200301-EDA_and_model-yuqi

Yuqi Miao ym2771

3/1/2020

data and manipulation

$$\log(\frac{\pi_i}{1-\pi_i}) = \mathbf{x}_i\boldsymbol{\beta}$$

validation using glm

questions or modify:

1. normalize or standardize?
2. how to standardize easily?

instead of using MSE, using pearson chi-square

validation

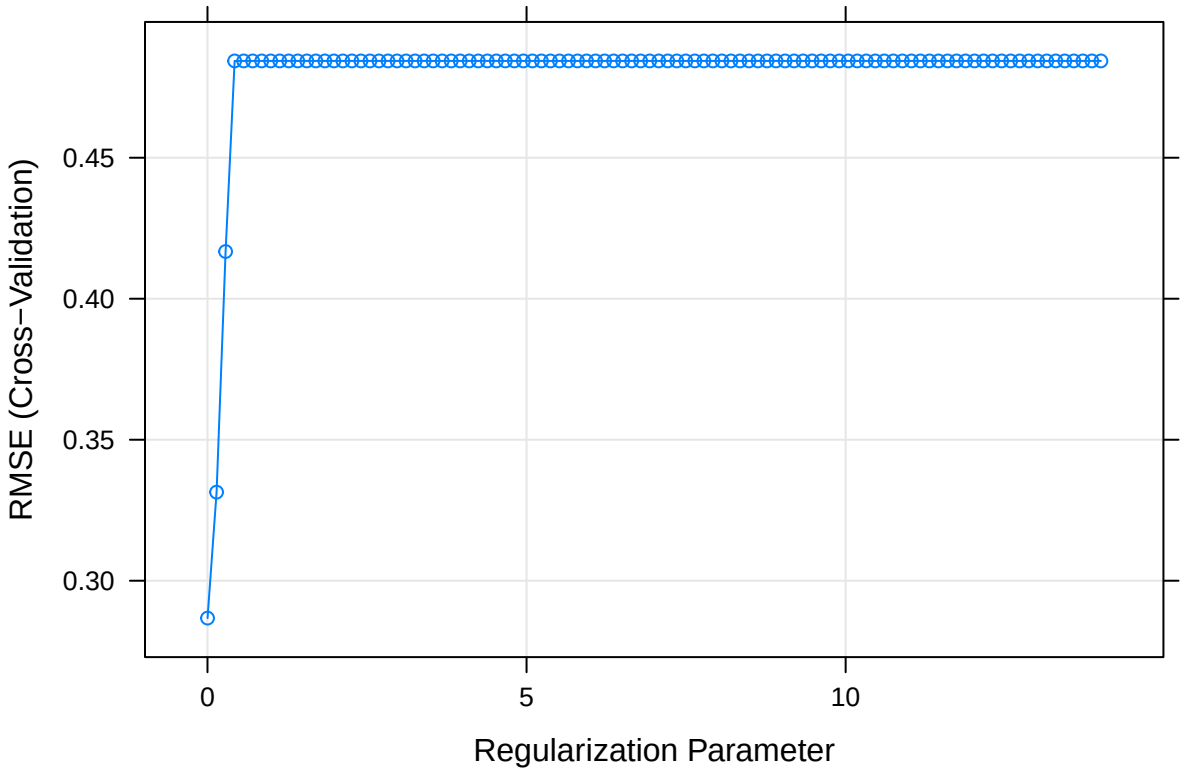
```
x.mat <- model.matrix(diagnosis~., cancer_package[-1])[,-1]
y.class <- cancer_package$diagnosis

ctrl1 <- trainControl(method = "cv", number = 5)
lasso.fit <- train(x.mat, y.class,
                   method = "glmnet",
                   tuneGrid = expand.grid(alpha = 1,
                                          lambda = tuning_grid),
                   # preProc = c("center", "scale"),
                   trControl = ctrl1)

lasso.fit$bestTune
```

```
##   alpha lambda
## 1     1  0.001
```

```
plot(lasso.fit)
```




```
min(lasso.fit$results$RMSE)
```

```
## [1] 0.2867455
```

```
co=coef(lasso.fit$finalModel,lasso.fit$bestTune$lambda)
co2=co@x

names(co2)=co@Dimnames[[1]]
co2 %>% as.data.frame() %>% knitr::kable()
```

	.
(Intercept)	-1.5315370
texture_mean	0.0216265
perimeter_mean	0.0156391
area_mean	-0.0007061
smoothness_mean	2.3356763
compactness_mean	-1.2677496
concavity_mean	0.2262630
concave points_mean	6.1420082
symmetry_mean	1.0106152
fractal_dimension_mean	-1.3592103