

Hitters exercise

Torben

August 24, 2018

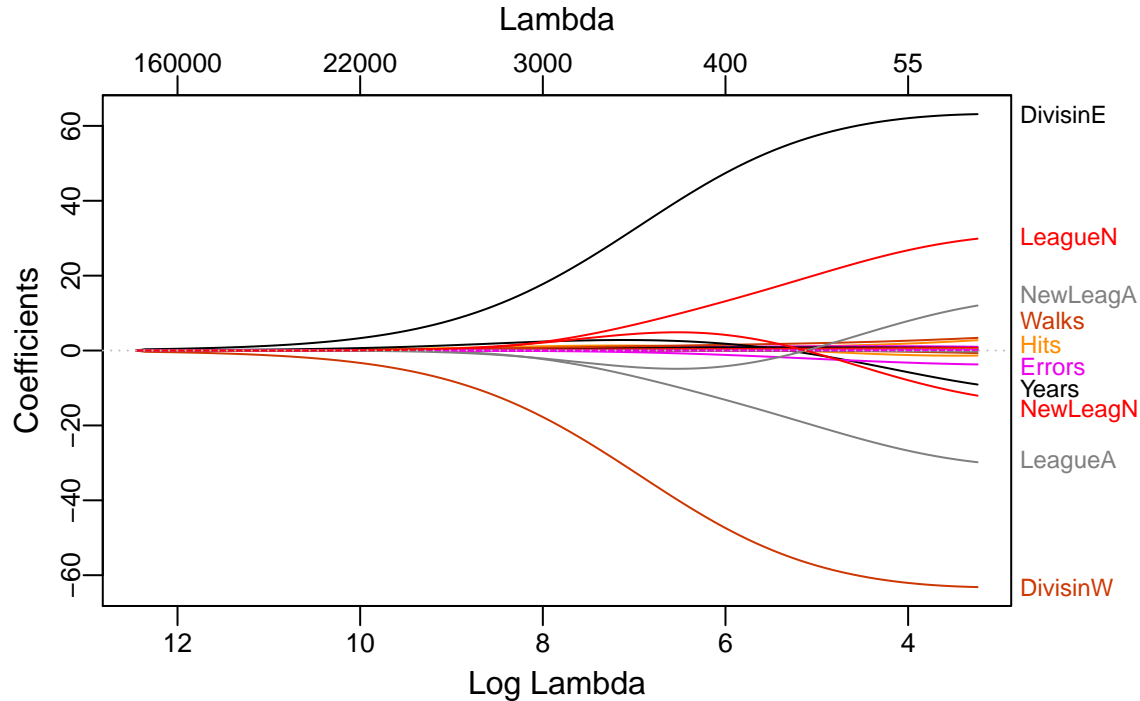
Load data and packages

```
library(tidyverse)
library(ISLR)
library(glmnetUtils)
library(plotmo)

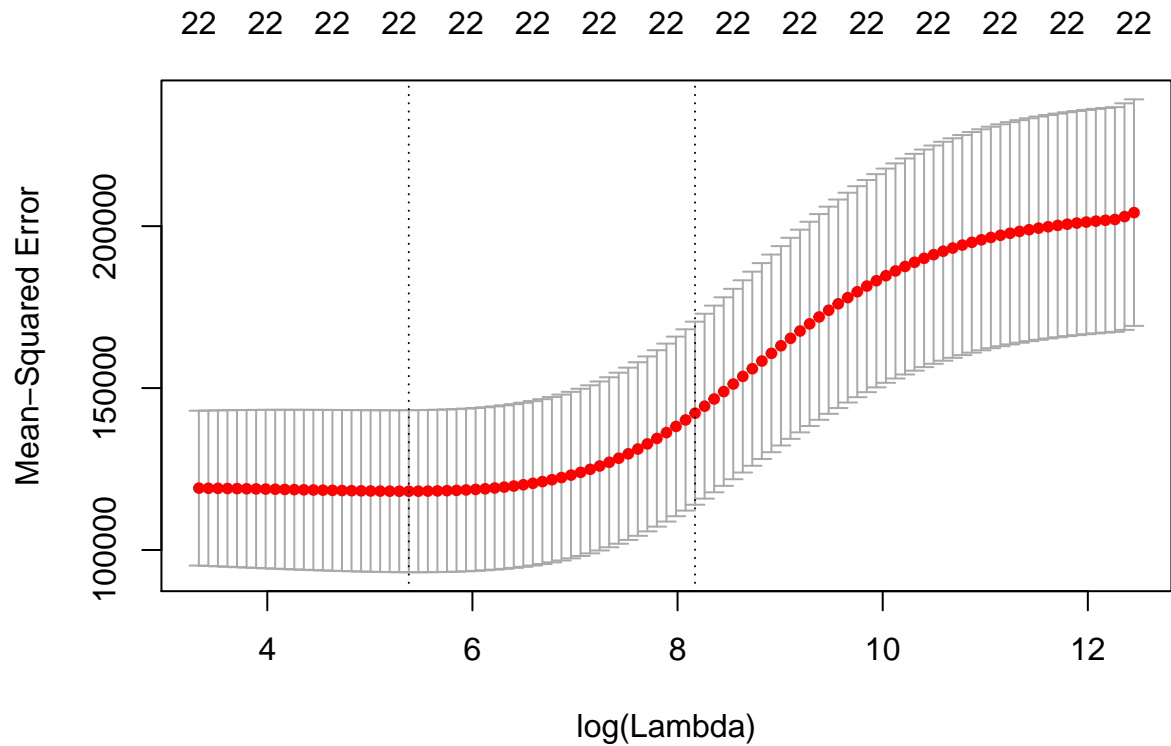
rmse <- function(pred, obs){
  sqrt(mean((pred-obs)^2, na.rm = TRUE))
}
```

Ridge Regression

```
salary_ridge <- glmnet(Salary ~ ., alpha=0, data = Hitters) ## alpha = 0; Ridge
# plot(salary_ridge)
plot_glmnet(salary_ridge)
```



```
salary_ridge_cv <- cv.glmnet(Salary ~ ., alpha=0, data = Hitters) ## alpha = 0; Ridge
plot(salary_ridge_cv)
```



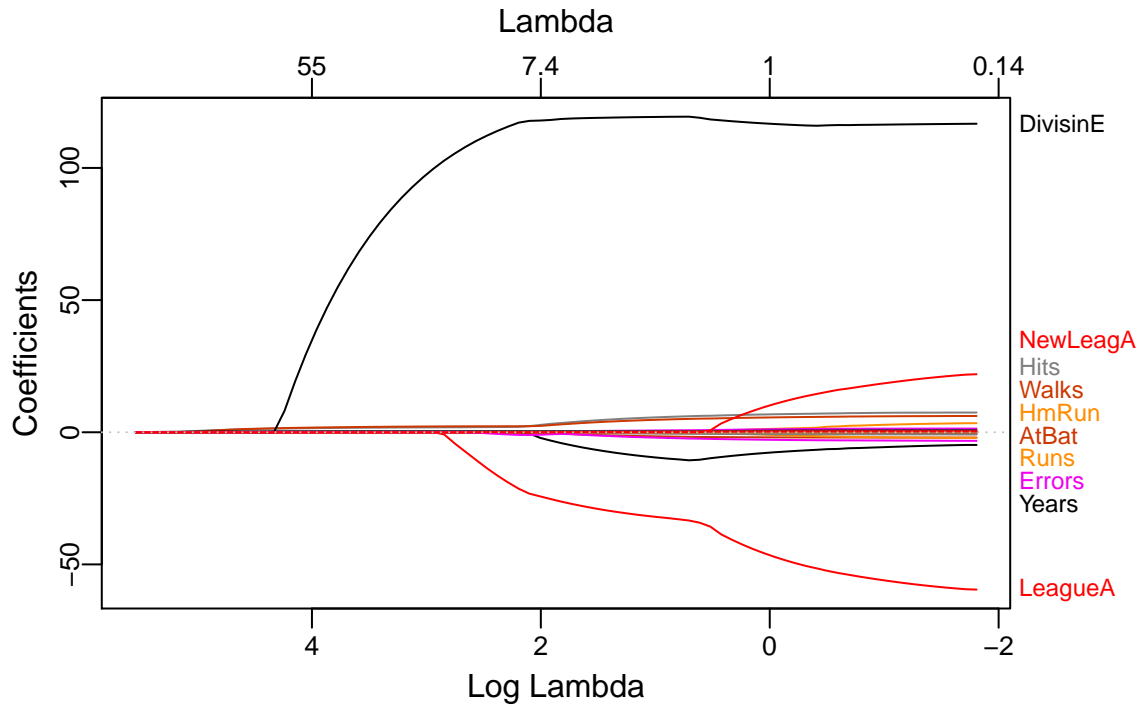
```
salary_ridge_lambda_1se <- salary_ridge_cv$lambda.1se
salary_ridge_1se_pred <- predict(salary_ridge, newdata = Hitters, s = salary_ridge_lambda_1se)

rmse(pred = salary_ridge_1se_pred, obs = Hitters$Salary)

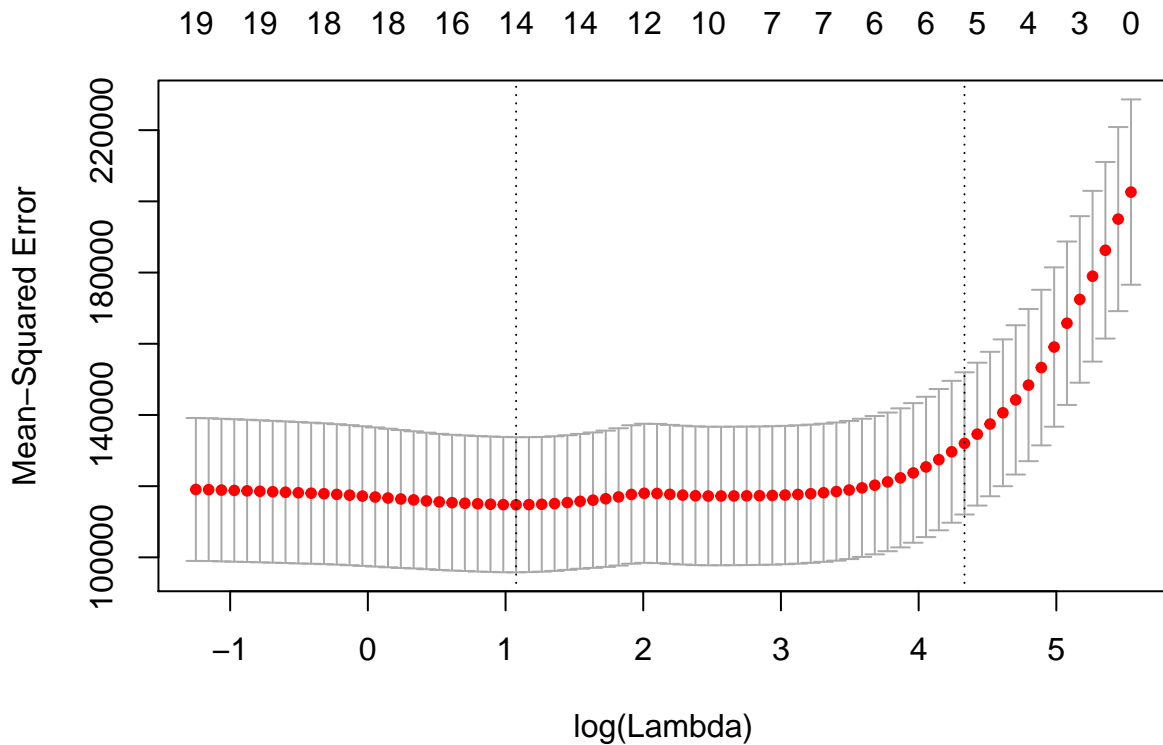
## [1] 371.2862
```

LASSO

```
salary_lasso <- glmnet(Salary ~ ., alpha=1, data = Hitters) ## alpha = 1; LASSO
# plot(salary_lasso)
plot_glmnet(salary_lasso)
```



```
salary_lasso_cv <- cv.glmnet(Salary ~ ., alpha=1, data = Hitters) ## alpha = 0; LASSO
plot(salary_lasso_cv)
```



```
salary_lasso_lambda_1se <- salary_lasso_cv$lambda.1se
salary_lasso_1se_pred <- predict(salary_lasso, s = salary_lasso_lambda_1se, newdata = Hitters)
rmse(pred = salary_lasso_1se_pred, obs = Hitters$Salary)
```

[1] 348.2684

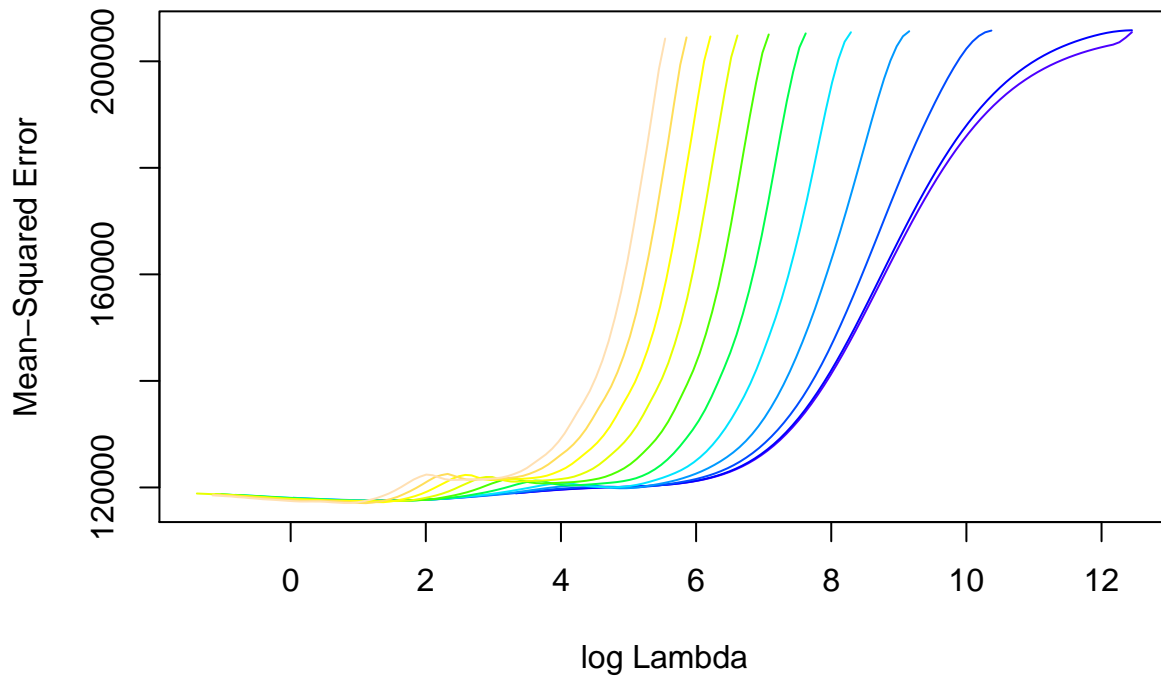
```
salary_lasso_active_coef <- glmnet::nonzeroCoef(coef(salary_lasso_cv, s = "lambda.1se"))[-1]
rownames(coef(salary_lasso_cv, s = "lambda.1se"))[salary_lasso_active_coef]
```

```
## [1] "Hits"      "Walks"     "CRuns"     "CRBI"     "PutOuts"
```

Elastic Net

```
enet_fit <- cva.glmnet(Salary ~ ., data = Hitters)
```

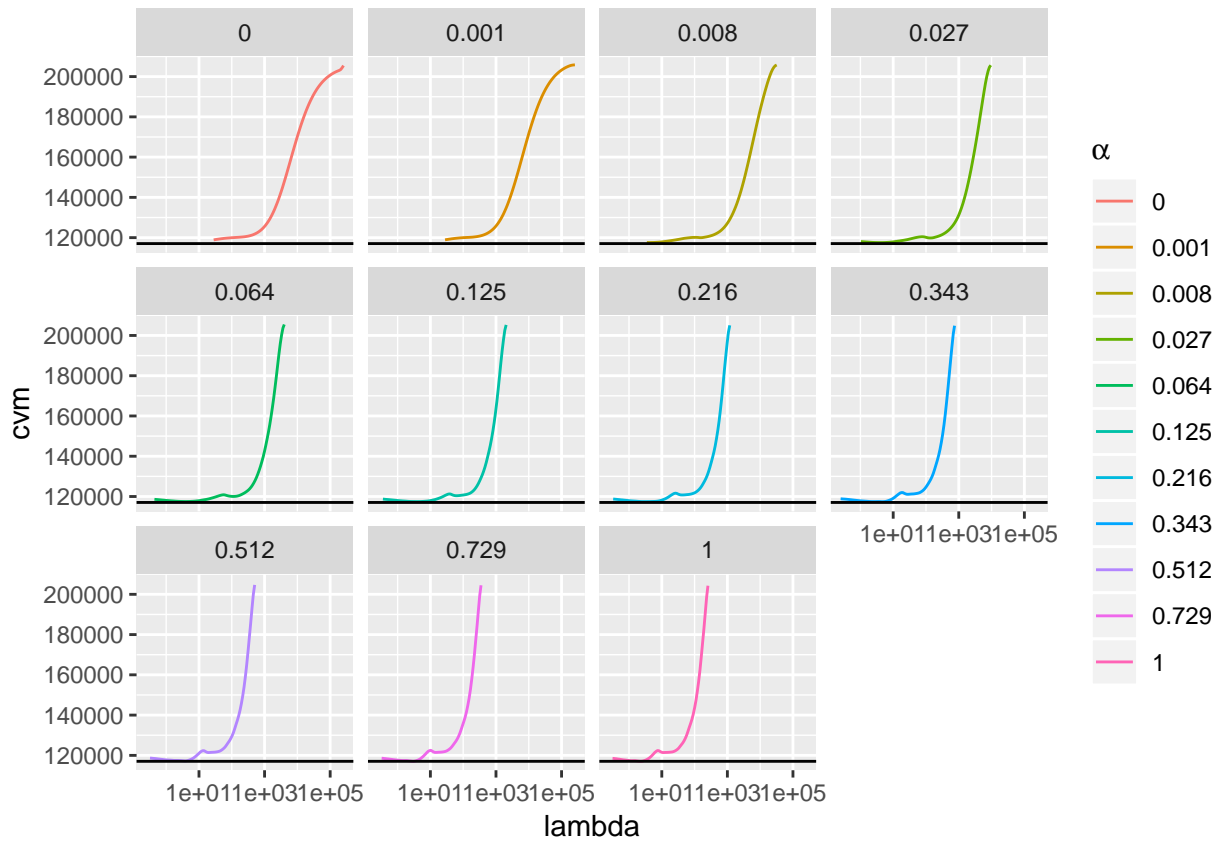
```
plot(enet_fit)
```



Adding some more info to outcome

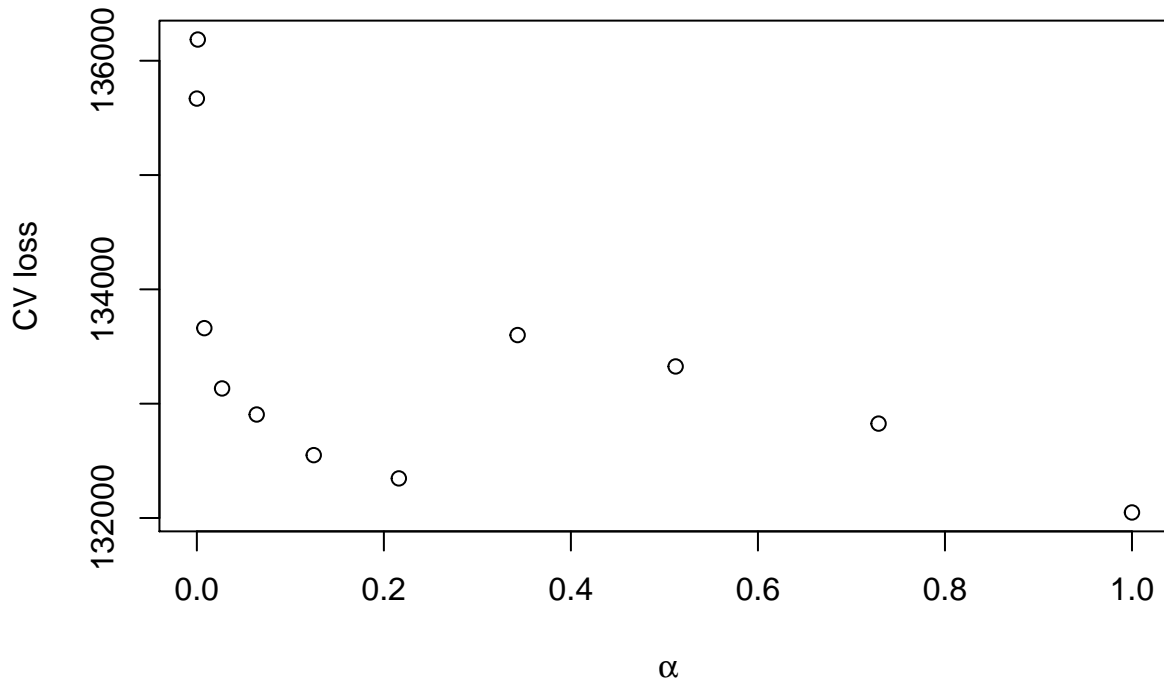
```
names(enet_fit$modlist) <- enet_fit$alpha
enet_fit_cv <- enet_fit$modlist %>% lapply(function(x) do.call("cbind", x[c(1:6)])) %>% as_tibble()
enet_fit_cv <- enet_fit_cv %>%
  bind_rows(.id = "alpha") %>%
  mutate(alpha = as.numeric(alpha))
```

```
enet_fit_cv %>%
  mutate(f_alpha = factor(alpha), cv_min = min(cvm)) %>%
  ggplot(aes(x = lambda, y = cvm, colour = f_alpha)) +
  scale_x_log10() +
  # geom_ribbon(aes(ymax = cvup, ymin = cvlo, fill = ..colour..), show.legend = FALSE, alpha = 0.1) +
  geom_line() +
  geom_hline(aes(yintercept = cv_min)) +
  labs(colour = expression(alpha)) +
  facet_wrap(~f_alpha) # + coord_cartesian(ylim = c(115196, 115196 + 1000))
```

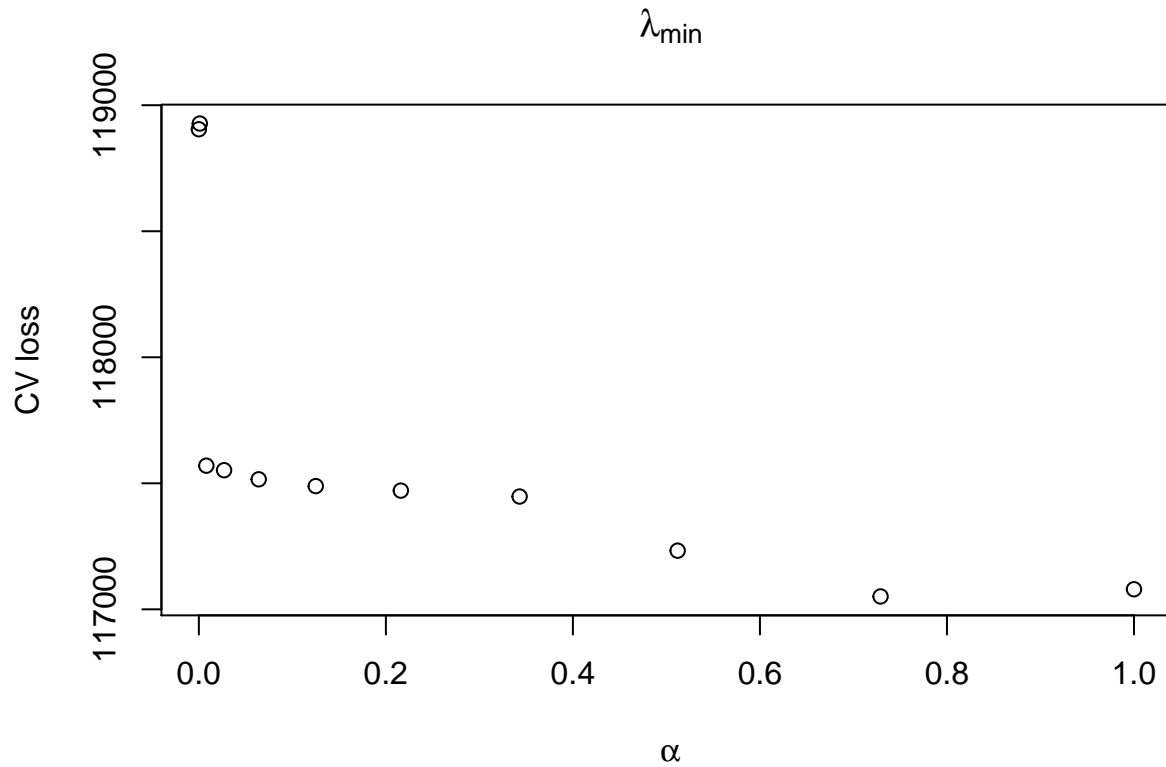


```
minlossplot(enet_fit, cv.type = "1se", main = ~lambda['1se'], xlab = ~alpha)
```

λ_{1se}



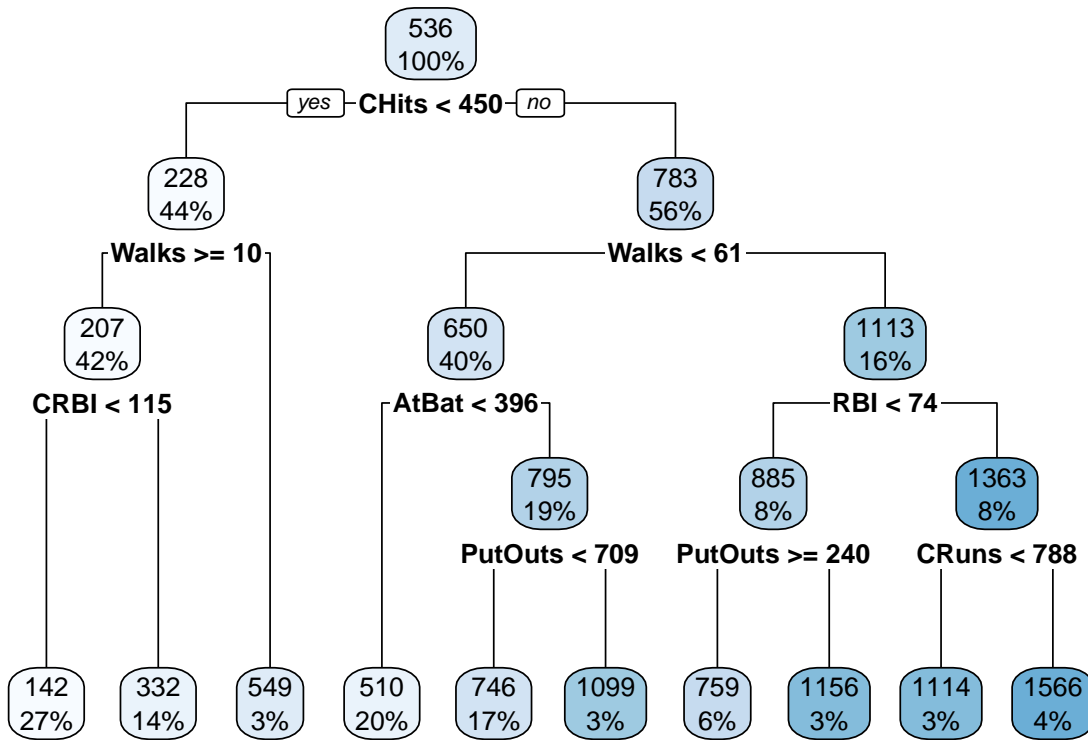
```
minlossplot(enet_fit, cv.type = "min", main = ~lambda[ $\lambda$ min], xlab = ~alpha)
```



rpart

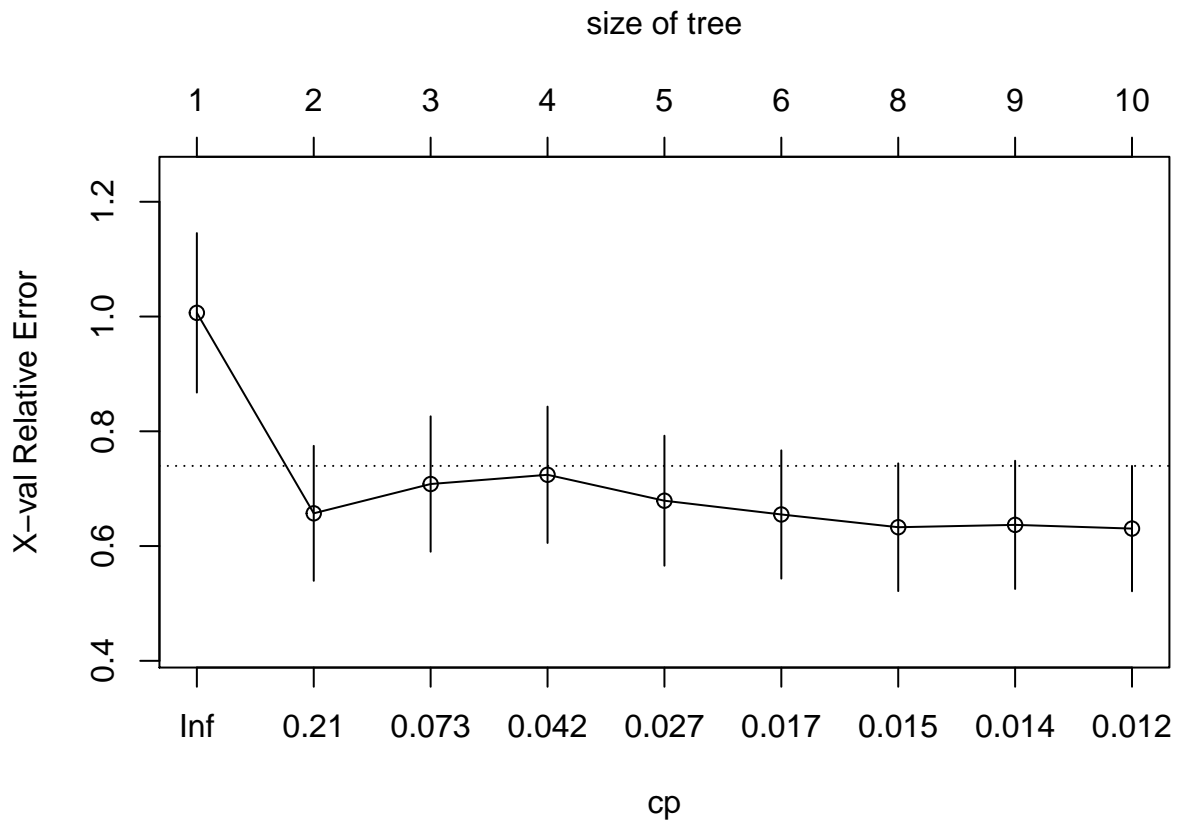
```
library(rpart)
library(rpart.plot)

set.seed(20182408)
salary_rpart_default <- rpart(Salary ~ ., data = Hitters)
rpart.plot(salary_rpart_default)
```



Plateau? Increase cp

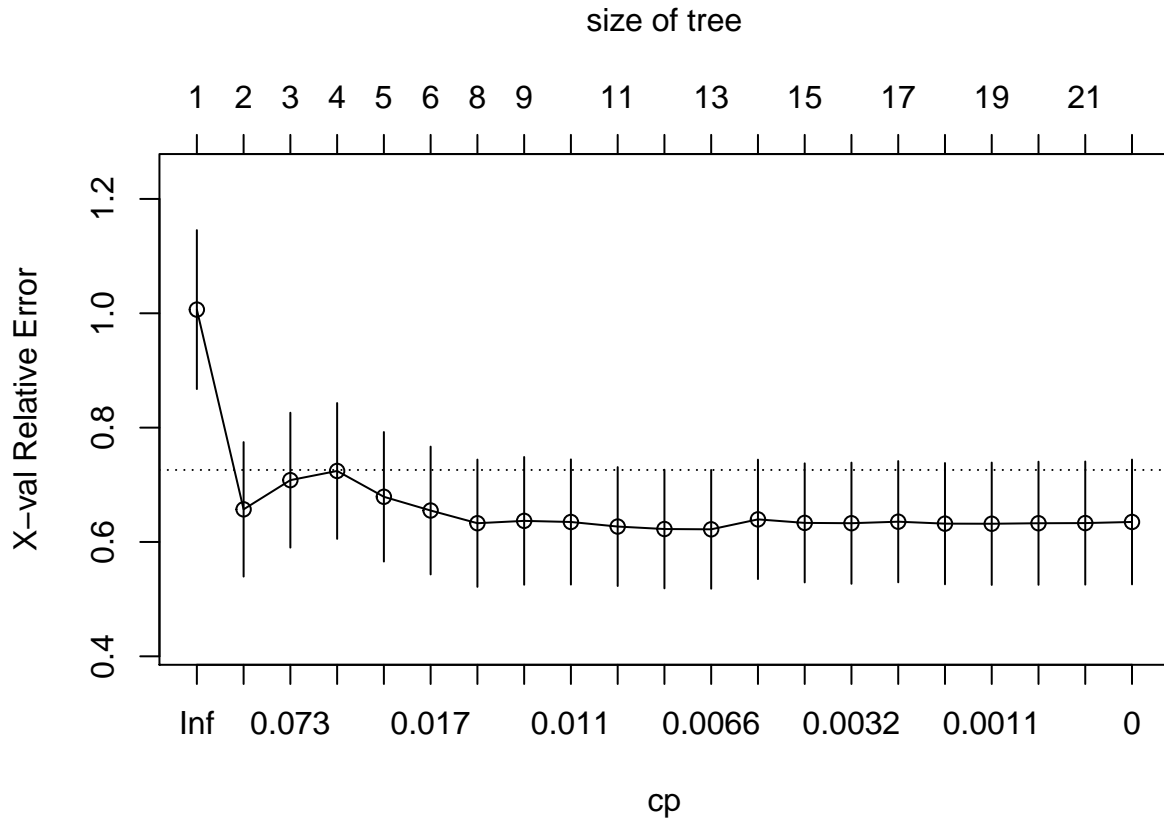
```
plotcp(salary_rpart_default)
```



```

set.seed(20182408)
salary_rpart <- rpart(Salary ~ ., data = Hitters, cp = 0)
plotcp(salary_rpart)

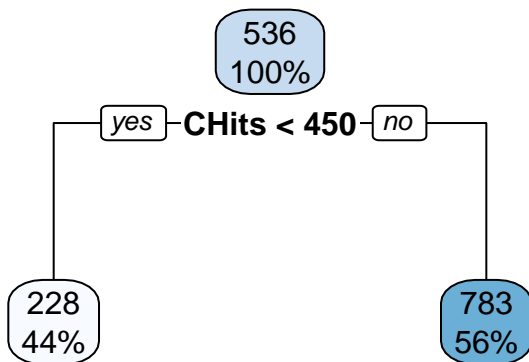
```



```

salary_rpart_pruned_0.21 <- prune(salary_rpart, cp = 0.21)
salary_rpart_pruned_0.073 <- prune(salary_rpart, cp = 0.073)
rpart.plot(salary_rpart_pruned_0.21)

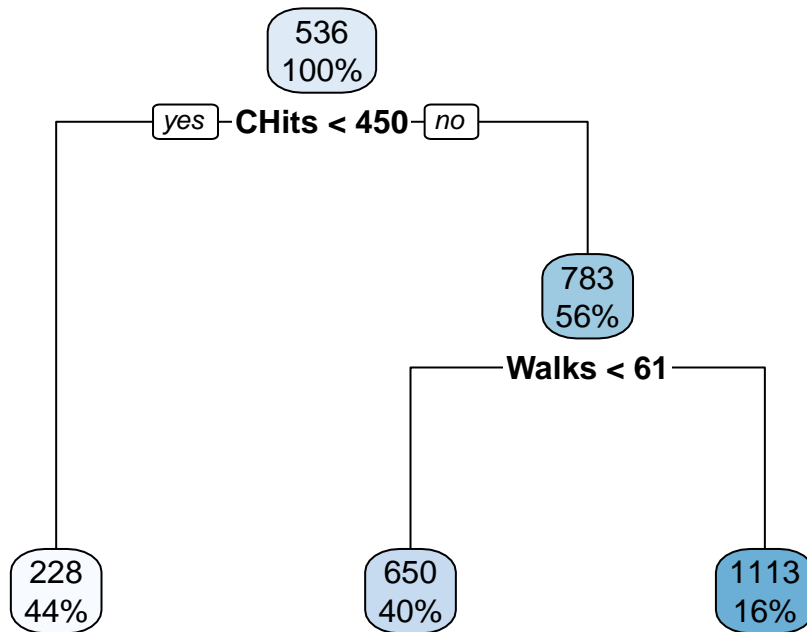
```



```

rpart.plot(salary_rpart_pruned_0.073)

```

```
salary_rpart_pred_0.21 <- predict(salary_rpart_pruned_0.21, newdata = Hitters)
salary_rpart_pred_0.073 <- predict(salary_rpart_pruned_0.073, newdata = Hitters)
salary_rpart_pred_default <- predict(salary_rpart_default, newdata = Hitters)
```

```
rmse(pred = salary_rpart_pred_0.21, obs = Hitters$Salary)
```

```
## [1] 355.9185
```

```
rmse(pred = salary_rpart_pred_0.073, obs = Hitters$Salary)
```

```
## [1] 319.8374
```

```
rmse(pred = salary_rpart_pred_default, obs = Hitters$Salary)
```

```
## [1] 263.2883
```