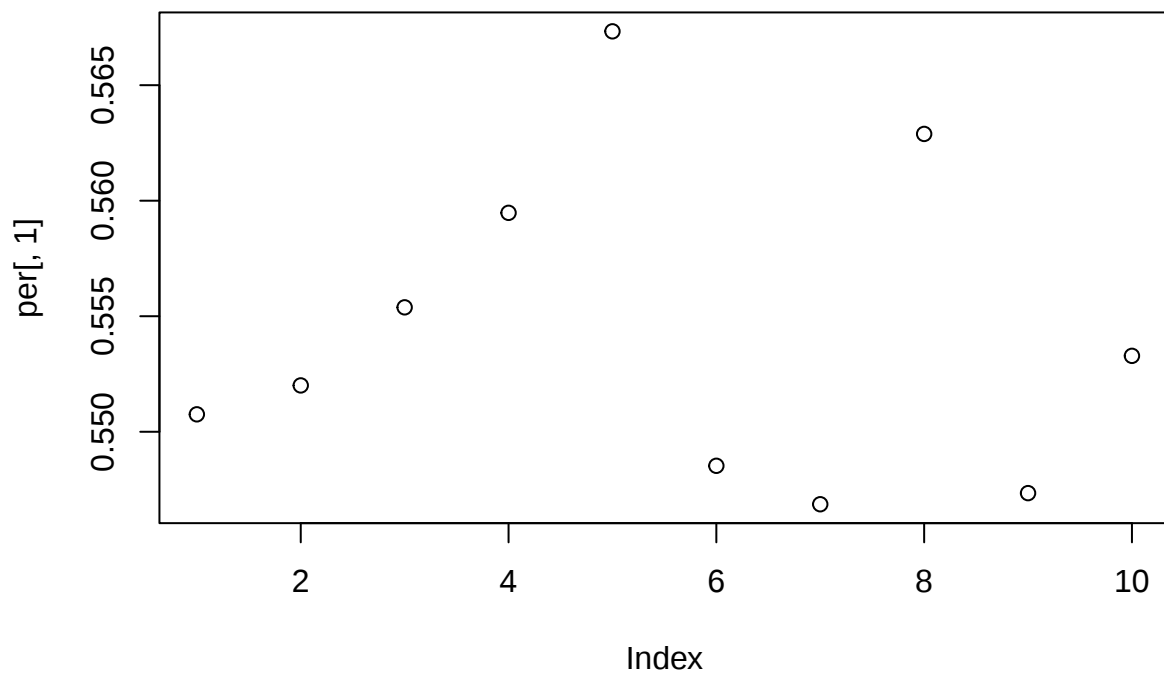
  pl[[q]] = indp * sqrt(indp * pg$global_imp)
  gl[[q]] = indg * sqrt(indg * xg$global_imp)
  per[q,] = colMeans(indp * sqrt(indp * pg$global_imp))
  gri[q,] = colMeans(indg * sqrt(indg * xg$global_imp))
  #colMeans(sqrt(pmax(pg$global_imp, 0)))
}
stopCluster(cl)
unregister_dopar()

summary(per)

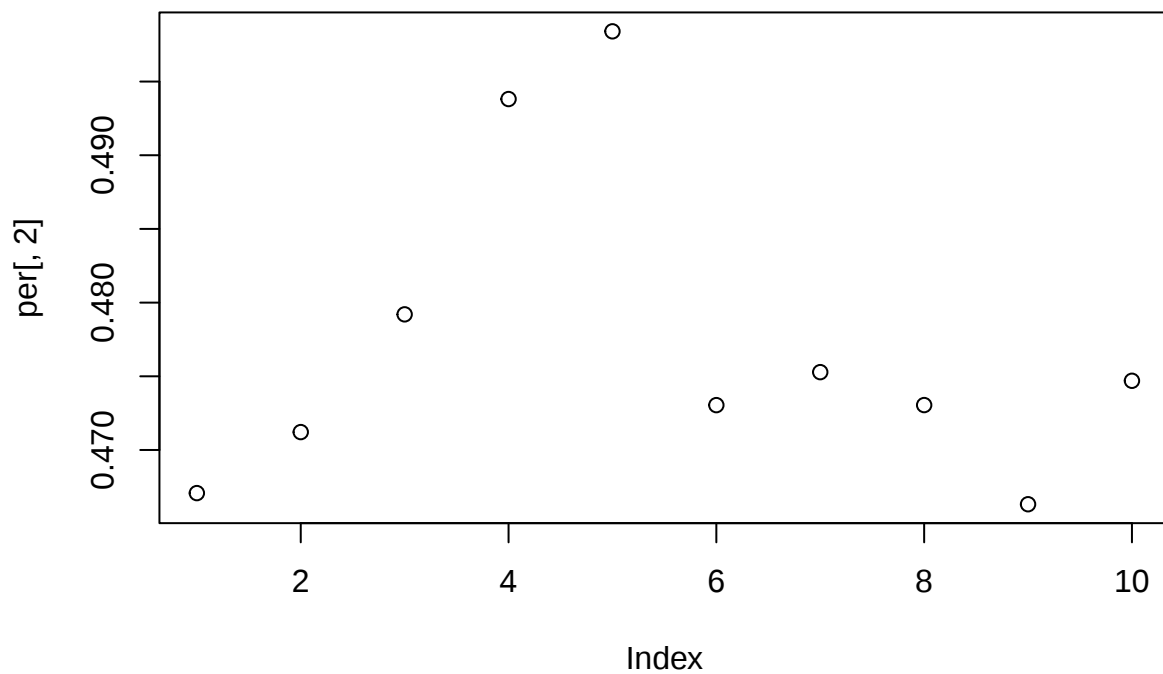
```

##	V1	V2	V3	V4
## Min.	:0.5469	Min. :0.4663	Min. : -0.0282783	Min. : -0.026017
## 1st Qu.	:0.5491	1st Qu.:0.4717	1st Qu.: -0.0082056	1st Qu.: -0.005691
## Median	:0.5526	Median :0.4739	Median : -0.0019098	Median : 0.012232
## Mean	:0.5544	Mean :0.4772	Mean : -0.0006032	Mean : 0.003758
## 3rd Qu.	:0.5585	3rd Qu.:0.4782	3rd Qu.: 0.0121301	3rd Qu.: 0.013606
## Max.	:0.5673	Max. :0.4984	Max. : 0.0146455	Max. : 0.015734

```
plot(per[,1])
```



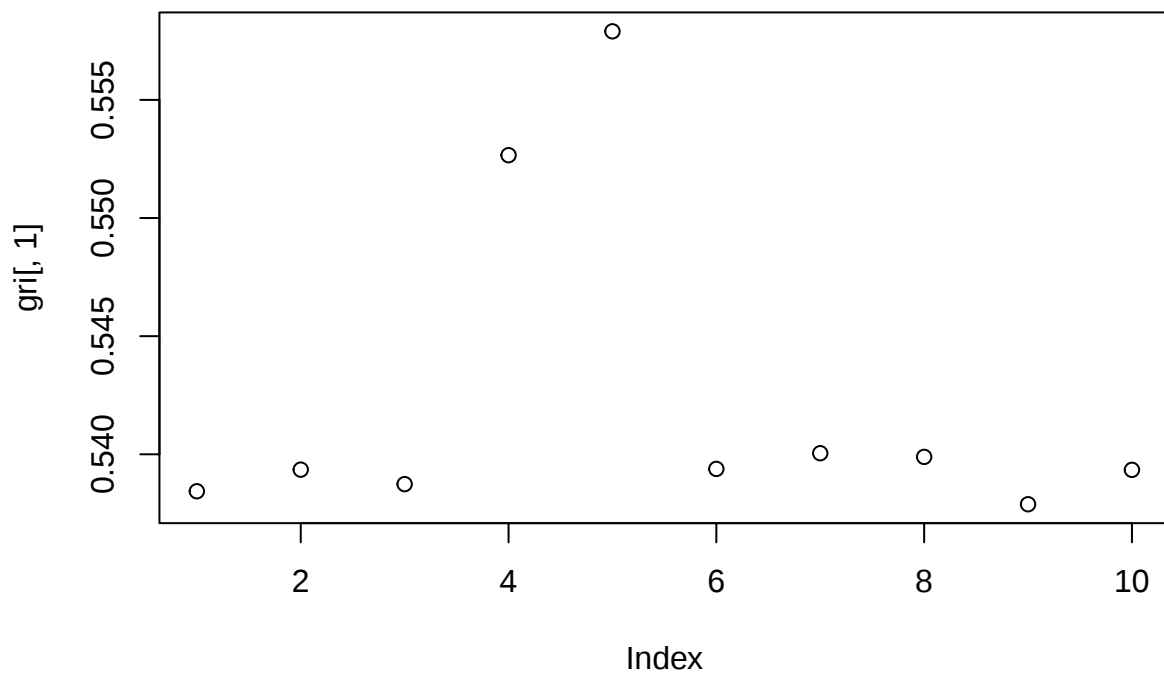
```
plot(per[,2])
```



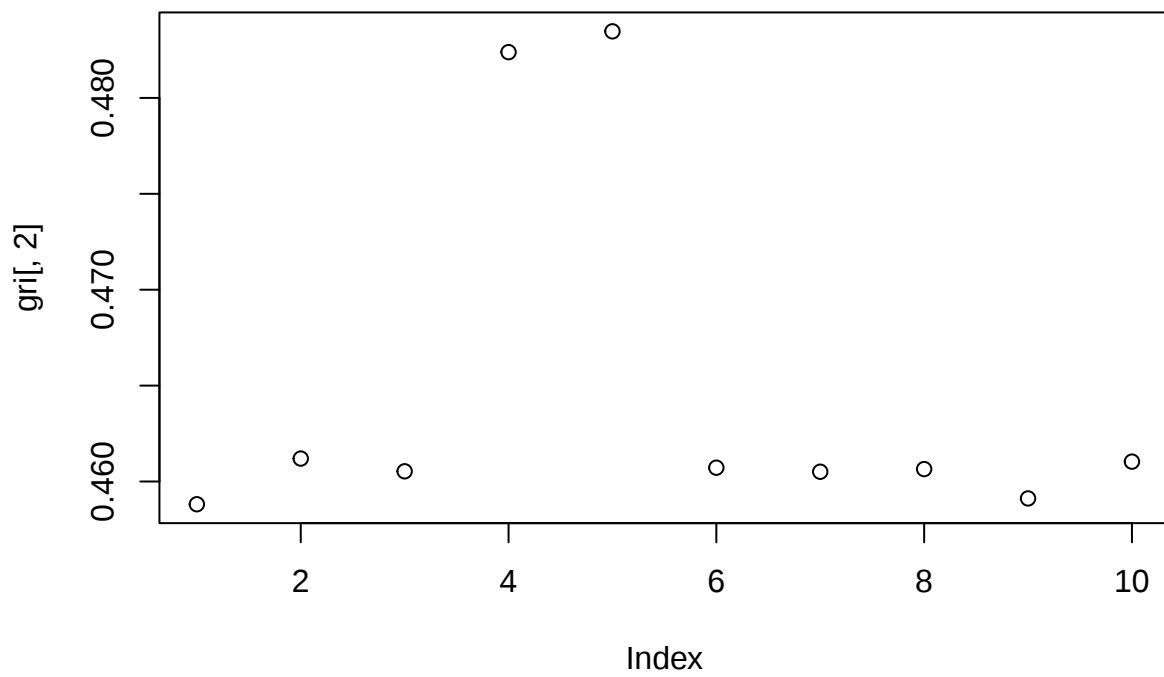
```
summary(gri)
```

##	V1	V2	V3	V4
##	Min. :0.5379	Min. :0.4588	Min. : -0.017991	Min. : -1.843e-02
##	1st Qu.:0.5389	1st Qu.:0.4605	1st Qu.: -0.006226	1st Qu.: -6.296e-03
##	Median :0.5394	Median :0.4607	Median : -0.002641	Median : -2.330e-07
##	Mean :0.5424	Mean :0.4648	Mean : -0.002322	Mean : -1.959e-03
##	3rd Qu.:0.5400	3rd Qu.:0.4612	3rd Qu.: 0.002651	3rd Qu.: 1.492e-03
##	Max. :0.5579	Max. :0.4835	Max. : 0.013572	Max. : 1.271e-02

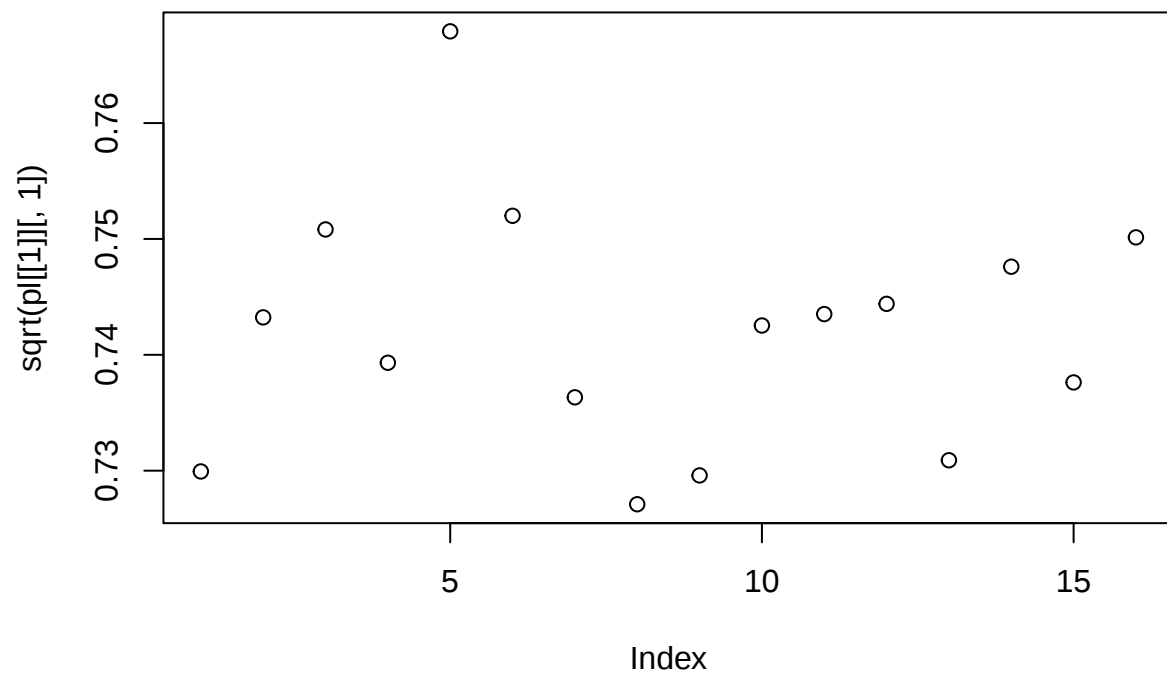
```
plot(gri[,1])
```



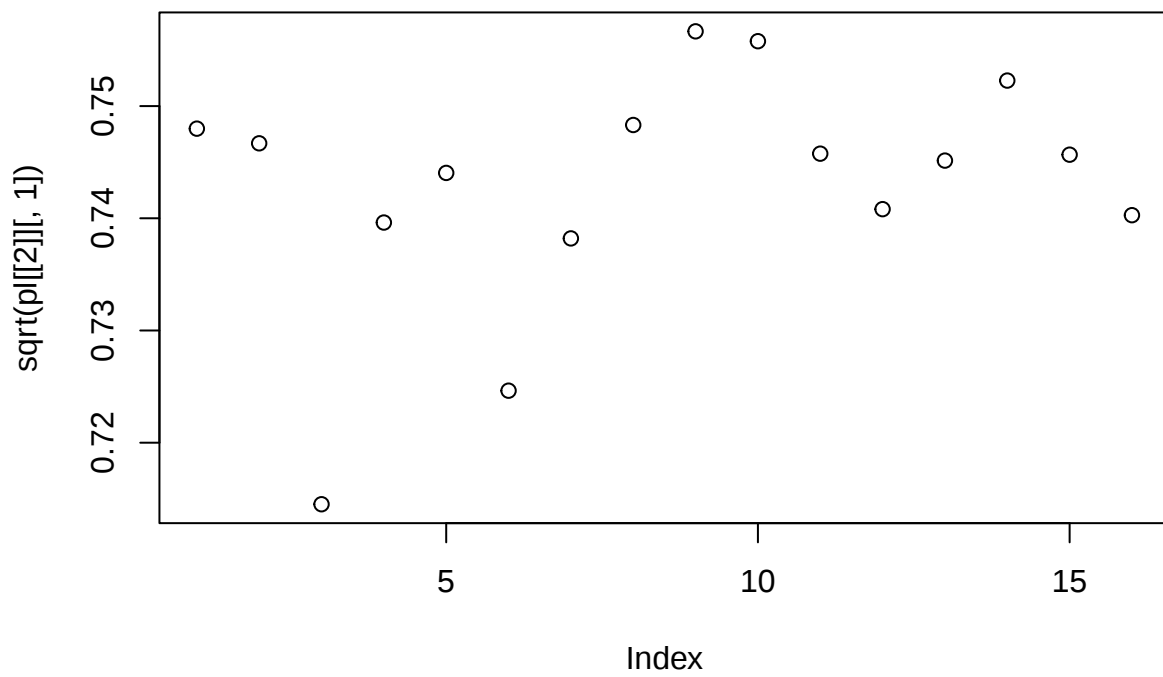
```
plot(gri[,2])
```



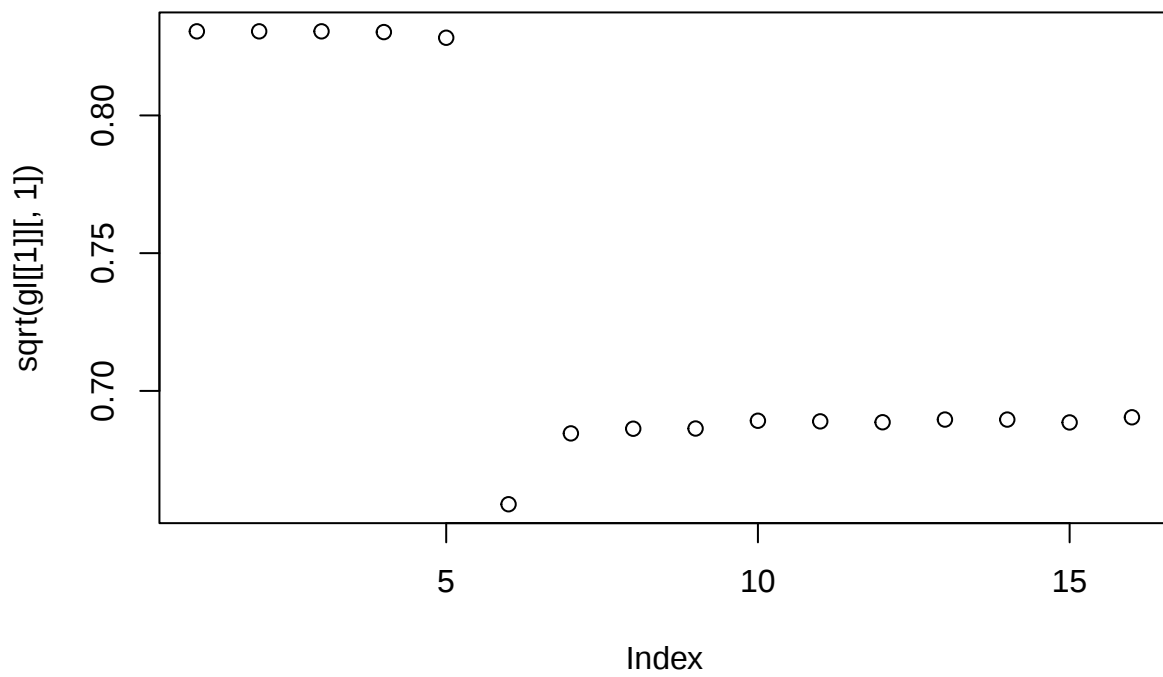
```
plot(sqrt(p1[[1]][,1]))
```



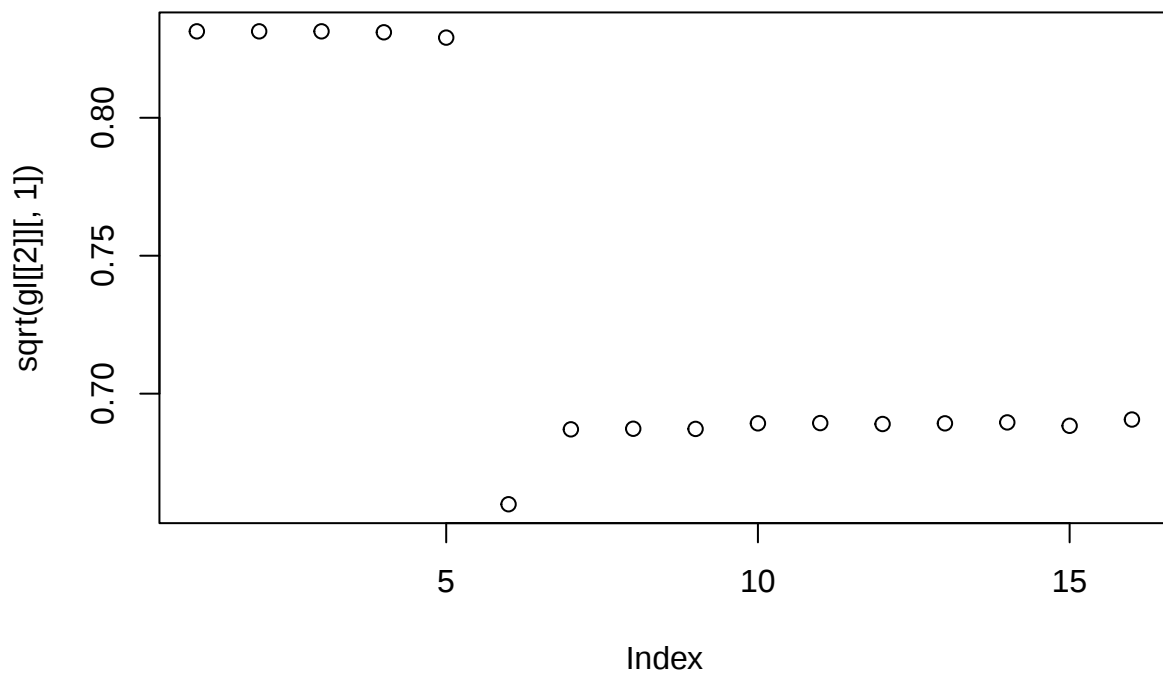
```
plot(sqrt(p1[[2]][, 1]))
```



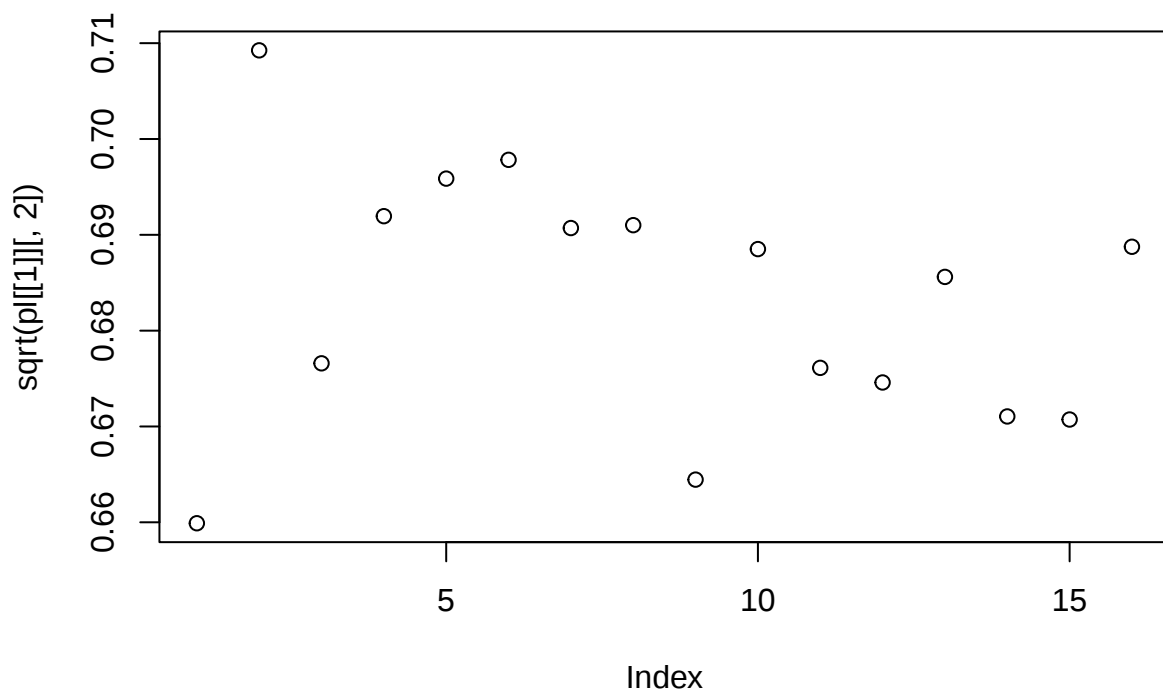
```
plot(sqrt(g1[[1]][,1]))
```



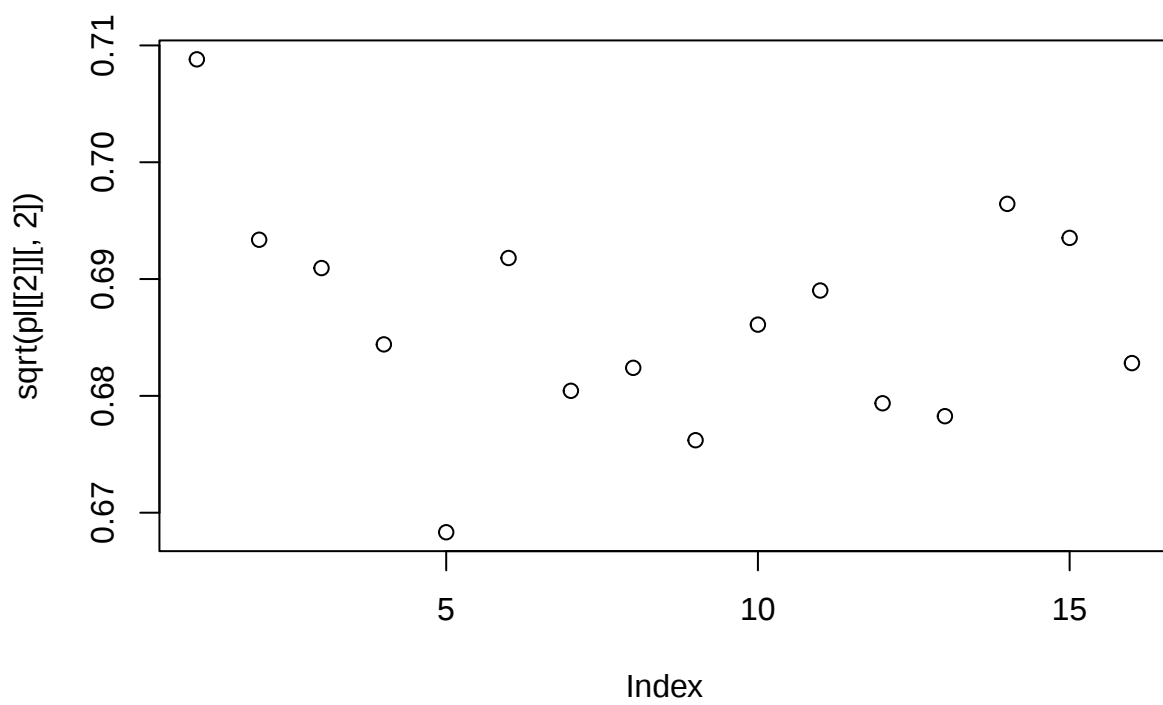
```
plot(sqrt(g1[[2]][, 1]))
```



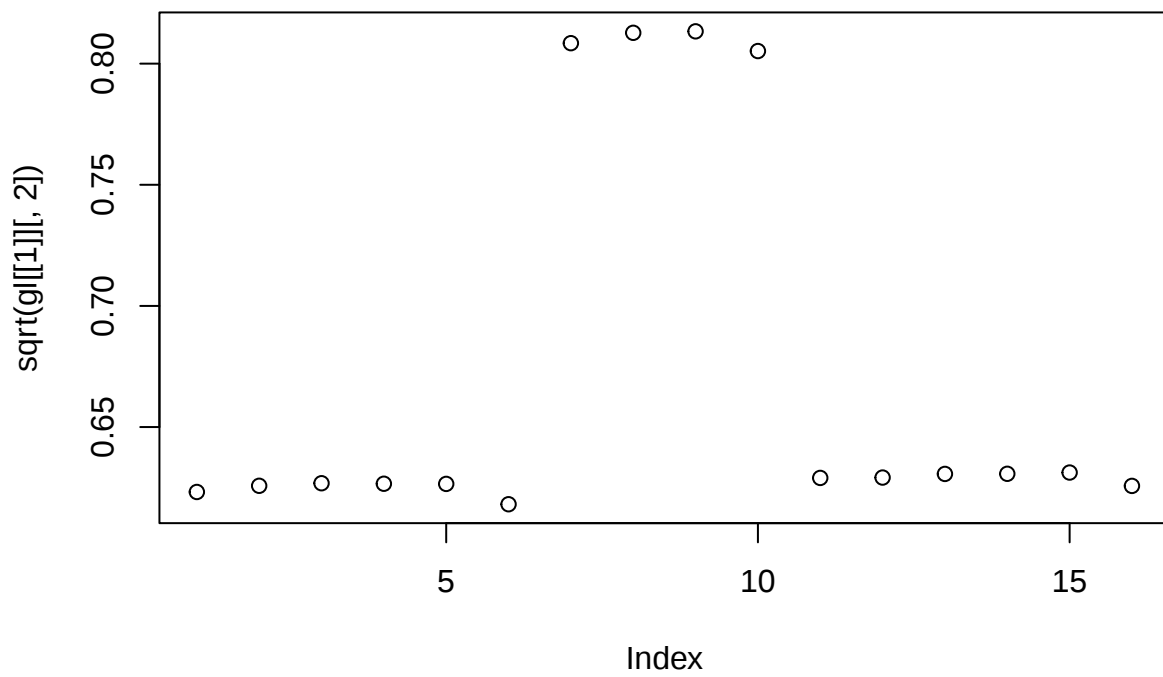
```
plot(sqrt(p1[[1]][,2]))
```



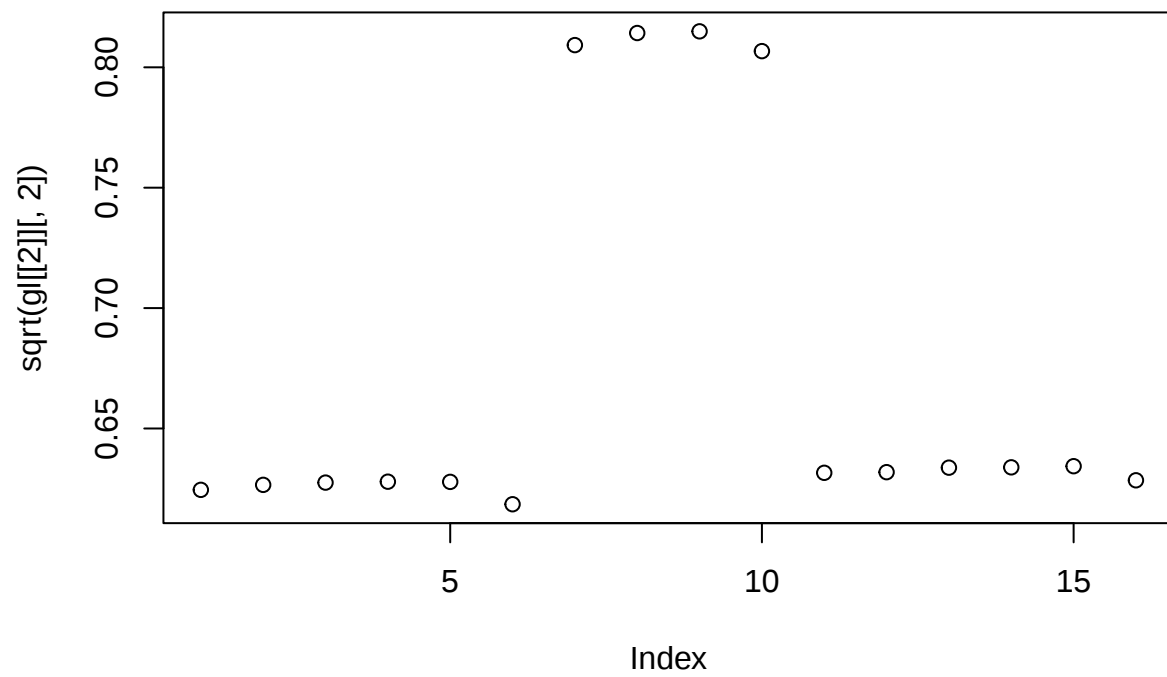
```
plot(sqrt(p1[[2]][, 2]))
```



```
plot(sqrt(g1[[1]][,2]))
```



```
plot(sqrt(g1[[2]][,2]))
```



```
Sys.time() - s
```

```
## Time difference of 11.57908 mins
```