Milestone 2: Core results 01

Anders

25/1/2023

Major goal

1. Predict outcome of cancer treatment with genetic profile as independent features
2. Features selection

Data

Clinical trail on breast cancer using two different drug combination

Evaluation

Preliminary experiments (on given data set) showed instability in prediction and feature selection between bootstrap samples. A more thorough evaluation is conducted. Two levels of evaluation is planed:

1. Comparison of models

1000 bootstrap models are fitted and then evaluated on the original sample. This gives a relative comparison of the various models with respect to data similar to the given data set. In addition to Correlation and MSE, frequency of selected features is compared.

2. Expected outcome of future patients

3 strategies are considered:
1. Repeated cross-validations (100 rep)
2. Bootstrap models with 0.632 (or 0.632?) adjustment
3. Test data set from a second trail (This trail have different responses)

Models

Lasso

Post Lasso

Ridge

Elastic Net

Boosting with stumps as base learner

RESULTS

Lasso

```r
load("/Users/anders/Documents/MASTER/Cancer/R_codeP01/lb_object_AllGenes01.RData")

# Correlation
cor_vec <- as.numeric(lb_object[[1]])
sum(is.na(cor_vec))/length(cor_vec)      # fraction of NA
```
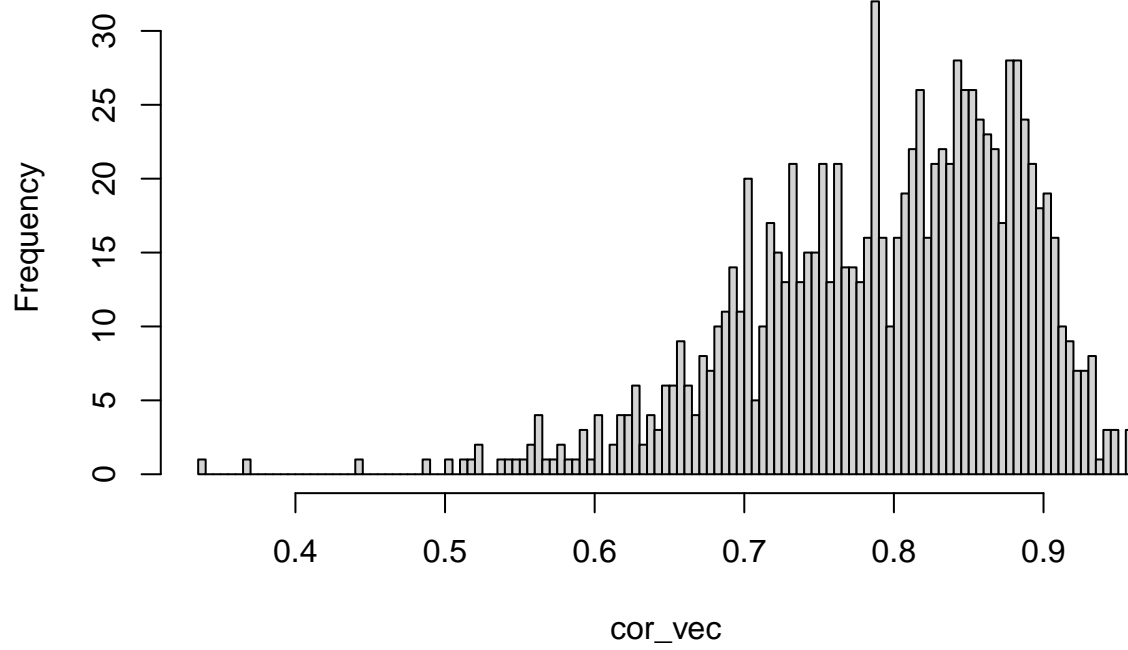
```
## [1] 0.002
```

```r
mean(cor_vec, na.rm=TRUE)
```

```
## [1] 0.7941413
```

```r
median(cor_vec, na.rm=TRUE)
```

```
## [1] 0.8101886
```

```r
var(cor_vec, na.rm=TRUE)
```

```
## [1] 0.008119272
```

```r
# par(mfrow=c(1,1))
hist(cor_vec, breaks = 100)
```

Histogram of cor_vec

Post Lasso

Ridge

Elastic Net

Boosting with stumps as base learner