Final Report

Group 1: Jiawen Chen, Brooke Felsheim, Elena Kharitonova, Jairui Tang, and Xinjie Qian

4/29/2022

Introduction

```
library(usethis)
library(devtools)
load_all("package/bikeSharing/")
```

```
## i Loading bikeSharing
```

```
set.seed(1)
```

```
str(london)
```

```
## 'data.frame':    2185 obs. of  14 variables:
##  $ Date        : chr  "01-01" "01-01" "01-01" "01-01" ...
##  $ Hour_chunks : Factor w/ 3 levels "[0,8)","[8,16)",..: 1 1 2 2 3 3 1 1 2 2 ...
##  $ Day         : num  1 1 1 1 1 1 2 2 2 2 ...
##  $ Is_weekend  : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 1 2 1 2 ...
##  $ Is_holiday  : Factor w/ 2 levels "0","1": 2 1 2 1 2 1 2 1 2 1 ...
##  $ Season      : Factor w/ 4 levels "Spring","Summer",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Min_temp    : num  3 5 3 5 3 5 1 9 1 9 ...
##  $ Max_temp    : num  9 10 9 10 9 10 6 11.5 6 11.5 ...
##  $ Min_humidity: num  76 81 76 81 76 81 71 82 71 82 ...
##  $ Max_humidity: num  87 93 87 93 87 93 93 94 93 94 ...
##  $ Year        : chr  "Year 1" "Year 2" "Year 1" "Year 2" ...
##  $ Wind_speed  : num  2.48 3.65 4.83 4.08 6.63 ...
##  $ Rain_or_snow: Factor w/ 2 levels "0","1": 1 2 2 2 2 2 1 2 1 2 ...
##  $ Bike_count  : int  2715 2962 4460 2450 2622 1009 438 475 7756 4263 ...
```

```
str(dc)
```

```
## 'data.frame':    2187 obs. of  14 variables:
##  $ Date        : chr  "01-01" "01-01" "01-01" "01-01" ...
##  $ Hour_chunks : Factor w/ 3 levels "[0,8)","[8,16)",..: 1 1 2 2 3 3 1 1 2 2 ...
##  $ Day         : num  1 1 1 1 1 1 2 2 2 2 ...
##  $ Is_weekend  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 1 2 1 2 ...
##  $ Is_holiday  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 2 1 ...
##  $ Season      : Factor w/ 4 levels "Spring","Summer",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Min_temp    : num  1.4 4.22 1.4 4.22 1.4 4.22 2.34 2.34 2.34 2.34 ...
##  $ Max_temp    : num  13.6 14.6 13.6 14.6 13.6 ...
##  $ Min_humidity: num  72 48 72 48 72 48 32 39 32 39 ...
##  $ Max_humidity: num  94 93 94 93 94 93 45 100 45 100 ...
##  $ Year        : chr  "Year 1" "Year 2" "Year 1" "Year 2" ...
##  $ Wind_speed  : num  0.208 1.458 3.958 3.75 4.791 ...
##  $ Rain_or_snow: Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 2 1 2 ...
##  $ Bike_count  : int  108 290 508 1218 369 786 96 55 1102 452 ...
```

```
str(seoul)
```

```
## 'data.frame':    1059 obs. of  13 variables:
##  $ Date        : chr  "01-01" "01-01" "01-01" "01-02" ...
##  $ Hour_chunks : Factor w/ 3 levels "[0,8)","[8,16)",..: 1 2 3 1 2 3 1 2 3 1 ...
##  $ Day         : num  1 1 1 2 2 2 3 3 3 4 ...
##  $ Is_weekend  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Is_holiday  : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
##  $ Season      : Factor w/ 4 levels "Spring","Summer",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Min_temp    : num  -5 -5 -5 -3.8 -3.8 -3.8 -7 -7 -7 -8.6 ...
##  $ Max_temp    : num  3.7 3.7 3.7 1.7 1.7 1.7 -0.4 -0.4 -0.4 -0.8 ...
##  $ Min_humidity: int  20 20 20 20 20 20 29 29 29 31 ...
##  $ Max_humidity: int  56 56 56 71 71 71 54 54 54 57 ...
##  $ Wind_speed  : num  0.9 1.85 1.61 0.65 2.26 ...
##  $ Rain_or_snow: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Bike_count  : int  1002 1633 1655 938 2610 2898 1022 2624 2866 1015 ...
```

```
london_train <- london[london$Year == "Year 1",]
london_test <- london[london$Year == "Year 2",]
```

Methods

Negative Binomial Generalized Linear Mixed Model

Random Forest

Results

Negative Binomial Generalized Linear Mixed Model

```
glmm_fit <- MCEM_algorithm( beta_initial = c(8.3, 1.5, 1.5, -0.25, -0.50, 0,
                                             0, -0.25, 0, 0, 0, 0, 0, -0.25),
                         theta_initial = 10,
                         s2gamma_initial = 0.2,
                         M = 1000,
                         burn.in = 200,
                         tol = 10^-4,
                         maxit = 100,
                         data = london_train
                         )
```

```
str(glmm_fit)
```

```
## List of 7
##  $ beta     : num [1:14] 8.353 1.534 1.415 -0.337 -0.393 ...
##  $ s2gamma  : num 0.0296
##  $ theta    : num 18.4
##  $ eps      : num 5.15e-05
##  $ qfunction: num -9520
##  $ day_ranef: num [1:365] 0.0653 -0.398 -0.5165 -0.2612 -0.0374 ...
##  $ iter     : num 23
```

```
glmm_model_fit(glmm_fit, london_train, scale_to_london_mean = "no")
```

```
##       RMSE      MAE        R2
## 1 1886.267 1291.106 0.8831618
```

```
glmm_model_fit(glmm_fit, london_test, scale_to_london_mean = "no")
```

```
##       RMSE      MAE        R2
## 1 2491.293 1647.064 0.8142036
```

```
glmm_model_fit(glmm_fit, dc, scale_to_london_mean = "yes")
```

```
##      RMSE     MAE       R2
## 1 845.741 605.217 0.521788
```

```
glmm_model_fit(glmm_fit, seoul, scale_to_london_mean = "yes")
```

```
##       RMSE      MAE        R2
## 1 3519.413 2719.021 0.4999935
```

Random Forest

```
rf_fit <- train_random_forest(data = london_train)
```

```
## Loading required package: ggplot2
```

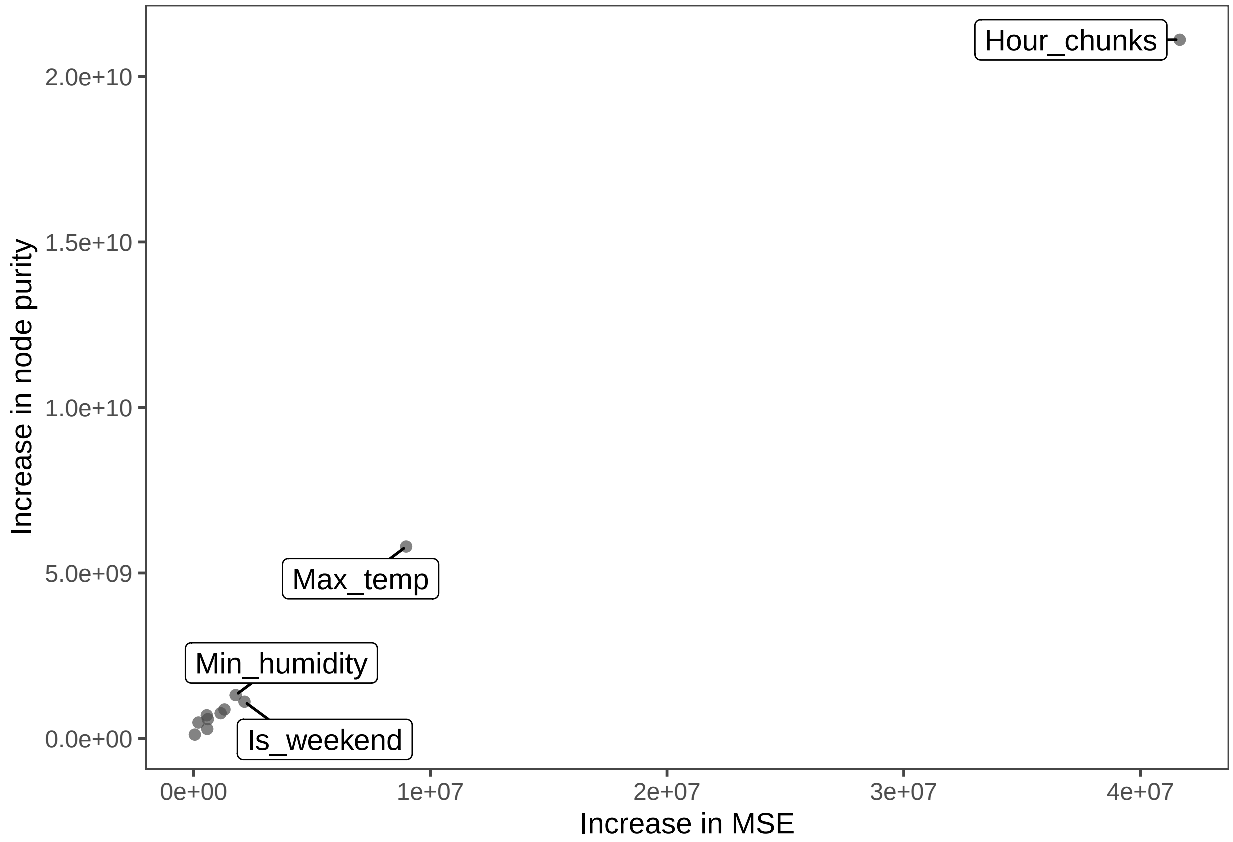
```
## Loading required package: lattice
```

```
rf_fit
```

```
## Random Forest
##
## 1095 samples
##   11 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 876, 878, 875, 875, 876
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##    2    2262.068  0.8841247  1721.956
##    6    1797.963  0.8980543  1174.833
##   11    1789.815  0.8964189  1167.026
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 11.
```

```
plot_rf_importance(london_train)
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
rf_model_fit(rf_fit, london_train, scale_to_london_mean = "no")
```

```
##       RMSE      MAE        R2
## 1 714.5723 445.2132 0.9838003
```

```
rf_model_fit(rf_fit, london_test, scale_to_london_mean = "no")
```

```
##      RMSE      MAE        R2
## 1 1780.33 1172.378 0.9063308
```

```
rf_model_fit(rf_fit, dc, scale_to_london_mean = "yes")
```

```
##       RMSE      MAE       R2
## 1 737.5336 524.7186 0.641438
```

```
rf_model_fit(rf_fit, seoul, scale_to_london_mean = "yes")
```

```
##       RMSE      MAE       R2
## 1 3465.028 2630.914 0.544965
```

Discussion