

firstAttempt

isilendil

2014.08.16.

load train set

```
train <- read.csv("~/Kaggle/Titanic:MachineLearningfromDisaster/titanic/train.csv")
```

load test set

```
test <- read.csv("~/Kaggle/Titanic:MachineLearningfromDisaster/titanic/test.csv")
```

take a look at the data set

```
str(train)
```

```
## 'data.frame':      891 obs. of  12 variables:
##  $ PassengerId: int   1  2  3  4  5  6  7  8  9 10 ...
##  $ Survived   : int   0  1  1  1  0  0  0  0  1  1 ...
##  $ Pclass     : int   3  1  3  1  3  3  1  3  3  2 ...
##  $ Name       : Factor w/ 891 levels "Abbing, Mr. Anthony",...: 109 191 358 277 16 559 520 629 416 58
##  $ Sex        : Factor w/ 2 levels "female","male": 2 1 1 1 2 2 2 2 1 1 ...
##  $ Age        : num   22 38 26 35 35 NA 54 2 27 14 ...
##  $ SibSp      : int   1  1  0  1  0  0  0  3  0  1 ...
##  $ Parch      : int   0  0  0  0  0  0  0  1  2  0 ...
##  $ Ticket     : Factor w/ 681 levels "110152","110413",...: 525 596 662 50 473 276 86 396 345 133 ...
##  $ Fare       : num    7.25 71.28 7.92 53.1 8.05 ...
##  $ Cabin      : Factor w/ 148 levels "", "A10", "A14",...: 1 83 1 57 1 1 131 1 1 1 ...
##  $ Embarked   : Factor w/ 4 levels "", "C", "Q", "S": 4 2 4 4 4 3 4 4 4 2 ...
```

```
str(test)
```

```
## 'data.frame':      418 obs. of  11 variables:
##  $ PassengerId: int  892 893 894 895 896 897 898 899 900 901 ...
##  $ Pclass     : int   3  3  2  3  3  3  3  2  3  3 ...
##  $ Name       : Factor w/ 418 levels "Abbott, Master. Eugene Joseph",...: 210 409 273 414 182 370 85
##  $ Sex        : Factor w/ 2 levels "female","male": 2 1 2 2 1 2 1 2 1 2 ...
##  $ Age        : num   34.5 47 62 27 22 14 30 26 18 21 ...
##  $ SibSp      : int   0  1  0  0  1  0  0  1  0  2 ...
##  $ Parch      : int   0  0  0  0  1  0  0  1  0  0 ...
##  $ Ticket     : Factor w/ 363 levels "110469","110489",...: 153 222 74 148 139 262 159 85 101 268 ...
##  $ Fare       : num    7.83 7 9.69 8.66 12.29 ...
##  $ Cabin      : Factor w/ 77 levels "", "A11", "A18",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Embarked   : Factor w/ 3 levels "C", "Q", "S": 2 3 2 3 3 3 2 3 1 3 ...
```

```
table(train$Survived)
```

```
##  
##    0    1  
## 549 342
```

```
prop.table(table(train$Survived))
```

```
##  
##      0      1  
## 0.6162 0.3838
```

predict

```
test$Survived <- rep(0, dim(test)[1])
```

generate result

```
submit <- data.frame(PassengerId = test$PassengerId, Survived = test$Survived)  
write.csv(submit, file = "/home/yanyg/Kaggle/Titanic:MachineLearningfromDisaster/titanic/firstAttempt.csv")
```