Stacked Regression: Boston dataset

Data

```
rm(list=ls())
library(MASS)

set.seed(123)
istrain = rbinom(n=nrow(Boston),size=1,prob=0.5)>0
train <- Boston[istrain,]
( n=nrow(train) )
```

```
## [1] 235
```

```
test = Boston[!istrain,-14]
test.y = Boston[!istrain,14]
( m=nrow(test) )
```

```
## [1] 271
```

Program the following algorithm:

The training and test data are

$$(x_1, y_1), \ldots, (x_n, y_n), \quad (x_1^*, y_1^*), \ldots, (x_m^*, y_m^*)$$

with $n = 235$ and $m = 271$ for the Boston data set.

The response variable is `medv`, and the predictor variables are `crim, zn, indus, chas, nox, rm, age, dis, rad, tax, ptratio, black, lstat`.

1. Fit a library of L models $\hat{f}_1, \ldots, \hat{f}_L$ to the training set, with $L = 2$ and $\hat{f}_1$ a linear model with all predictors

```
fit1 = lm(medv ~ ., train)
# to see the names of the predictors,
attr(summary(fit1)$term,"term.labels")
```

```
##  [1] "crim"    "zn"      "indus"   "chas"    "nox"     "rm"      "age"
##  [8] "dis"     "rad"     "tax"     "ptratio" "black"   "lstat"
```

and $\hat{f}_2$ a classification tree with all predictors and default settings

```
library(rpart)
fit2 = rpart(medv ~ ., train)
```

2. Let $\hat{f}_l^{-i}(x_i)$ be the prediction at x_i using model l fitted to the training data with the ith training observation (x_i, y_i) removed

3. Obtain the weights by least squares

$$\hat{w}_1, \ldots, \hat{w}_L = \underset{w_1,\ldots,w_L}{\arg\min} \sum_{i=1}^{n} \left[y_i - \sum_{l=1}^{L} w_l \hat{f}_l^{-i}(x_i) \right]^2$$

4. Compute the predictions for the test set as

$$\hat{f}_{\text{stack}}(x_i^*) = \sum_{l=1}^{L} \hat{w}_l \hat{f}_l(x_i^*), \quad i = 1, \ldots, m$$

5. Compute the mean squared error for the test set

$$\text{MSE}_{\text{Tr}}^{\text{stack}} = \frac{1}{m} \sum_{i=1}^{m} (y_i^* - \hat{f}_{\text{stack}}(x_i^*))^2$$

and compare with

$$\text{MSE}_{\text{Tr}}^{l} = \frac{1}{m} \sum_{i=1}^{m} (y_i^* - \hat{f}_l(x_i^*))^2, \quad l = 1, \ldots, L.$$

```
## MSE stack:  21.95002
## MSE lm:  28.83767
## MSE rpart:  24.39206
```