report

ZJH

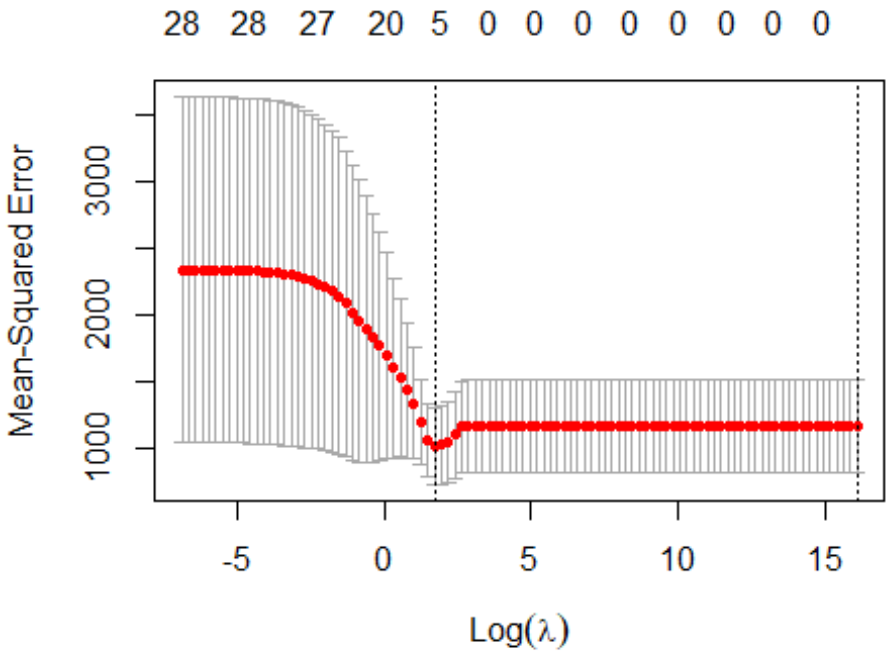
2022/3/29

```
set.seed(100)
tea <- readRDS('nonglin_tea.RDS')
analyze_data <- readRDS('analyze.RDS')
library(tidyverse)

numeric_dat <- analyze_data %>% select(.,-c(key,level,observe_ys,observ
e_date:sample_label,FAA:total_catechins,cutting_height,tea_buds_cm,year,
season,temp_differ))
train_idx <- sample(1:141,70)
test_idx <- !(1:141 %in%train_idx)
train <- numeric_dat[train_idx,]
test <- numeric_dat[test_idx,]
xtrain <- model.matrix(polyphenol~., train)[,-1]
ytrain <- train$polyphenol

ytest <- test$polyphenol
xtest <- model.matrix(polyphenol~., test)[,-1]

lambdas_to_try <- 10^seq(-3, 7, length.out = 100)
lasso_cv <- cv.glmnet(xtrain, ytrain, alpha = 1, lambda = lambdas_to_tr
y)
plot(lasso_cv)
```




```
best_lambda_lasso <- lasso_cv$lambda.min
best_lambda_lasso

## [1] 5.462277

lasso_mod <- glmnet(xtrain, ytrain, alpha = 1, lambda = best_lambda_las
so)
predict.glmnet(lasso_mod, type = 'coefficients')

## 31 x 1 sparse Matrix of class "dgCMatrix"
##                                s0
## (Intercept)       1.192076e+02
## duration_oc       1.891797e-01
## Total_leaf       -2.571372e-01
## GORatio           .
## OTRatio           .
## G1L               .
## G2L               .
## G3L               .
## G4L               .
## G5L               .
## G6L               .
## G7L               .
## G8L              -1.094133e+00
## G9L               .
## O1L               .
## O2L               .
```

```
## O3L              -5.807523e-02
## O4L              -9.718511e-01
## O5L               .
## O6L               .
## O7L               .
## O8L               6.790721e-01
## O9L               .
## leaf_number      -1.937634e+00
## avg_inter_node    .
## open_plane        .
## buds_weight_100   .
## acu_mean_temp     3.435788e-04
## rain              .
## growth_mean_temp  .
## Growth_length     .
```

```
fit <- lm(polyphenol~leaf_number+duration_oc+Total_leaf+G8L+O3L+O4L+O8L
+acu_mean_temp,data = test)
lasso_pred <- predict(lasso_mod,  newx = xtest)
lm_pred <- predict(fit)

mean((lasso_pred-ytest)^2)
```

```
## [1] 1089.807
```

```
mean((lm_pred-ytest)^2)
```

```
## [1] 738.1017
```