

# Machine Learning

Pakakorn Kaeoluan

2023-06-08

## 1. Data preparation

To prepare train and test data with 80% and 20% of raw data

```
library(caret)

## Loading required package: ggplot2
## Loading required package: lattice

traintestsplit <- function(data){
  set.seed(42)
  n <- nrow(data)
  id <- sample(n,size =0.8*n)
  train_data <-data[id,]
  test_data <- data[-id,]
  return(list(train_data,test_data))
}

split_data <- traintestsplit(mtcars)
```

## 2. Calculation

### 2.1 Model : linear regression

```
# train model
lm_model <- train(mpg~hp,
                  data = split_data[[1]],
                  method="lm")

# test model
p <- predict(lm_model,newdata = split_data[[2]])

error<-split_data[[2]]$mpg-p
RMSE <- sqrt(mean(error**2))
```

### 2.2 Model : linear regression

```
#Prep data
mtcars$am <- factor(mtcars$am,
                   levels = c(0,1),
                   labels = c("auto","manual"))
```

```

split_data <- traintestsplit(mtcars)

# train model
glm_model <- train(am~mpg,
                   data = split_data[[1]],
                   method="glm") #generalized linear model

## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

# test model
p <- predict(glm_model,newdata = split_data[[2]]) #predicted value

acc <- mean(p == split_data[[2]]$am)
acc

## [1] 0.5714286
p

## [1] auto   manual manual manual auto   auto   auto
## Levels: auto manual
glm_model

## Generalized Linear Model
##
## 25 samples
## 1 predictor
## 2 classes: 'auto', 'manual'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 25, 25, 25, 25, 25, 25, ...
## Resampling results:
##
##   Accuracy   Kappa
## 0.7884863 0.5532344

```