

report

ZJH

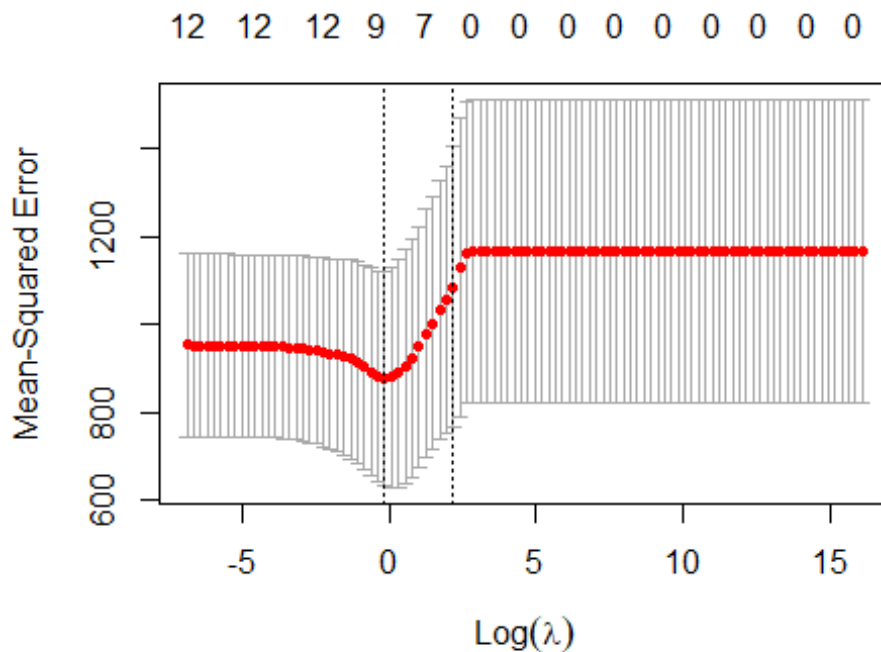
2022/3/29

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

numeric_dat <- analyze_data %>% select(., -c(key, level, observe_ys, observe_date:sample_label, FAA:total_catechins, G1L:09L, 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_try)
plot(lasso_cv)
```



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

## [1] 0.8497534

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

## 13 x 1 sparse Matrix of class "dgCMatrix"
##
##              s0
## (Intercept)    123.248609229
## duration_oc     0.283093707
## Total_leaf     -0.773099465
## GORatio         .
## OTRatio        -36.272840050
## leaf_number    -14.668985328
## avg_inter_node -10.065369534
## open_plane      .
## buds_weight_100 0.201941770
## acu_mean_temp   0.002401389
## rain            .
## growth_mean_temp 2.704565774
## Growth_length   .

fit <- lm(polyphenol~leaf_number+duration_oc+Total_leaf+acu_mean_temp,data = test)
```

```
lasso_pred <- predict(lasso_mod, newx = xtest)
lm_pred <- predict(fit)

mean((lasso_pred-ytest)^2)

## [1] 917.5125

mean((lm_pred-ytest)^2)

## [1] 913.1562
```

效果仍糟