Model Evaluation - Metrics

source("~/MCD/MCD/m.R")  
  
library(mlbench)  
data(PimaIndiansDiabetes)  
data(longley)  
  
print("Loaded")

## [1] "Loaded"

There are many diﬀerent metrics that you can use to evaluate your machine learning algorithms in R. When you use caret to evaluate your models, the default metrics used are accuracy for classiﬁcation problems and RMSE for regression. But caret supports a range of other popular evaluation metrics.

## Accuracy and Kappa

Accuracy and Kappa are the default metrics used to evaluate algorithms on binary and multiclass classiﬁcation datasets in caret. Accuracy is the percentage of correctly classiﬁed instances out of all instances. It is more useful on a binary classiﬁcation than multiclass classiﬁcation problem because it can be less clear exactly how the accuracy breaks down across those classes (e.g. you need to go deeper with a confusion matrix).

Kappa or Cohen’s Kappa is like classiﬁcation accuracy, except that it is normalized at the baseline of random chance on your dataset. It is a more useful measure to use on problems that have an imbalance in the classes (e.g. a 70% to 30% split for classes 0 and 1 and you can achieve 70% accuracy by predicting all instances are for class 0). In the example below the Pima Indians diabetes dataset is used. It has a class break down of 65% to 35% for negative and positive outcomes.

m <- m\_raw(PimaIndiansDiabetes) %>% m\_response(~ diabetes) %>% m\_generalized\_linear\_regression() %>% m\_cross\_validation(number = 5) %>% m\_accuracy\_metric() %>% m\_train() %>% m\_run()

## [1] "PimaIndiansDiabetes"

print(m$train)

## Generalized Linear Model   
##   
## 768 samples  
## 8 predictor  
## 2 classes: 'neg', 'pos'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 614, 614, 615, 615, 614   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.7695442 0.4656824  
##   
##

## RMSE and R-squared

RMSE and R2 are the default metrics used to evaluate algorithms on regression datasets in caret. RMSE or Root Mean Squared Error is the average deviation of the predictions from the observations. It is useful to get a gross idea of how well (or not) an algorithm is doing, in the units of the output variable.

R2 spoken as R Squared (also called the coeﬃcient of determination) provides a goodness-of-ﬁt measure for the predictions to the observations. This is a value between 0 and 1 for no-ﬁt and perfect ﬁt respectively. In this example the longley economic dataset is used. The output variable for this dataset is a number employed people in the population. It is not clear whether this is an actual count (e.g. in millions) or a percentage.

m <- m\_raw(longley) %>% m\_response(~ Employed) %>% m\_linear\_regression() %>% m\_cross\_validation(number = 5) %>% m\_rmse\_metric() %>% m\_train() %>% m\_run()

## [1] "longley"

print(m$train)

## Linear Regression   
##   
## 16 samples  
## 6 predictor  
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 12, 12, 14, 13, 13   
## Resampling results:  
##   
## RMSE Rsquared   
## 0.3868618 0.9883114  
##   
## Tuning parameter 'intercept' was held constant at a value of TRUE  
##

## Area Under ROC Curve

## Logarithmic Loss