BA810: Team 2: Group Project

Luke Hong, Carmen Cheung, Tavishi Pandey, Zichen Wang, Jiazhi Jia, Mona Ma

10/01/2021

Setup and Load the Airbnb dataset and Train/Test split

```
library(data.table)
library(ggplot2)
library(ggthemes)
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-2
```

```
library(caTools)
# [, -1] means take all columns of the matrix except the first column, which is an index
dd <- fread("ab_updated.csv")[, -1]
# set seed for splitting training / test data
set.seed(810)
sample <- sample.split(dd$num, SplitRatio = .70)
train <- subset(dd , sample == TRUE)
test <- subset(dd, sample == FALSE)
# the [, -1] means take all columns of the matrix except the first column, which is an intercept
# splitting x and y train/test
x.train <- model.matrix( ~ . - host_since - zipcode, train)[, -2][, -1]
y.train <- train$log_price
x.test <- model.matrix( ~ . - host_since - zipcode, test)[, -2][, -1]
y.test <- test$log_price
```

Lasso Regression

Predict responses / Compute MSEs for Lasso

```
# use the cv.glmnet command to automatically select the best value for the lambda hyper-parameter.
fit.lasso <- cv.glmnet(x.train, y.train, alpha = 1, nfolds = 10)

# computing MSE on the training/test data
yhat.train.lasso <- predict(fit.lasso, x.train)
yhat.test.lasso <- predict(fit.lasso, x.test)

mse.train.lasso <- mean((y.train - yhat.train.lasso)^2)
mse.test.lasso <- mean((y.test - yhat.test.lasso)^2)

mse.train.lasso
```

```
## [1] 0.2077325
```

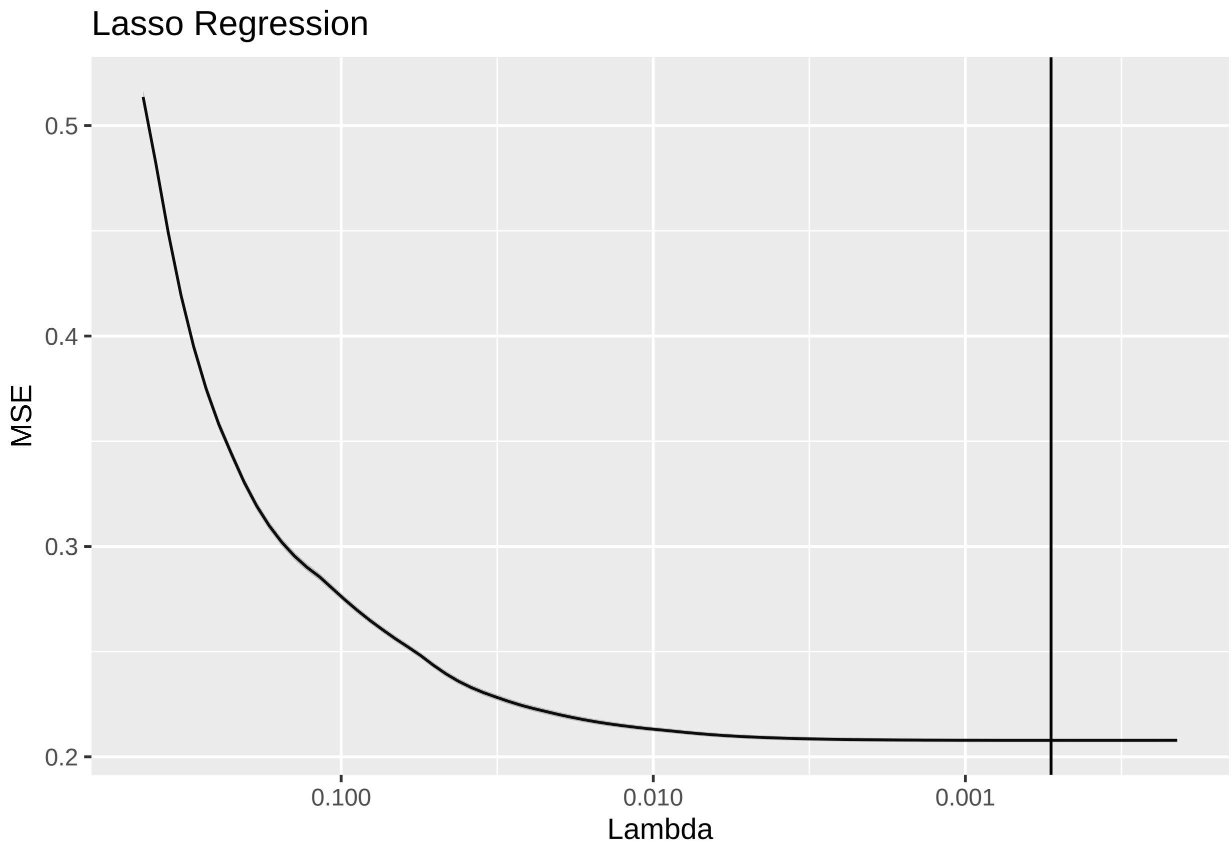
```r
mse.test.lasso
```

```
## [1] 0.2106664
```

Lasso Lambda and MSE Graph

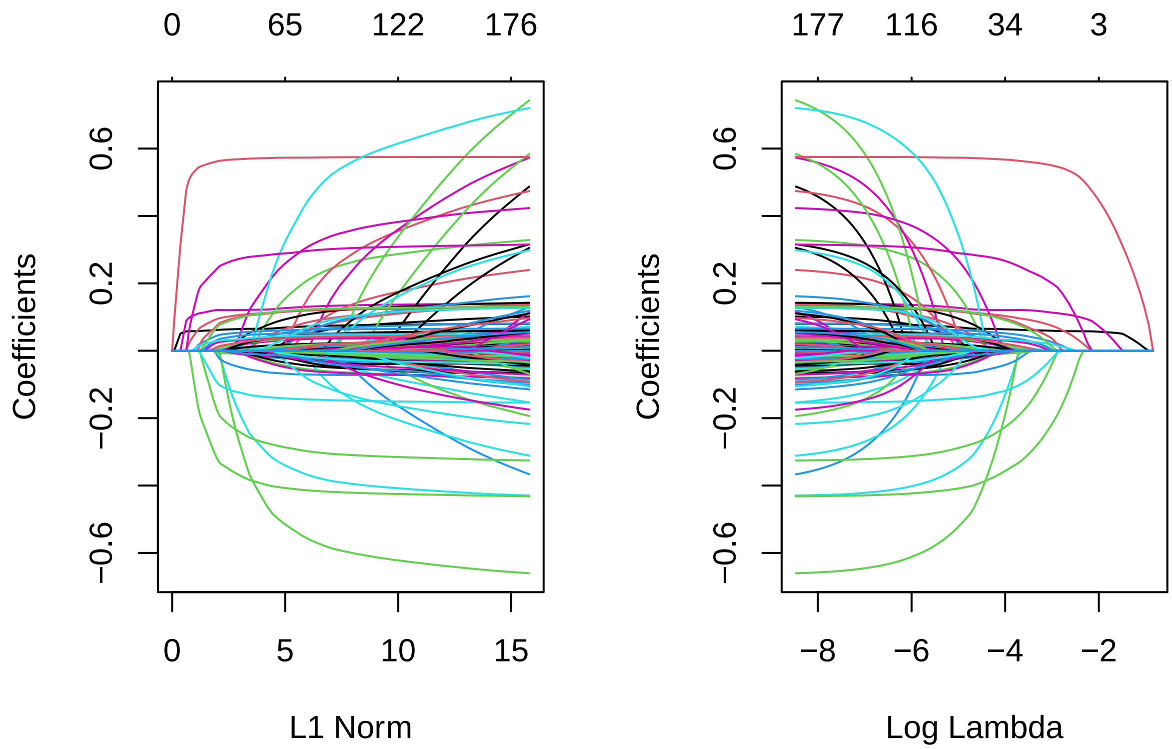
```r
# manual function to reverse on x axis and scale at the same time on ggplot2
library("scales")
reverselog_trans <- function(base = exp(1)) {
    trans <- function(x) -log(x, base)
    inv <- function(x) base^(-x)
    trans_new(paste0("reverselog-", format(base)), trans, inv,
              log_breaks(base = base),
              domain = c(1e-100, Inf))
}
# The broom package takes the messy output of built-in functions in R, and turns them into tidy tibbles
library(broom)

tidied_cv <- tidy(fit.lasso)
glance_cv <- glance(fit.lasso)
g <- ggplot(tidied_cv, aes(lambda, estimate)) +
  geom_line() + scale_x_continuous(trans=reverselog_trans(10))
g <- g + geom_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = .25)
g <- g +
  geom_vline(xintercept = glance_cv$lambda.min) + ggtitle("Lasso Regression") +
  xlab("Lambda") + ylab("MSE")
g
```



Lasso Coefficients Graph

```
# each colored line represents the value taken by a different coefficient in your model, plot to show t
op <- par(mfrow=c(1, 2))
# L1 norm is the regularization term for Lasso
plot(fit.lasso$glmnet.fit, "norm",   label=TRUE)
plot(fit.lasso$glmnet.fit, "lambda", label=TRUE)
```




```
par(op)
```

Lasso Coefficients

```
# examine the coefficients associated with min lambda.min, the value of lambda that gives minimum mean
# lambda.1se, the value of lambda that gives the most regularized model such that the cross-validated e
# One line of reasoning suggests using 1se because it hedges against overfitting by selecting a larger
lasso_coef <- coef(fit.lasso, s = fit.lasso$lambda.min)
lasso_coef_df <- data.frame(name = lasso_coef@Dimnames[[1]][lasso_coef@i + 1], coefficient = lasso_coef(
lasso_coef_df <- lasso_coef_df[order(lasso_coef_df[,2]),]

lasso_coef_df
```

```
##                                        name   coefficient
## 26                      property_type_Hostel -6.537507e-01
## 44                    `room_type_Shared room` -4.305645e-01
## 23                        property_type_Dorm -4.263334e-01
## 27                         property_type_Hut -3.298526e-01
## 55                              city_Chicago -3.240001e-01
## 34                        property_type_Tent -2.922666e-01
## 102                               beachfront -2.081280e-01
## 94                              air_purifier -1.703202e-01
## 136                              lake_access -1.609626e-01
## 57                                   city_LA -1.535623e-01
## 28                    `property_type_In-law` -1.388668e-01
```

```
## 123                         free_parking_on_street -1.054627e-01
## 140                                      microwave -9.406879e-02
## 146                 path_to_entrance_lit_at_night -9.308866e-02
## 73                                  self_check_in -8.885417e-02
## 121        fixed_grab_bars_for_showers_and_toilet -8.180876e-02
## 80                              wireless_internet -7.887103e-02
## 4                              host_response_rate -7.793355e-02
## 118                                   ev_charger -7.430146e-02
## 115                                doorman_entry -6.637377e-02
## 125                            garden_or_backyard -6.636731e-02
## 86                                       kitchen -6.426921e-02
## 120                                firm_mattress -5.612909e-02
## 111                         disabled_parking_spot -4.376888e-02
## 45                                bed_type_Airbed -4.346731e-02
## 18                            property_type_Cabin -4.245442e-02
## 143                                outlet_covers -4.065635e-02
## 8                                           beds -3.975853e-02
## 68                      free_parking_on_premises -3.821602e-02
## 157                              smoking_allowed -3.797718e-02
## 104                                         cats -3.750648e-02
## 155                               `ski_in/ski_out` -3.635191e-02
## 48                   cancellation_policy_moderate -3.114337e-02
## 159                           table_corner_guards -3.111563e-02
## 167                                window_guards -3.030258e-02
## 74                                smoke_detector -2.895310e-02
## 105                       children_books_and_toys -2.830168e-02
## 95                                     baby_bath -2.589733e-02
## 71                           lock_on_bedroom_door -2.558131e-02
## 98                                       bathtub -2.488808e-02
## 166                       wide_hallway_clearance -2.438925e-02
## 129                                      hangers -2.415256e-02
## 119                              fireplace_guards -2.248289e-02
## 101                              beach_essentials -2.129775e-02
## 72                     pets_live_on_this_property -2.110975e-02
## 93                         `accessible-height-bed` -1.674291e-02
## 69                                host_greets_you -1.557147e-02
## 77  `translation_missing:_en.hosting_amenity_50` -1.546216e-02
## 25                      `property_type_Guest suite` -1.513396e-02
## 149                              private_entrance -1.504831e-02
## 89                                         stove -1.500148e-02
## 106                           children_dinnerware -1.278246e-02
## 67                                 first_aid_kit -1.156566e-02
## 131                                    hot_water -1.105145e-02
## 70                      laptop_friendly_workspace -1.064124e-02
## 83                                    essentials -1.009136e-02
## 138                        long_term_stays_allowed -9.267491e-03
## 47                                bed_type_Futon -9.229721e-03
## 163                         wide_clearance_to_bed -9.084921e-03
## 78                          wheelchair_accessible -8.639570e-03
## 76  `translation_missing:_en.hosting_amenity_49` -7.735661e-03
## 141                                        other -7.212209e-03
## 134                                         iron -6.146625e-03
## 92                      `twentyfour-hour-check-in` -5.793289e-03
## 158                             `step-free_access` -5.693596e-03
```

```
## 151                               refrigerator -4.354918e-03
## 135                                     keypad -4.275497e-03
## 19                   `property_type_Camper/RV` -3.134233e-03
## 63                    carbon_monoxide_detector -2.259116e-03
## 164         `wide_clearance_to_shower_&_toilet` -2.104000e-03
## 156                                 smart_lock -1.939705e-03
## 154                                safety_card -1.753517e-03
## 91                                      washer -1.525481e-03
## 96                                baby_monitor -8.994539e-04
## 13                     property_type_Apartment -7.965155e-04
## 103                                 bed_linens -7.093798e-04
## 5                            number_of_reviews -4.727484e-04
## 53                           cleaning_fee_True -1.103995e-11
## 10                      host_has_profile_pic_t -5.764853e-12
## 12                    host_identity_verified_t -2.139934e-14
## 54                                 city_Boston  1.171682e-04
## 153                   `room-darkening_shades`  2.054219e-04
## 150                        private_living_room  8.921791e-04
## 49                  cancellation_policy_strict  1.876498e-03
## 144                                       oven  2.240831e-03
## 139                    luggage_dropoff_allowed  4.377478e-03
## 145                    `pack_play/travel_crib`  5.224179e-03
## 99                   bathtub_with_shower_chain  6.396286e-03
## 128                       handheld_shower_head  6.558147e-03
## 85                                     heating  6.999668e-03
## 87                                        pool  7.278795e-03
## 6                         review_scores_rating  7.870704e-03
## 162              `well-lit_path_to_entrance`  8.626830e-03
## 66                            fire_extinguisher  8.932445e-03
## 65                       `family/kid_friendly`  9.545790e-03
## 79                               wide_entryway  1.012782e-02
## 132                           hot_water_kettle  1.073695e-02
## 37                     property_type_Townhouse  1.103180e-02
## 147                                pocket_wifi  1.196511e-02
## 160                             `washer/dryer`  1.204814e-02
## 137                                    lockbox  1.228082e-02
## 130                                    hot_tub  1.312763e-02
## 108                               coffee_maker  1.344555e-02
## 124                               game_console  1.352614e-02
## 107                    cleaning_before_checkout  1.493024e-02
## 11                    host_identity_verified_f  1.546096e-02
## 112                      dishes_and_silverware  1.635588e-02
## 60                            air_conditioning  1.760461e-02
## 81                                   breakfast  1.812348e-02
## 56                                     city_DC  2.072644e-02
## 122            flat_smooth_pathway_to_front_door  2.487535e-02
## 59                          instant_bookable_f  2.550237e-02
## 88                                     shampoo  2.616256e-02
## 97                  babysitter_recommendations  3.008649e-02
## 110                                       crib  3.054049e-02
## 109                             cooking_basics  3.228367e-02
## 117                        ethernet_connection  3.242087e-02
## 114                                       dogs  3.618092e-02
## 64                         elevator_in_building  3.789952e-02
```

```
## 127                                  hair_dryer  3.805364e-02
## 142                                  other_pets  4.031454e-02
## 52                            cleaning_fee_False  4.363631e-02
## 165                                wide_doorway  4.582488e-02
## 116                                       dryer  4.637995e-02
## 29                      property_type_Lighthouse  4.699311e-02
## 90                                          tv  4.778893e-02
## 100                                    bbq_grill  4.787636e-02
## 84                                          gym  4.792148e-02
## 152                  `roll-in_shower_with_chair`  4.818915e-02
## 61                   `buzzer/wireless_intercom`  4.831871e-02
## 161                                  waterfront  4.956793e-02
## 168                                 stair_gates  5.008072e-02
## 41                          property_type_Villa  5.328951e-02
## 17                       property_type_Bungalow  5.674829e-02
## 75                          suitable_for_events  5.899697e-02
## 46                               bed_type_Couch  6.494720e-02
## 62                                     cable_tv  6.537226e-02
## 2                                  accommodates  7.963604e-02
## 22                    property_type_Condominium  8.027202e-02
## 148                            private_bathroom  8.600407e-02
## 126                         ground_floor_access  9.299023e-02
## 39                      property_type_Treehouse  9.421557e-02
## 14                `property_type_Bed & Breakfast`  9.798327e-02
## 33             `property_type_Serviced apartment`  1.004836e-01
## 3                                     bathrooms  1.262735e-01
## 113                                  dishwasher  1.265201e-01
## 31                          property_type_Other  1.279357e-01
## 133                             indoor_fireplace  1.296031e-01
## 82                                      doorman  1.328125e-01
## 9                        host_has_profile_pic_f  1.356890e-01
## 7                                      bedrooms  1.399332e-01
## 30                           property_type_Loft  1.404397e-01
## 50             cancellation_policy_super_strict_30  1.538621e-01
## 15                           property_type_Boat  2.281178e-01
## 36                           property_type_Tipi  2.530297e-01
## 40                `property_type_Vacation home`  2.759620e-01
## 42                           property_type_Yurt  2.896183e-01
## 58                                      city_SF  3.135850e-01
## 16                `property_type_Boutique hotel`  3.207937e-01
## 20               `property_type_Casa particular`  4.111677e-01
## 35                      property_type_Timeshare  4.160817e-01
## 21                         property_type_Castle  4.529490e-01
## 38                          property_type_Train  5.103047e-01
## 24                  `property_type_Earth House`  5.349895e-01
## 43                   `room_type_Entire home/apt`  5.749637e-01
## 32                `property_type_Parking Space`  6.699858e-01
## 51             cancellation_policy_super_strict_60  7.013073e-01
## 1                                   (Intercept)  3.299046e+00
```

```
#print.data.frame(lasso_coef_df)
#write.csv(lasso_coef_df,file="lasso_coef_df.csv")
```

Ridge Regression

Predict responses / Compute MSEs for Ridge

```r
# use the cv.glmnet command to automatically select the best value for the lambda hyper-parameter.
# cv.glmnet returns a cv.glmnet object, a list with all the ingredients of the cross-validated fit
fit.ridge <- cv.glmnet(x.train, y.train, alpha = 0, nfolds = 10)

# computing MSE on the training/test data
yhat.train.ridge <- predict(fit.ridge, x.train)
yhat.test.ridge <- predict(fit.ridge, x.test)

mse.train.ridge <- mean((y.train - yhat.train.ridge)^2)
mse.test.ridge <- mean((y.test - yhat.test.ridge)^2)

mse.train.ridge
```

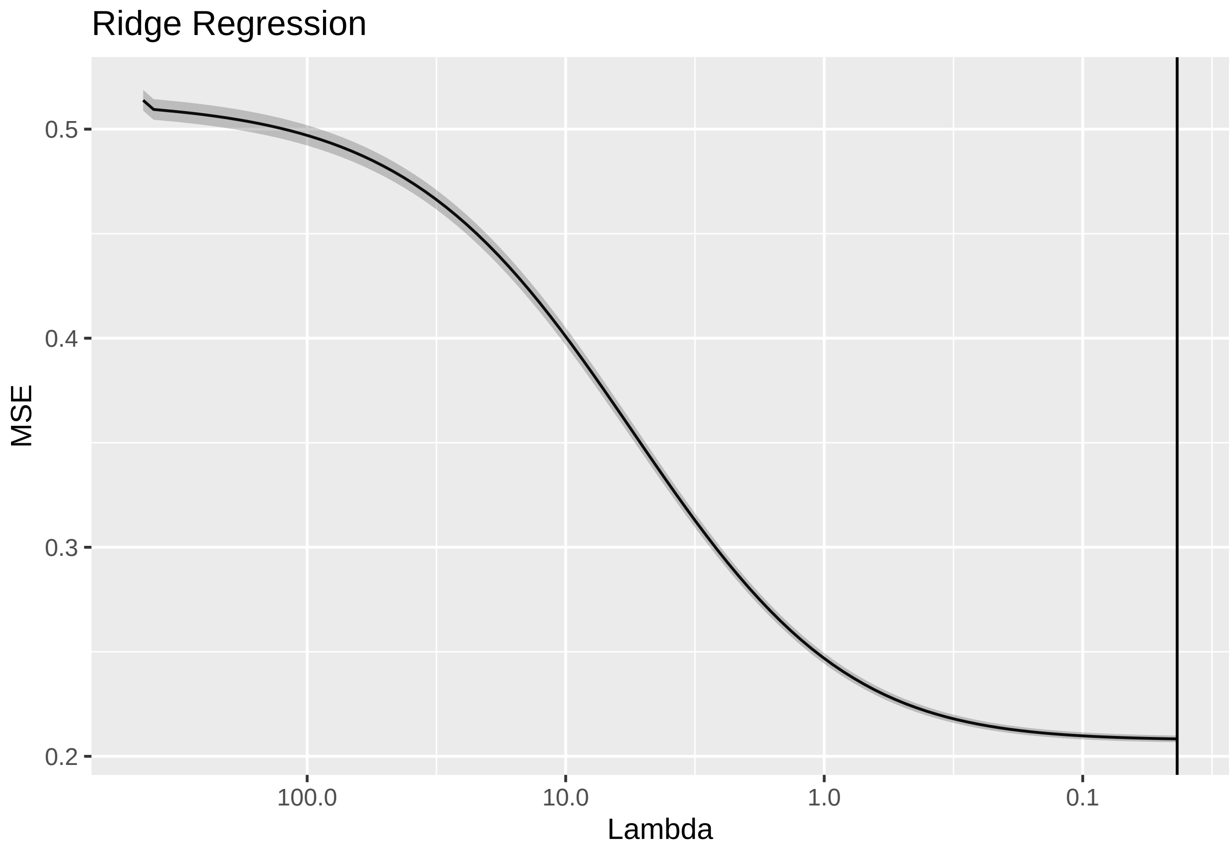
```
## [1] 0.2084374
```

```r
mse.test.ridge
```

```
## [1] 0.2116062
```

Ridge: Lambda and MSE Graph

```r
tidied_cv <- tidy(fit.ridge)
glance_cv <- glance(fit.ridge)
g <- ggplot(tidied_cv, aes(lambda, estimate)) +
  geom_line() + scale_x_continuous(trans=reverselog_trans(10))
g <- g + geom_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = .25)
g <- g +
  geom_vline(xintercept = glance_cv$lambda.min) + ggtitle("Ridge Regression") +
  xlab("Lambda") + ylab("MSE")
g
```



Ridge Coefficients

```
ridge_coef <- coef(fit.ridge, s = fit.ridge$lambda.min)
ridge_coef_df <- data.frame(name = ridge_coef@Dimnames[[1]][ridge_coef@i + 1], coefficient = ridge_coef@
ridge_coef_df <- ridge_coef_df[order(ridge_coef_df[,2]),]

ridge_coef_df
```

```
##                                          name   coefficient
## 28                        property_type_Hostel -0.7096236844
## 48                      `room_type_Shared room` -0.6526095967
## 24                          property_type_Dorm -0.4676235027
## 30                           property_type_Hut -0.3813503720
## 37                          property_type_Tent -0.3328564988
## 62                                city_Chicago -0.2753736262
## 47                     `room_type_Private room` -0.2426156612
## 107                               air_purifier -0.2167968880
## 116                                 beachfront -0.2156975728
## 156                                lake_access -0.1753778529
## 31                       `property_type_In-law` -0.1614403643
## 140                       free_parking_on_street -0.1240364905
## 64                                     city_LA -0.1180671693
## 167                 path_to_entrance_lit_at_night -0.1164966084
## 134                                 ev_charger -0.1025825113
## 138       fixed_grab_bars_for_showers_and_toilet -0.0982219989
## 177                               `ski_in/ski_out` -0.0872895907
```

```
## 160                                        microwave -0.0809484327
## 131                                    doorman_entry -0.0754660264
## 4                                 host_response_rate -0.0741274783
## 18                               property_type_Cabin -0.0731167882
## 10                            host_has_profile_pic_t -0.0705226012
## 91                                 wireless_internet -0.0697848062
## 127                              disabled_parking_spot -0.0688564491
## 142                               garden_or_backyard -0.0659238962
## 83                                     self_check_in -0.0641356230
## 22                              property_type_Chalet -0.0634610101
## 137                                    firm_mattress -0.0593273053
## 98                                           kitchen -0.0580839511
## 49                                   bed_type_Airbed -0.0492408222
## 163                                    outlet_covers -0.0484540182
## 179                                  smoking_allowed -0.0422951642
## 77                          free_parking_on_premises -0.0389192145
## 119                                             cats -0.0381386076
## 26                     `property_type_Guest suite` -0.0371318715
## 190                                    window_guards -0.0357131659
## 189                            wide_hallway_clearance -0.0353124261
## 19                       `property_type_Camper/RV` -0.0346968462
## 181                               table_corner_guards -0.0345880559
## 108                                        baby_bath -0.0338506029
## 115                                  beach_essentials -0.0322246468
## 121                          children_books_and_toys -0.0320058240
## 80                               lock_on_bedroom_door -0.0290716291
## 85                                    smoke_detector -0.0269528551
## 82                         pets_live_on_this_property -0.0265307978
## 112                                          bathtub -0.0257471170
## 173                                     refrigerator -0.0253328241
## 148                                          hangers -0.0244771355
## 55                       cancellation_policy_moderate -0.0233629291
## 136                                 fireplace_guards -0.0227111903
## 101                                            stove -0.0214694537
## 60                                 cleaning_fee_True -0.0194730972
## 8                                               beds -0.0188890818
## 187             `wide_clearance_to_shower_&_toilet` -0.0186270252
## 155                                           keypad -0.0170895936
## 186                              wide_clearance_to_bed -0.0168482723
## 105                          `accessible-height-bed` -0.0164614076
## 78                                   host_greets_you -0.0164176424
## 88  `translation_missing:_en.hosting_amenity_50` -0.0163766858
## 122                               children_dinnerware -0.0162256922
## 178                                       smart_lock -0.0158286585
## 109                                     baby_monitor -0.0157420108
## 68                                instant_bookable_t -0.0148764432
## 51                                    bed_type_Futon -0.0146247805
## 29                               property_type_House -0.0141593313
## 158                           long_term_stays_allowed -0.0141439092
## 76                                     first_aid_kit -0.0141375557
## 171                                 private_entrance -0.0140539216
## 13                           property_type_Apartment -0.0137669306
## 151                                        hot_water -0.0137208027
## 27                          property_type_Guesthouse -0.0133716645
```

```
## 161                                    other -0.0128826263
## 180                      `step-free_access` -0.0122410865
## 87  `translation_missing:_en.hosting_amenity_49` -0.0113983467
## 94                               essentials -0.0100475887
## 79                laptop_friendly_workspace -0.0092233651
## 12                  host_identity_verified_t -0.0080517379
## 89                     wheelchair_accessible -0.0074690029
## 97                                 internet -0.0060549835
## 104              `twentyfour-hour-check-in` -0.0059675716
## 154                                    iron -0.0056968899
## 106              `accessible-height-toilet` -0.0038717544
## 176                             safety_card -0.0037296998
## 117                              bed_linens -0.0036359846
## 72                  carbon_monoxide_detector -0.0031706317
## 168                         patio_or_balcony -0.0011792111
## 5                          number_of_reviews -0.0004750107
## 172                      private_living_room -0.0003054303
## 103                                   washer -0.0002391618
## 40                    property_type_Townhouse  0.0010342624
## 81                              pets_allowed  0.0012538605
## 157                                 lockbox  0.0014624428
## 175                 `room-darkening_shades`  0.0015903749
## 135               extra_blankets_and_pillows  0.0039059327
## 52                 `bed_type_Pull-out Sofa`  0.0040216559
## 84                         single_level_home  0.0040426504
## 159                  luggage_dropoff_allowed  0.0053568134
## 54              cancellation_policy_flexible  0.0060868757
## 149                              high_chair  0.0063175807
## 6                       review_scores_rating  0.0075795570
## 165                 `pack_play/travel_crib`  0.0077988615
## 11                  host_identity_verified_f  0.0080997655
## 53                      `bed_type_Real Bed`  0.0084961615
## 99                                      pool  0.0088803976
## 75                         fire_extinguisher  0.0102209807
## 56                cancellation_policy_strict  0.0112591795
## 145                      hand_or_paper_towel  0.0112609111
## 96                                   heating  0.0117090201
## 147                     handheld_shower_head  0.0118484572
## 185             `well-lit_path_to_entrance`  0.0121461054
## 150                                 hot_tub  0.0133899639
## 152                         hot_water_kettle  0.0139823874
## 69                          air_conditioning  0.0143383327
## 67                        instant_bookable_f  0.0144876008
## 146                               hand_soap  0.0146459944
## 111                              bath_towel  0.0150561373
## 74                     `family/kid_friendly`  0.0154717200
## 141                             game_console  0.0162550570
## 92                                breakfast  0.0166609794
## 118                               body_soap  0.0169822916
## 169                             pocket_wifi  0.0189844413
## 90                              wide_entryway  0.0189910986
## 182                            toilet_paper  0.0192464059
## 59                        cleaning_fee_False  0.0200607826
## 120                          changing_table  0.0200868182
```

```
## 164                                 oven  0.0206006084
## 123              cleaning_before_checkout  0.0211783810
## 124                          coffee_maker  0.0225288583
## 128                 dishes_and_silverware  0.0245918696
## 100                               shampoo  0.0259971400
## 113              bathtub_with_shower_chain  0.0261129101
## 65                               city_NYC  0.0261282637
## 125                        cooking_basics  0.0285091946
## 61                            city_Boston  0.0297923933
## 139      flat_smooth_pathway_to_front_door  0.0335786718
## 110             babysitter_recommendations  0.0337605175
## 126                                  crib  0.0358479035
## 73                   elevator_in_building  0.0359597524
## 133                   ethernet_connection  0.0365389900
## 144                            hair_dryer  0.0371825250
## 130                                  dogs  0.0378512084
## 17                  property_type_Bungalow  0.0435908837
## 183                         `washer/dryer`  0.0442660863
## 132                                 dryer  0.0446556641
## 102                                    tv  0.0467777665
## 44                     property_type_Villa  0.0485231435
## 70             `buzzer/wireless_intercom`  0.0490873006
## 95                                    gym  0.0497156944
## 162                            other_pets  0.0507902734
## 184                            waterfront  0.0515576139
## 188                          wide_doorway  0.0518036649
## 63                                city_DC  0.0543507591
## 114                             bbq_grill  0.0549308258
## 191                           stair_gates  0.0559978163
## 50                         bed_type_Couch  0.0629955720
## 86                     suitable_for_events  0.0635980326
## 71                               cable_tv  0.0646734821
## 2                            accommodates  0.0687960136
## 23              property_type_Condominium  0.0689944074
## 9                   host_has_profile_pic_f  0.0709624634
## 14        `property_type_Bed & Breakfast`  0.0755153666
## 170                      private_bathroom  0.0916220029
## 166              paid_parking_off_premises  0.0996930929
## 143                   ground_floor_access  0.1131279464
## 36     `property_type_Serviced apartment`  0.1147436127
## 129                             dishwasher  0.1156905679
## 34                     property_type_Other  0.1233835437
## 174            `roll-in_shower_with_chair`  0.1234691775
## 3                               bathrooms  0.1236261499
## 93                                 doorman  0.1240677056
## 153                       indoor_fireplace  0.1256646303
## 33                      property_type_Loft  0.1281672495
## 42                 property_type_Treehouse  0.1287814770
## 7                                bedrooms  0.1306180894
## 32                property_type_Lighthouse  0.1690854914
## 57     cancellation_policy_super_strict_30  0.1773585430
## 15                      property_type_Boat  0.2157646876
## 43          `property_type_Vacation home`  0.2918413724
## 45                      property_type_Yurt  0.3016187347
```

```
## 39                          property_type_Tipi  0.3054526586
## 46                   `room_type_Entire home/apt`  0.3149083450
## 16               `property_type_Boutique hotel`  0.3153063349
## 66                                     city_SF  0.3257794557
## 38                     property_type_Timeshare  0.4063196873
## 21                        property_type_Castle  0.4475704926
## 20              `property_type_Casa particular`  0.4648299490
## 25                  `property_type_Earth House`  0.6000140808
## 41                         property_type_Train  0.6030287703
## 35                `property_type_Parking Space`  0.7007091713
## 58           cancellation_policy_super_strict_60  0.7388750167
## 1                                  (Intercept)  3.6584185155
```

Predict responses / Compute MSEs for Elastic Net

```r
fit.elastic <- cv.glmnet(x.train, y.train, alpha = 0.5, nfolds = 10)

yhat.train.elastic <- predict(fit.elastic, x.train)
yhat.test.elastic <- predict(fit.elastic, x.test)

mse.train.elastic <- mean((y.train - yhat.train.elastic)^2)
mse.test.elastic <- mean((y.test - yhat.test.elastic)^2)

mse.train.elastic
```

```
## [1] 0.2074488
```

```r
mse.test.elastic
```

```
## [1] 0.2104755
```

Elastic Net Coefficients

```r
elastic_coef <- coef(fit.elastic, s = fit.elastic$lambda.min)
elastic_coef_df <- data.frame(name = elastic_coef@Dimnames[[1]][elastic_coef@i + 1], coefficient = elas
elastic_elastic_df <- elastic_coef_df[order(elastic_coef_df[,2]),]

elastic_coef_df
```

```
##                                         name   coefficient
## 1                                (Intercept)  3.3786604758
## 2                               accommodates  0.0798125953
## 3                                  bathrooms  0.1264918961
## 4                         host_response_rate -0.0788099133
## 5                          number_of_reviews -0.0004774579
## 6                       review_scores_rating  0.0078809617
## 7                                   bedrooms  0.1402013262
## 8                                       beds -0.0403519651
## 9                       host_has_profile_pic_f  0.1064022114
## 10                      host_has_profile_pic_t -0.0322465127
## 11                    host_identity_verified_f  0.0117520499
## 12                    host_identity_verified_t -0.0040579137
```

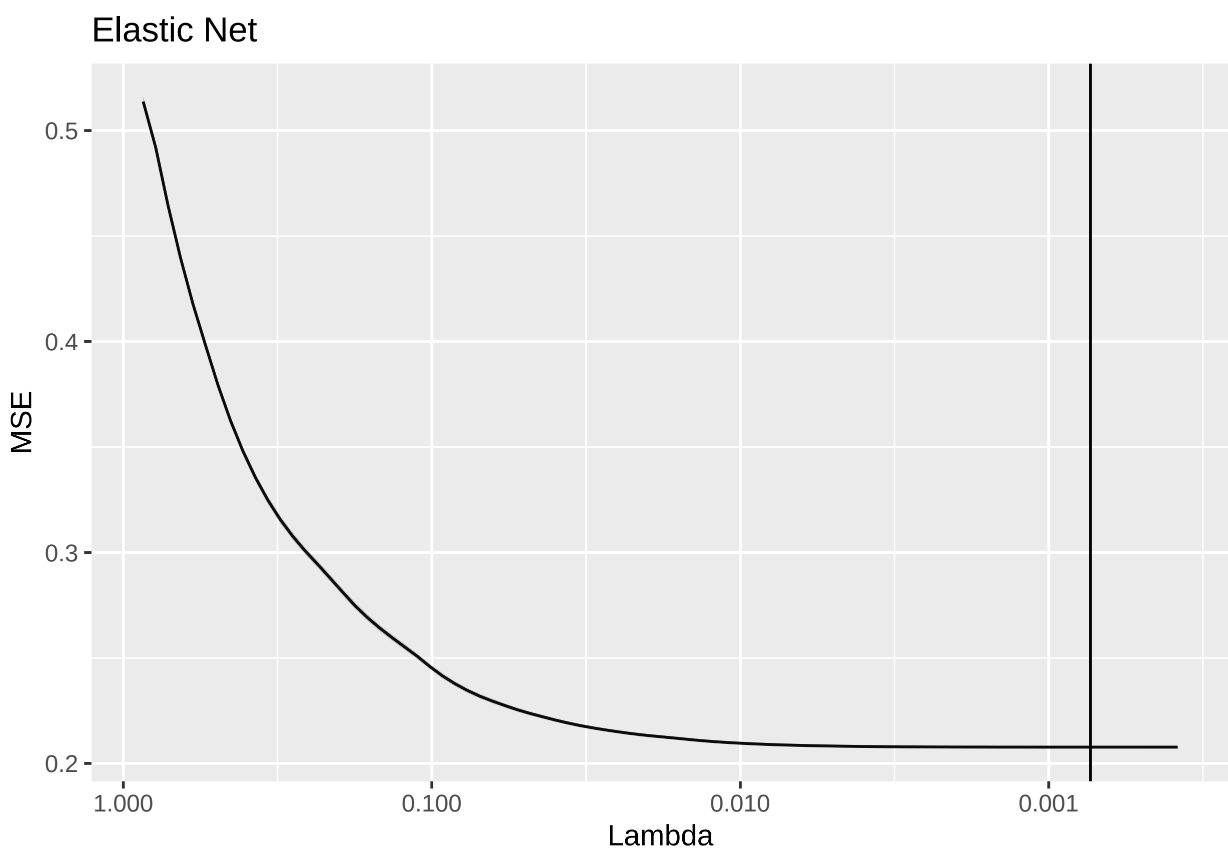
```
## 13                     property_type_Apartment -0.0009917444
## 14            `property_type_Bed & Breakfast`  0.0996773505
## 15                          property_type_Boat  0.2338897131
## 16             `property_type_Boutique hotel`  0.3247024328
## 17                      property_type_Bungalow  0.0596204792
## 18                         property_type_Cabin -0.0472243854
## 19                  `property_type_Camper/RV` -0.0068832926
## 20            `property_type_Casa particular`  0.4493635627
## 21                        property_type_Castle  0.4635139860
## 22                        property_type_Chalet -0.0066265299
## 23                   property_type_Condominium  0.0808613202
## 24                          property_type_Dorm -0.4281456743
## 25                `property_type_Earth House`  0.5546023274
## 26                `property_type_Guest suite` -0.0178352723
## 27                    property_type_Guesthouse  0.0010257657
## 28                        property_type_Hostel -0.6574458805
## 29                           property_type_Hut -0.3484526154
## 30                     `property_type_In-law` -0.1459431383
## 31                    property_type_Lighthouse  0.0820264266
## 32                          property_type_Loft  0.1414115046
## 33                         property_type_Other  0.1296045565
## 34              `property_type_Parking Space`  0.7067046849
## 35         `property_type_Serviced apartment`  0.1097140217
## 36                          property_type_Tent -0.3019698738
## 37                     property_type_Timeshare  0.4198562180
## 38                          property_type_Tipi  0.2799570371
## 39                     property_type_Townhouse  0.0120246569
## 40                         property_type_Train  0.5482931656
## 41                     property_type_Treehouse  0.1116472511
## 42              `property_type_Vacation home`  0.2872186794
## 43                         property_type_Villa  0.0558476355
## 44                          property_type_Yurt  0.3027421677
## 45                `room_type_Entire home/apt`  0.5576970769
## 46                   `room_type_Private room` -0.0168783613
## 47                    `room_type_Shared room` -0.4481439309
## 48                             bed_type_Airbed -0.0457286521
## 49                              bed_type_Couch  0.0679286780
## 50                              bed_type_Futon -0.0106073109
## 51                cancellation_policy_moderate -0.0310447882
## 52                  cancellation_policy_strict  0.0025115395
## 53        cancellation_policy_super_strict_30  0.1581264012
## 54        cancellation_policy_super_strict_60  0.7112636276
## 55                          cleaning_fee_False  0.0272960099
## 56                           cleaning_fee_True -0.0166989568
## 57                                 city_Boston  0.0014387416
## 58                                city_Chicago -0.3244783127
## 59                                     city_DC  0.0213691479
## 60                                     city_LA -0.1535081545
## 61                                     city_SF  0.3143567584
## 62                          instant_bookable_f  0.0156331950
## 63                          instant_bookable_t -0.0101246695
## 64                            air_conditioning  0.0180378575
## 65                `buzzer/wireless_intercom`  0.0487444142
## 66                                    cable_tv  0.0654082043
```

```
## 67                        carbon_monoxide_detector -0.0023689783
## 68                            elevator_in_building  0.0381476616
## 69                           `family/kid_friendly`  0.0099056986
## 70                              fire_extinguisher  0.0095819251
## 71                                  first_aid_kit -0.0119362201
## 72                       free_parking_on_premises -0.0386007801
## 73                                host_greets_you -0.0162958314
## 74                      laptop_friendly_workspace -0.0109217997
## 75                           lock_on_bedroom_door -0.0259017127
## 76                                   pets_allowed -0.0002639540
## 77                     pets_live_on_this_property -0.0227584040
## 78                                 self_check_in -0.0919126673
## 79                              single_level_home  0.0014103461
## 80                                 smoke_detector -0.0289454971
## 81                            suitable_for_events  0.0594434325
## 82  `translation_missing:_en.hosting_amenity_49` -0.0077133668
## 83  `translation_missing:_en.hosting_amenity_50` -0.0155831900
## 84                           wheelchair_accessible -0.0099998737
## 85                                  wide_entryway  0.0131940829
## 86                              wireless_internet -0.0799814534
## 87                                      breakfast  0.0186536705
## 88                                       doorman  0.1339940401
## 89                                     essentials -0.0101697283
## 90                                            gym  0.0479057136
## 91                                        heating  0.0075852989
## 92                                       internet  0.0003869363
## 93                                        kitchen -0.0650831837
## 94                                           pool  0.0077628348
## 95                                        shampoo  0.0265156672
## 96                                          stove -0.0210951511
## 97                                             tv  0.0479856165
## 98                                         washer -0.0063492301
## 99                    `twentyfour-hour-check-in` -0.0061527636
## 100                      `accessible-height-bed` -0.0172415472
## 101                                  air_purifier -0.1820362321
## 102                                     baby_bath -0.0286787526
## 103                                  baby_monitor -0.0070772684
## 104                     babysitter_recommendations  0.0324394763
## 105                                       bathtub -0.0255692139
## 106                      bathtub_with_shower_chain  0.0138349584
## 107                                      bbq_grill  0.0490200544
## 108                               beach_essentials -0.0264967169
## 109                                    beachfront -0.2127509293
## 110                                    bed_linens -0.0010399862
## 111                                          cats -0.0374053897
## 112                                 changing_table  0.0039793609
## 113                       children_books_and_toys -0.0302835245
## 114                            children_dinnerware -0.0135922920
## 115                       cleaning_before_checkout  0.0172488910
## 116                                   coffee_maker  0.0180275472
## 117                                 cooking_basics  0.0368201415
## 118                                           crib  0.0321413533
## 119                           disabled_parking_spot -0.0508907056
## 120                          dishes_and_silverware  0.0257303961
```

```
## 121                               dishwasher  0.1264376766
## 122                                     dogs  0.0381242151
## 123                            doorman_entry -0.0665574025
## 124                                    dryer  0.0506663986
## 125                      ethernet_connection  0.0338473708
## 126                               ev_charger -0.0809063446
## 127                         fireplace_guards -0.0251613877
## 128                            firm_mattress -0.0586734959
## 129     fixed_grab_bars_for_showers_and_toilet -0.0871870497
## 130          flat_smooth_pathway_to_front_door  0.0284764077
## 131                    free_parking_on_street -0.1101558834
## 132                              game_console  0.0152833382
## 133                        garden_or_backyard -0.0686609486
## 134                       ground_floor_access  0.1021159806
## 135                               hair_dryer  0.0389876865
## 136                      handheld_shower_head  0.0088200334
## 137                                  hangers -0.0242133596
## 138                               high_chair  0.0022063772
## 139                                  hot_tub  0.0134911061
## 140                                hot_water -0.0116587554
## 141                         hot_water_kettle  0.0114434139
## 142                          indoor_fireplace  0.1297745332
## 143                                     iron -0.0065718017
## 144                                   keypad -0.0031076492
## 145                              lake_access -0.1678597598
## 146                                  lockbox  0.0142943267
## 147                   long_term_stays_allowed -0.0108776918
## 148                   luggage_dropoff_allowed  0.0050612226
## 149                                microwave -0.0982115428
## 150                                    other -0.0099543106
## 151                               other_pets  0.0461550878
## 152                            outlet_covers -0.0425176223
## 153                                     oven  0.0099274842
## 154                   `pack_play/travel_crib`  0.0064931471
## 155             path_to_entrance_lit_at_night -0.0992232767
## 156                          patio_or_balcony -0.0003786993
## 157                              pocket_wifi  0.0151644888
## 158                          private_bathroom  0.0917938557
## 159                          private_entrance -0.0151527485
## 160                       private_living_room  0.0015031706
## 161                             refrigerator -0.0149670491
## 162               `roll-in_shower_with_chair`  0.0716481311
## 163                  `room-darkening_shades`  0.0016018489
## 164                              safety_card -0.0021745921
## 165                          `ski_in/ski_out` -0.0524883030
## 166                               smart_lock -0.0010376609
## 167                          smoking_allowed -0.0385811200
## 168                        `step-free_access` -0.0084337487
## 169                       table_corner_guards -0.0359440915
## 170                             `washer/dryer`  0.0159315831
## 171                               waterfront  0.0521245648
## 172               `well-lit_path_to_entrance`  0.0097771827
## 173                     wide_clearance_to_bed -0.0131552080
## 174        `wide_clearance_to_shower_&_toilet` -0.0089249993
```

```
## 175                                 wide_doorway  0.0504327265
## 176                        wide_hallway_clearance -0.0293259271
## 177                                 window_guards -0.0318844415
## 178                                   stair_gates  0.0527039409
```

```
tidied_cv <- tidy(fit.elastic)
glance_cv <- glance(fit.elastic)
g <- ggplot(tidied_cv, aes(lambda, estimate)) +
  geom_line() + scale_x_continuous(trans=reverselog_trans(10))
g <- g + geom_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = .25)
g <- g +
  geom_vline(xintercept = glance_cv$lambda.min) + ggtitle("Elastic Net") +
  xlab("Lambda") + ylab("MSE")
g
```




```
# explicitly control the fold that each observation is assigned to via the foldid argument
foldid <- sample(1:10, size = length(y.train), replace = TRUE)
cv.1  <- cv.glmnet(x.train, y.train, foldid = foldid, alpha = 1)
cv.5 <- cv.glmnet(x.train, y.train, foldid = foldid, alpha = 0.5)
cv.0  <- cv.glmnet(x.train, y.train, foldid = foldid, alpha = 0)

par(mfrow = c(2,2))
plot(cv.1, col = "red", xlim=rev(c(-9,1)))
legend("topright", legend = c("Lasso/alpha=1"),
       pch = 19, col = c("red"))
```

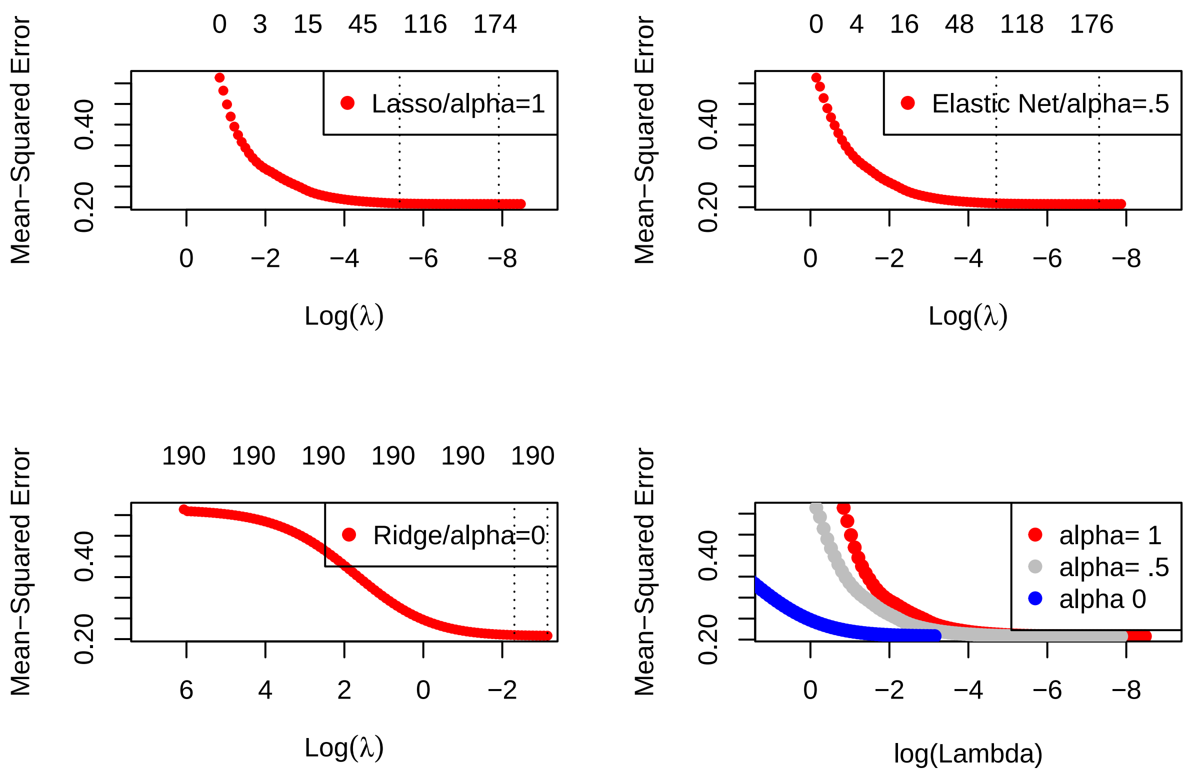
```
plot(cv.5, col = "grey", xlim=rev(c(-9,1)))
legend("topright", legend = c("Elastic Net/alpha=.5"),
       pch = 19, col = c("red"));

plot(cv.0, col = "blue", xlim=rev(c(-3,7)))
legend("topright", legend = c("Ridge/alpha=0"),
       pch = 19, col = c("red"))

plot(log(cv.1$lambda), cv.1$cvm , pch = 19, col = "red",
     xlab = "log(Lambda)", ylab = cv.1$name, xlim=rev(c(-9,1)))

points(log(cv.5$lambda), cv.5$cvm, pch = 19, col = "grey")

points(log(cv.0$lambda) , cv.0$cvm , pch = 19, col = "blue")
legend("topright", legend = c("alpha= 1", "alpha= .5", "alpha 0"),
       pch = 19, col = c("red","grey","blue"))
```




```
# The intervals estimate variance of the loss metric (red dots). They're computed using CV.
# The vertical dotted lines show the locations of labda min and lambda 1se.
# The numbers across the top are the number of nonzero coefficient estimates.
```

Linear Regression

Forward / Backward Selection

```
library(data.table)
library(caret)
```

```
## Loading required package: lattice
```

```
library(olsrr)
```

```
##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers
```

```
dd<-fread('ab_updated.csv')
dd<-dd[,c(2:5,7:8,10:196)]
dd<-na.omit(dd)
dd.sample.size <- 72955
dd <- dd[sample(nrow(dd), dd.sample.size)]
# 70 - 30 split
smp_size <- floor(0.70 * nrow(dd))
train_ind <- sample(seq_len(nrow(dd)), size = smp_size)
train<-dd[train_ind, ]
test<-dd[-train_ind,]

trainControl<-trainControl(method='cv',number=2)
linearmodel2<-train(log_price~.,data=train,method='leapForward',
                    tuneGrid=data.frame(nvmax=1:170),
                    preProcess=NULL,trControl=trainControl)
```

```
## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 21
## linear dependencies found

## Reordering variables and trying again:

## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 19
## linear dependencies found

## Reordering variables and trying again:

## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 17
## linear dependencies found

## Reordering variables and trying again:
```

```
Forward_result<-data.table(Variable=colnames(summary(linearmodel2$finalModel)[["which"]]),
                           Order_Add=(max(colSums(summary(linearmodel2$finalModel)[["which"]]))-colSums

linearmodel3<-train(log_price~.,data=train,method='leapBackward',
                    tuneGrid=data.frame(nvmax=1:170),
                    preProcess=NULL,trControl=trainControl)
```

```
## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 20
## linear dependencies found

## Reordering variables and trying again:

## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 19
## linear dependencies found

## Reordering variables and trying again:

## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 17
## linear dependencies found

## Reordering variables and trying again:
```

```
Backward_result<-data.table(Variable=colnames(summary(linearmodel3$finalModel)[["which"]]),
                            Order_Remove=colSums(summary(linearmodel3$finalModel)[["which"]]))

Forward_result<-Forward_result[order(Forward_result[,2]),]
Backward_result<-Backward_result[order(-Backward_result[,2]),]
print(head(Forward_result,5))
```

```
##                         Variable Order_Add
## 1:                   (Intercept)         0
## 2: `room_type_Entire home/apt`          0
## 3:                      bedrooms         1
## 4:                       city_SF         2
## 5:                  accommodates         3
```

```
print(head(Backward_result,5))
```

```
##                         Variable Order_Remove
## 1:                   (Intercept)          170
## 2: `room_type_Entire home/apt`           170
## 3:                      bedrooms          169
## 4:                  accommodates          168
## 5:    `room_type_Private room`           167
```

```
y_forward=predict(linearmodel2,test)
MSEtest_forward<-colMeans((test[,1]-y_forward)^2)
print(MSEtest_forward)
```

```
## log_price
## 0.2150757
```

```
y_bavkward=predict(linearmodel3,test)
MSEtest_backward<-colMeans((test[,1]-y_bavkward)^2)
print(MSEtest_backward)
```

```
## log_price
## 0.2134275
```

Predict responses / Compute MSEs for Linear Regression

```
y_forward=predict(linearmodel2,test)
MSEtest_forward<-colMeans((test[,1]-y_forward)^2)
print(MSEtest_forward)
```

```
## log_price
## 0.2150757
```

```
y_bavkward=predict(linearmodel3,test)
MSEtest_backward<-colMeans((test[,1]-y_bavkward)^2)
print(MSEtest_backward)
```

```
## log_price
## 0.2134275
```

Boosting

Predict responses / Compute MSEs for Boosting

```
### Generalized Boosted Regression Modeling (GBM)
# Setup
library(caTools)
library(data.table)
library(ggplot2)
library(ggthemes)
library(glmnet)
library(rpart)
library(rpart.plot)
theme_set(theme_bw())
library(scales)
library(gbm)
```

```
## Loaded gbm 2.1.8
```

```
# Splite Test and Train

ab <- fread("ab_updated.csv", stringsAsFactors = T)
ab <- ab[,-1] # remove the ID
ab <- ab[,-5] # remove host_since

set.seed(810)
sample = sample.split(ab$num, SplitRatio = .70)
ab.train = subset(ab , sample == TRUE)
ab.test = subset(ab, sample == FALSE)

ab.train.sample.size <- 5000
ab.train.sample <- ab.train[sample(nrow(ab.train), ab.train.sample.size)]

# Set sample dataset
f1 <- as.formula(log_price ~.)
x1.train.sample <- model.matrix(f1, ab.train.sample)[, -1]
```

```r
y.train <- ab.train$log_price
y.train.sample <- ab.train.sample$log_price

x1.test <- model.matrix(f1, ab.test)[, -1]
y.test <- ab.test$log_price


# We will fit a boosted forest.
fit.btree <- gbm(f1,
                 data = ab.train.sample,
                 distribution = "gaussian",
                 n.trees = 1000,
                 interaction.depth = 2,
                 shrinkage = 0.01,
                 cv.folds = 5)

relative.influence(fit.btree)
```

```
## n.trees not given. Using 467 trees.
```

```
##                              accommodates
##                                3243.43869
##                                 bathrooms
##                                2060.77185
##                        host_response_rate
##                                   0.00000
##                         number_of_reviews
##                                 675.92583
##                      review_scores_rating
##                                  15.97592
##                                   zipcode
##                               22975.42802
##                                  bedrooms
##                                5362.12404
##                                      beds
##                                   0.00000
##                    host_has_profile_pic_f
##                                   0.00000
##                    host_has_profile_pic_t
##                                   0.00000
##                  host_identity_verified_f
##                                   0.00000
##                  host_identity_verified_t
##                                   0.00000
##                   property_type_Apartment
##                                   0.00000
##          `property_type_Bed & Breakfast`
##                                   0.00000
##                        property_type_Boat
##                                   0.00000
##           `property_type_Boutique hotel`
##                                   0.00000
##                    property_type_Bungalow
```

```
##                                   0.00000
##                        property_type_Cabin
##                                   0.00000
##               `property_type_Camper/RV`
##                                   0.00000
##         `property_type_Casa particular`
##                                   0.00000
##                       property_type_Castle
##                                   0.00000
##                         property_type_Cave
##                                   0.00000
##                       property_type_Chalet
##                                   0.00000
##                  property_type_Condominium
##                                   0.00000
##                         property_type_Dorm
##                                   0.00000
##             `property_type_Earth House`
##                                   0.00000
##             `property_type_Guest suite`
##                                   0.00000
##                   property_type_Guesthouse
##                                   0.00000
##                       property_type_Hostel
##                                   0.00000
##                        property_type_House
##                                   0.00000
##                          property_type_Hut
##                                   0.00000
##                  `property_type_In-law`
##                                   0.00000
##                       property_type_Island
##                                   0.00000
##                   property_type_Lighthouse
##                                   0.00000
##                         property_type_Loft
##                                   0.00000
##                        property_type_Other
##                                   0.00000
##           `property_type_Parking Space`
##                                   0.00000
##     `property_type_Serviced apartment`
##                                   0.00000
##                         property_type_Tent
##                                   0.00000
##                    property_type_Timeshare
##                                   0.00000
##                         property_type_Tipi
##                                   0.00000
##                    property_type_Townhouse
##                                   0.00000
##                        property_type_Train
##                                   0.00000
##                    property_type_Treehouse
```

```
##                                    0.00000
##             `property_type_Vacation home`
##                                    0.00000
##                       property_type_Villa
##                                    0.00000
##                        property_type_Yurt
##                                    0.00000
##               `room_type_Entire home/apt`
##                                16931.06678
##                  `room_type_Private room`
##                                    0.00000
##                   `room_type_Shared room`
##                                  337.60930
##                            bed_type_Airbed
##                                    0.00000
##                             bed_type_Couch
##                                    0.00000
##                             bed_type_Futon
##                                    0.00000
##                  `bed_type_Pull-out Sofa`
##                                    0.00000
##                      `bed_type_Real Bed`
##                                    0.00000
##               cancellation_policy_flexible
##                                    0.00000
##               cancellation_policy_moderate
##                                    0.00000
##                 cancellation_policy_strict
##                                    0.00000
##       cancellation_policy_super_strict_30
##                                    0.00000
##       cancellation_policy_super_strict_60
##                                    0.00000
##                         cleaning_fee_False
##                                    0.00000
##                          cleaning_fee_True
##                                    0.00000
##                                city_Boston
##                                    0.00000
##                               city_Chicago
##                                    0.00000
##                                    city_DC
##                                    0.00000
##                                    city_LA
##                                    0.00000
##                                   city_NYC
##                                    0.00000
##                                    city_SF
##                                    0.00000
##                         instant_bookable_f
##                                    0.00000
##                         instant_bookable_t
##                                    0.00000
##                           air_conditioning
```

```
##                                         0.00000
##                   `buzzer/wireless_intercom`
##                                         0.00000
##                                        cable_tv
##                                         0.00000
##                      carbon_monoxide_detector
##                                         0.00000
##                          elevator_in_building
##                                         0.00000
##                         `family/kid_friendly`
##                                         0.00000
##                             fire_extinguisher
##                                         0.00000
##                                 first_aid_kit
##                                         0.00000
##                      free_parking_on_premises
##                                         0.00000
##                               host_greets_you
##                                         0.00000
##                     laptop_friendly_workspace
##                                         0.00000
##                          lock_on_bedroom_door
##                                         0.00000
##                                  pets_allowed
##                                         0.00000
##                    pets_live_on_this_property
##                                         0.00000
##                                 self_check_in
##                                         0.00000
##                             single_level_home
##                                         0.00000
##                                smoke_detector
##                                         0.00000
##                           suitable_for_events
##                                         0.00000
## `translation_missing:_en.hosting_amenity_49`
##                                         0.00000
## `translation_missing:_en.hosting_amenity_50`
##                                         0.00000
##                         wheelchair_accessible
##                                         0.00000
##                                 wide_entryway
##                                         0.00000
##                             wireless_internet
##                                         0.00000
##                                     breakfast
##                                         0.00000
##                                       doorman
##                                         0.00000
##                                    essentials
##                                         0.00000
##                                           gym
##                                         0.00000
##                                       heating
```

```
##                                  0.00000
##                                 internet
##                                  0.00000
##                                  kitchen
##                                  0.00000
##                                     pool
##                                  0.00000
##                                  shampoo
##                                  0.00000
##                                    stove
##                                  0.00000
##                                       tv
##                                  0.00000
##                                   washer
##                                  0.00000
##              `twentyfour-hour-check-in`
##                                  0.00000
##                 `accessible-height-bed`
##                                  0.00000
##              `accessible-height-toilet`
##                                  0.00000
##                             air_purifier
##                                  0.00000
##                                baby_bath
##                                  0.00000
##                             baby_monitor
##                                  0.00000
##                babysitter_recommendations
##                                  0.00000
##                               bath_towel
##                                  0.00000
##                                  bathtub
##                                  0.00000
##                bathtub_with_shower_chain
##                                  0.00000
##                                 bbq_grill
##                                  0.00000
##                         beach_essentials
##                                  0.00000
##                               beachfront
##                                  0.00000
##                               bed_linens
##                                  0.00000
##                                body_soap
##                                  0.00000
##                                     cats
##                                  0.00000
##                           changing_table
##                                  0.00000
##                  children_books_and_toys
##                                  0.00000
##                      children_dinnerware
##                                  0.00000
##                 cleaning_before_checkout
```

```
##                                    0.00000
##                                coffee_maker
##                                    0.00000
##                              cooking_basics
##                                    0.00000
##                                        crib
##                                    0.00000
##                       disabled_parking_spot
##                                    0.00000
##                       dishes_and_silverware
##                                    0.00000
##                                  dishwasher
##                                    0.00000
##                                        dogs
##                                    0.00000
##                               doorman_entry
##                                    0.00000
##                                       dryer
##                                    0.00000
##                         ethernet_connection
##                                    0.00000
##                                  ev_charger
##                                    0.00000
##                  extra_blankets_and_pillows
##                                    0.00000
##                            fireplace_guards
##                                    0.00000
##                               firm_mattress
##                                    0.00000
##      fixed_grab_bars_for_showers_and_toilet
##                                    0.00000
##           flat_smooth_pathway_to_front_door
##                                    0.00000
##                      free_parking_on_street
##                                    0.00000
##                                game_console
##                                    0.00000
##                          garden_or_backyard
##                                    0.00000
##                         ground_floor_access
##                                    0.00000
##                                  hair_dryer
##                                    0.00000
##                         hand_or_paper_towel
##                                    0.00000
##                                   hand_soap
##                                    0.00000
##                        handheld_shower_head
##                                    0.00000
##                                     hangers
##                                    0.00000
##                                  high_chair
##                                    0.00000
##                                     hot_tub
```

```
##                                  0.00000
##                                hot_water
##                                  0.00000
##                         hot_water_kettle
##                                  0.00000
##                         indoor_fireplace
##                                 10.66594
##                                     iron
##                                  0.00000
##                                   keypad
##                                  0.00000
##                              lake_access
##                                  0.00000
##                                  lockbox
##                                  0.00000
##                  long_term_stays_allowed
##                                  0.00000
##                  luggage_dropoff_allowed
##                                  0.00000
##                                microwave
##                                  0.00000
##                                    other
##                                  0.00000
##                               other_pets
##                                  0.00000
##                            outlet_covers
##                                  0.00000
##                                     oven
##                                  0.00000
##                  `pack_play/travel_crib`
##                                  0.00000
##                paid_parking_off_premises
##                                  0.00000
##            path_to_entrance_lit_at_night
##                                  0.00000
##                         patio_or_balcony
##                                  0.00000
##                              pocket_wifi
##                                  0.00000
##                         private_bathroom
##                                  0.00000
##                         private_entrance
##                                  0.00000
##                      private_living_room
##                                  0.00000
##                             refrigerator
##                                  0.00000
##              `roll-in_shower_with_chair`
##                                  0.00000
##                  `room-darkening_shades`
##                                  0.00000
##                              safety_card
##                                  0.00000
##                         `ski_in/ski_out`
```

```
##                                              0.00000
##                                           smart_lock
##                                              0.00000
##                                      smoking_allowed
##                                              0.00000
##                                  `step-free_access`
##                                              0.00000
##                                  table_corner_guards
##                                              0.00000
##                                         toilet_paper
##                                              0.00000
##                                      `washer/dryer`
##                                              0.00000
##                                           waterfront
##                                              0.00000
##                         `well-lit_path_to_entrance`
##                                              0.00000
##                                wide_clearance_to_bed
##                                              0.00000
##                 `wide_clearance_to_shower_&_toilet`
##                                              0.00000
##                                         wide_doorway
##                                              0.00000
##                              wide_hallway_clearance
##                                              0.00000
##                                        window_guards
##                                              0.00000
##                                          stair_gates
##                                              0.00000
```

```r
# Train MSE
yhat.btree <- predict(fit.btree, ab.train.sample, n.trees = 467)
mse.btree <- mean((yhat.btree - y.train.sample) ^ 2)

print(mse.btree)
```

```
## [1] 0.164818
```

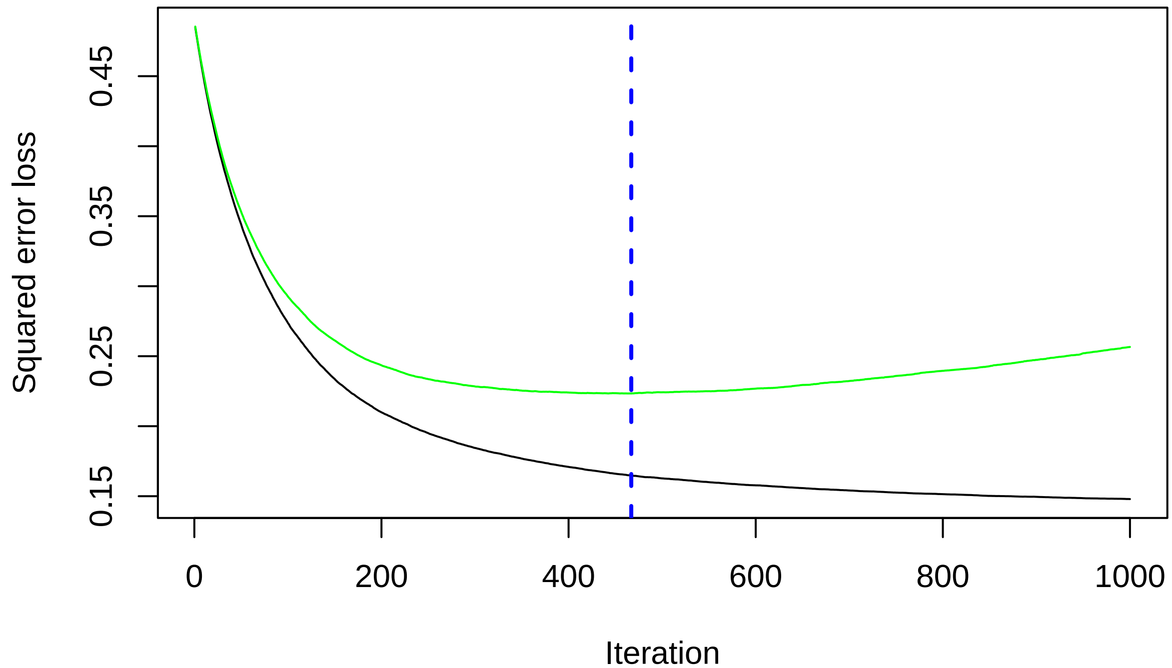
```r
# Test MSE
yhat.test.btree <- predict(fit.btree, ab.test,n.trees = 467)
test.mse.btree <- mean((yhat.test.btree - y.test) ^ 2)
print(test.mse.btree)
```

```
## [1] 0.2110714
```

```r
# get MSE and compute RMSE
min_MSE <- which.min(fit.btree$cv.error)
sqrt(fit.btree$cv.error[min_MSE])
```

```
## [1] 0.4726127
```

```r
# plot loss function as a result of n trees added to the ensemble
gbm.perf(fit.btree, method = "cv")
```




```
## [1] 467
```

Random Forest

Predict responses / Compute MSEs for Random Forest

```r
library(data.table)
library(ggplot2)
library(ggthemes)
library(glmnet)
library(caTools)
theme_set(theme_bw())

#read the file.
dd <- fread("/Users/jiazhijia/Desktop/810/ab.csv")
dd<-dd[,c(1:73,693:815)]#some columns are rather useless to our model and would also slow down the spee

#some column names are not readable for R.
names(dd)<-gsub(' ','_',names(dd))
names(dd)<-gsub('/','_',names(dd))
names(dd)<-gsub('-','_',names(dd))
names(dd)<-gsub('&','',names(dd))
```

```r
names(dd)<-gsub('24_','',names(dd))
names(dd)<-gsub(',','',names(dd))
names(dd)<-gsub("'",'',names(dd))
names(dd)<-gsub(':','',names(dd))

#test and train
set.seed(810)
sample = sample.split(dd$num, SplitRatio = .70)
train= subset(dd , sample == TRUE)
test= subset(dd, sample == FALSE)

#dataset too large! split into small data combine them in the end.
sample_train = sample.split(train$num, SplitRatio = .60)
train1= subset(train, sample_train == TRUE)
train2= subset(train, sample_train == FALSE)

#build model,setting ntrees=500,try 14 variable in each split.
library(randomForest)
```

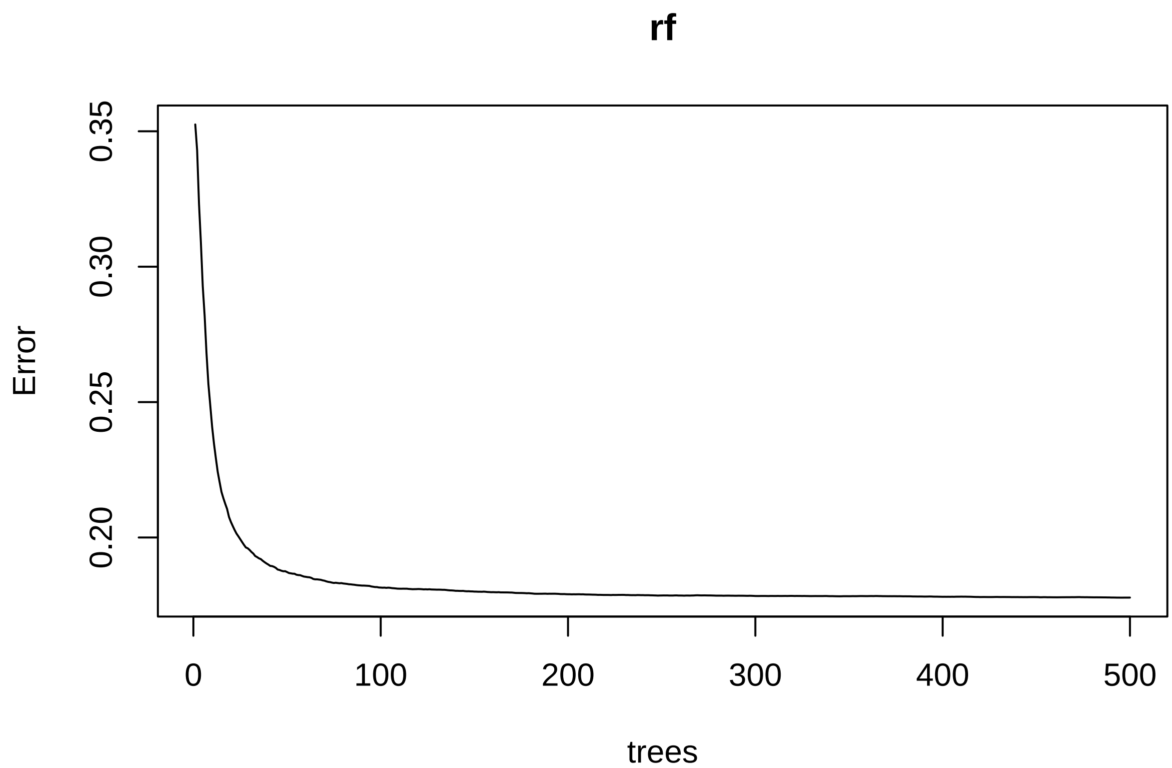
```
## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin
```

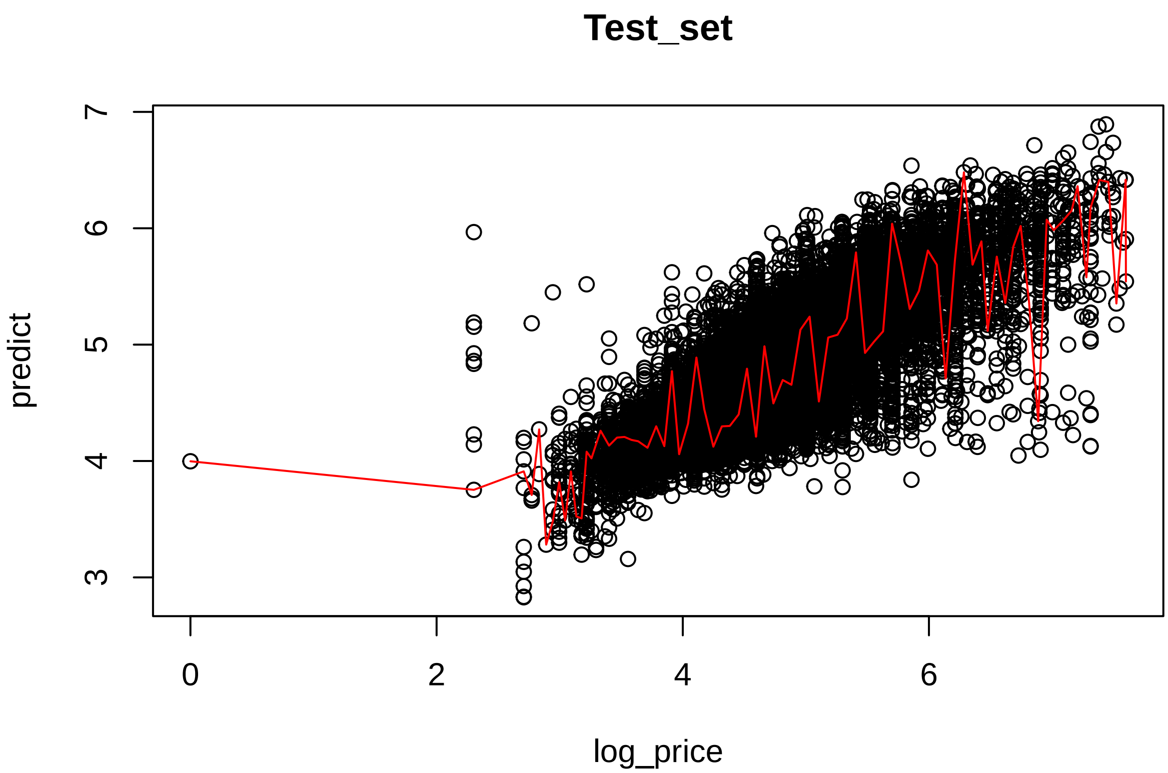
```r
rf <- randomForest(log_price~., data=train1, proximity=TRUE,ntree=500,mtry=sqrt(ncol(train1)),na.action
plot(rf)
```




```
#see the performence on test data.
pred=predict(rf,test)
y_test<-test$log_price
mse_test=mean((y_test-pred)^ 2,na.rm=TRUE)
print(mse_test)
```

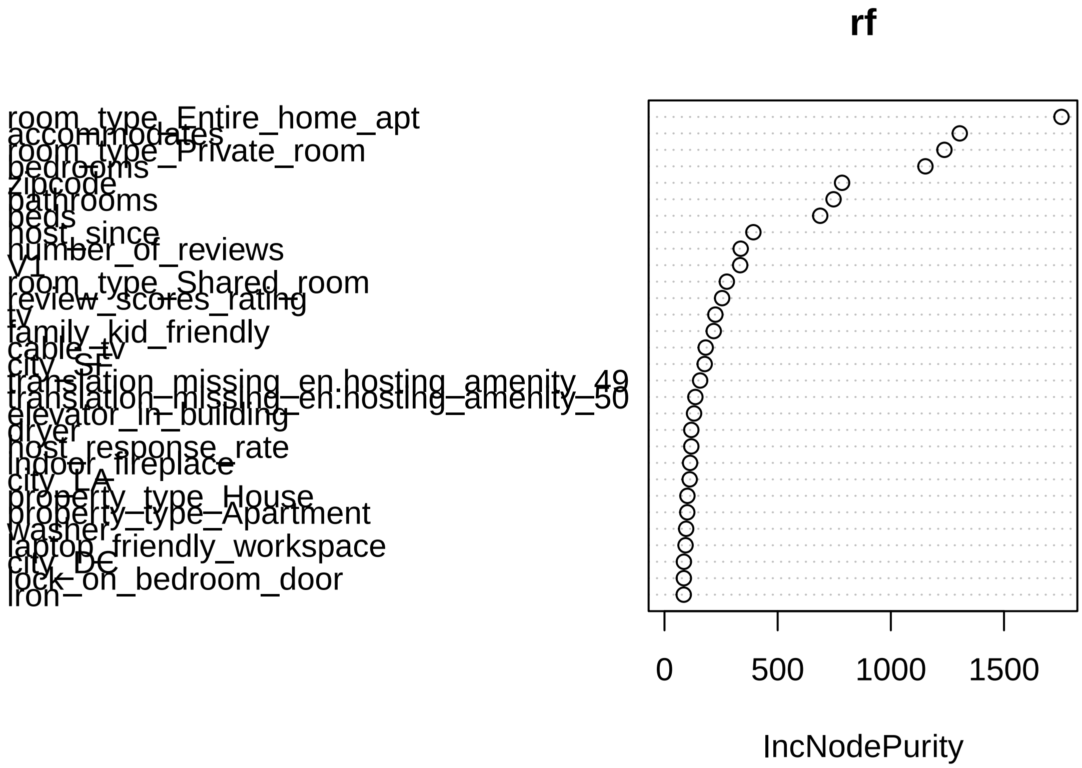
```
## [1] 0.1823137
```

```
#using plot to see the outcome.
plot(test$log_price,pred,main='Test_set',xlab='log_price',ylab='predict') +lines(lowess(test$log_price,p
```




```
## integer(0)
```

```
varImpPlot(rf)
```


rf
room_type_Entire_home_apt
accommodates
room_type_Private_room
bedrooms
zipcode
bathrooms
beds
host_since
number_of_reviews
V1
room_type_Shared_room
review_scores_rating
tv
family_kid_friendly
cable_tv
city_SF
translation_missing_en.hosting_amenity_49
translation_missing_en.hosting_amenity_50
elevator_in_building
dryer
host_response_rate
indoor_fireplace
city_LA
property_type_House
property_type_Apartment
washer
laptop_friendly_workspace
city_DC
lock_on_bedroom_door
iron
0
500
1000
1500
IncNodePurity