# Demensionality Reduction

Venukanan Subenthiran

2022-11-17

# Feature Selection Using Forward Selection for Numeric Variables

```
#install.packages("olsrr")
library(olsrr)
```

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

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

```
lm_model_numeric <- lm(is_fraud~trans_date_trans_time+cc_num+amt+zip+lat+long+city_pop+dob+unix_
time+merch_lat+merch_long, data = fraudTotal.db)

FWD_fit_num <- ols_step_forward_p(lm_model_numeric, penter=0.05)

FWD_fit_num
```

```
##
##                                 Selection Summary
## ------------------------------------------------------------------------------------------
##         Variable                            Adj.
## Step           Entered           R-Square    R-Square      C(p)           AIC          RMSE
## ------------------------------------------------------------------------------------------
##    1    amt                        0.0438      0.0438    894.4275    -4574143.0543    0.0704
##    2    dob                        0.0440      0.0440    550.3516    -4574486.9962    0.0704
##    3    unix_time                  0.0442      0.0441    236.4201    -4574800.8618    0.0704
##    4    trans_date_trans_time      0.0443      0.0443     34.8863    -4575002.3817    0.0704
##    5    zip                        0.0443      0.0443     22.6123    -4575014.6555    0.0704
##    6    merch_long                 0.0443      0.0443     14.5038    -4575022.7640    0.0704
##    7    lat                        0.0443      0.0443     11.9669    -4575025.3009    0.0704
## ------------------------------------------------------------------------------------------
```

# Feature Selection Using Forward Selection for Categorical Variables

# Dementionality Reduction Using Random Forest Method of Categorical Variables

## Chi Square test of merchant and category

```
chisq.test(fraudTotal.db$merchant, fraudTotal.db$category)
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$merchant and fraudTotal.db$category
## X-squared = 23821955, df = 8996, p-value < 2.2e-16
```

```
# Category variable to be removed
```

## Chi Square test of first and last

```
chisq.test(fraudTotal.db$first, fraudTotal.db$last)
```

```
## Warning in chisq.test(fraudTotal.db$first, fraudTotal.db$last): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$first and fraudTotal.db$last
## X-squared = 331405447, df = 171690, p-value < 2.2e-16
```

```
# Last variable to be removed
```

## Chi Square test of first and gender

```
chisq.test(fraudTotal.db$first, fraudTotal.db$gender)
```

```
## Warning in chisq.test(fraudTotal.db$first, fraudTotal.db$gender): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$first and fraudTotal.db$gender
## X-squared = 1852394, df = 354, p-value < 2.2e-16
```

## Chi Square test of street and city

```
chisq.test(fraudTotal.db$street, fraudTotal.db$city)
```

```
## Warning in chisq.test(fraudTotal.db$street, fraudTotal.db$city): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$street and fraudTotal.db$city
## X-squared = 1676416570, df = 903190, p-value < 2.2e-16
```

## Chi Square test of city and state

```
chisq.test(fraudTotal.db$city, fraudTotal.db$state)
```

```
## Warning in chisq.test(fraudTotal.db$city, fraudTotal.db$state): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$city and fraudTotal.db$state
## X-squared = 87829535, df = 45250, p-value < 2.2e-16
```

## Chi Square test of category and job

```
chisq.test(fraudTotal.db$category, fraudTotal.db$job)
```

```
## Warning in chisq.test(fraudTotal.db$category, fraudTotal.db$job): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$category and fraudTotal.db$job
## X-squared = 62987, df = 6448, p-value < 2.2e-16
```

# Chi Square test of gender and job

```
chisq.test(fraudTotal.db$gender, fraudTotal.db$job)
```

```
## Warning in chisq.test(fraudTotal.db$gender, fraudTotal.db$job): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$gender and fraudTotal.db$job
## X-squared = 1034523, df = 496, p-value < 2.2e-16
```

# Chi Square test of category and trans_num

```
chisq.test(fraudTotal.db$category, fraudTotal.db$trans_num)
```

```
## Warning in chisq.test(fraudTotal.db$category, fraudTotal.db$trans_num): Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  fraudTotal.db$category and fraudTotal.db$trans_num
## X-squared = 24081122, df = 24081109, p-value = 0.4992
```