

# svm\_CV.R

Magilan

Mon Oct 08 22:48:08 2018

```
library(e1071)
library(caret)
```

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

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

```
# Data Input
```

```
data <- read.csv("C:/Users/Magilan/Desktop/ML_project/austin_weather.csv",header = TRUE)
data1=na.omit(data,invert=FALSE)
attach(data1)
```

```
# Data Partitioning
```

```
index <- createDataPartition(Rain, p = 0.7, list = FALSE)
train.df <- data1[index,-c(1,20,22)]
test.df <- data1[-index,-c(1,20,21,22)]
test.Y <- data1[-index,21]
```

```
# SVM Model with Linear Kernel
```

```
model.svm <- svm(Rain ~ . , data = train.df)
```

```
pred.svm <- predict(model.svm, test.df, type = "C-classification")
head(pred.svm)
```

```
##  2  4  5 10 12 13
## no no no no no no
## Levels: no yes
```

```
confusionMatrix(pred.svm,test.Y)
```

```
## Confusion Matrix and Statistics
```

```
##
##           Reference
## Prediction  no yes
##      no  232  39
##      yes   25  94
##
##           Accuracy : 0.8359
##           95% CI : (0.7953, 0.8713)
##      No Information Rate : 0.659
##      P-Value [Acc > NIR] : 3.98e-15
##
##           Kappa : 0.6254
##  McNemar's Test P-Value : 0.1042
##
##           Sensitivity : 0.9027
##           Specificity : 0.7068
##           Pos Pred Value : 0.8561
##           Neg Pred Value : 0.7899
##           Prevalence : 0.6590
##           Detection Rate : 0.5949
##      Detection Prevalence : 0.6949
##           Balanced Accuracy : 0.8047
##
##           'Positive' Class : no
##
```

```
# Cross Validation

model.cv <- train(
  Rain ~., data = train.df[, -c(1, 20, 22)], method = "svmLinear",
  trControl = trainControl("cv", number = 10),
  preProcess = c("center", "scale"),
  tuneLength = 20
)
model.cv
```

```
## Support Vector Machines with Linear Kernel
##
## 915 samples
## 17 predictor
## 2 classes: 'no', 'yes'
##
## Pre-processing: centered (17), scaled (17)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 824, 824, 823, 824, 823, 823, ...
## Resampling results:
##
## Accuracy Kappa
## 0.8502628 0.6619508
##
## Tuning parameter 'C' was held constant at a value of 1
```

```
k=model.cv$bestTune
k
```

```
## C
## 1 1
```

```
pred.cv = predict(model.cv, test.df)
confusionMatrix(pred.cv, test.Y)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction no yes
##      no  229  36
##      yes   28  97
##
##              Accuracy : 0.8359
##              95% CI : (0.7953, 0.8713)
##      No Information Rate : 0.659
##      P-Value [Acc > NIR] : 3.98e-15
##
##              Kappa : 0.6295
##  Mcnemar's Test P-Value : 0.3816
##
##              Sensitivity : 0.8911
##              Specificity : 0.7293
##              Pos Pred Value : 0.8642
##              Neg Pred Value : 0.7760
##              Prevalence : 0.6590
##              Detection Rate : 0.5872
##      Detection Prevalence : 0.6795
##              Balanced Accuracy : 0.8102
##
##              'Positive' Class : no
##
```

```
# SVM AND RANDOM FOREST GIVES THE BEST ACCURACY APROX. 84.1%
```