Machine Learning Cheatsheet

Contents

[Decision Trees 1](#_Toc448030858)

[Overall Steps 1](#_Toc448030859)

# Decision Trees

## Overall Steps

1. Balance the Data, e.g.

safe\_loans\_raw = loans[loans[target] == 1]

risky\_loans\_raw = loans[loans[target] == -1]

# Since there are less risky loans than safe loans, find the ratio of the sizes

# and use that percentage to undersample the safe loans.

percentage = len(risky\_loans\_raw)/float(len(safe\_loans\_raw))

safe\_loans = safe\_loans\_raw.sample(percentage, seed = 1)

risky\_loans = risky\_loans\_raw

loans\_data = risky\_loans.append(safe\_loans)

print "Percentage of safe loans :", len(safe\_loans) / float(len(loans\_data))

print "Percentage of risky loans :", len(risky\_loans) / float(len(loans\_data))

print "Total number of loans in our new dataset :", len(loans\_data)

1. snext step