## Quiz 4



**5/5** points earned (100%)

Quiz passed!

Continue Course (/learn/practical-machine-learning/supplement/PvInj/course-project-instructions-read-first)

Back to Week 4 (/learn/practical-machine-learning/home/week/4)



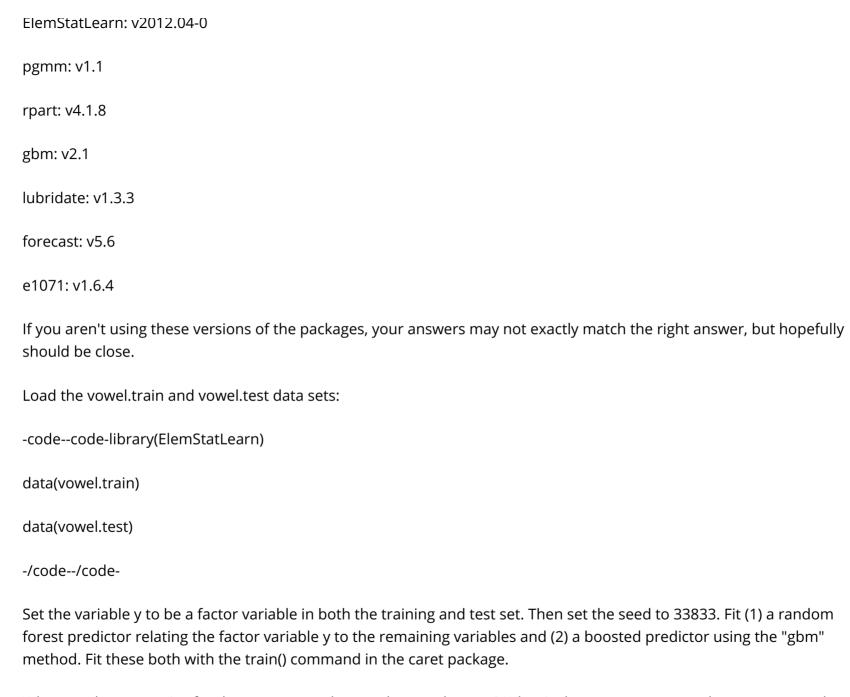
1/1 points

1

For this quiz we will be using several R packages. R package versions change over time, the right answers have been checked using the following versions of the packages.

AppliedPredictiveModeling: v1.1.6

caret: v6.0.47



What are the accuracies for the two approaches on the test data set? What is the accuracy among the test set samples where the two methods agree?

RF Accuracy = 0.9987

GBM Accuracy = 0.5152

Agreement Accuracy = 0.9985

RF Accuracy = 0.6082

GBM Accuracy = 0.5152

Agreement Accuracy = 0.5152

RF Accuracy = 0.9881

GBM Accuracy = 0.8371

Agreement Accuracy = 0.9983

RF Accuracy = 0.6082

GBM Accuracy = 0.5152

Agreement Accuracy = 0.6361

**Correct Response** 

2.

Load the Alzheimer's data using the following commands

-code-library(caret)

library(gbm)

set.seed(3433)

library(AppliedPredictiveModeling)

data(AlzheimerDisease)

adData = data.frame(diagnosis,predictors)

inTrain = createDataPartition(adData\$diagnosis, p = 3/4)[[1]]

training = adData[ inTrain,]

testing = adData[-inTrain,]

-/code-

Set the seed to 62433 and predict diagnosis with all the other variables using a random forest ("rf"), boosted trees ("gbm") and linear discriminant analysis ("lda") model. Stack the predictions together using random forests ("rf"). What is the resulting accuracy on the test set? Is it better or worse than each of the individual predictions?



Stacked Accuracy: 0.80 is better than random forests and Ida and the same as boosting.

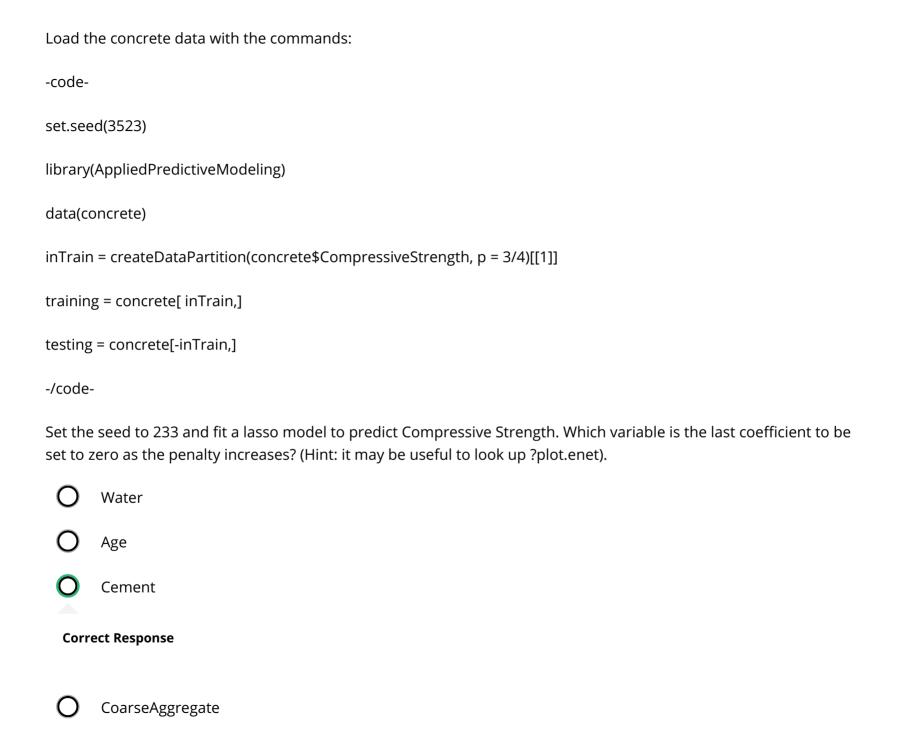
**Correct Response** 

0	Stacked Accuracy: 0.80 is better than all three other methods
0	Stacked Accuracy: 0.88 is better than all three other methods
0	Stacked Accuracy: 0.80 is worse than all the other methods.



1/1 points

3.





1/1 points

## 4.

Load the data on the number of visitors to the instructors blog from here:

https://d396qusza40orc.cloudfront.net/predmachlearn/gaData.csv (https://d396qusza40orc.cloudfront.net/predmachlearn/gaData.csv)

Using the commands:

-code-library(lubridate) # For year() function below

dat = read.csv("~/Desktop/gaData.csv")

training = dat[year(dat\$date) < 2012,]

testing = dat[(year(dat\$date)) > 2011,]

tstrain = ts(training\$visitsTumblr)

-/code-

Fit a model using the bats() function in the forecast package to the training time series. Then forecast this model for the remaining time points. For how many of the testing points is the true value within the 95% prediction interval bounds?



96%

**O** 94%

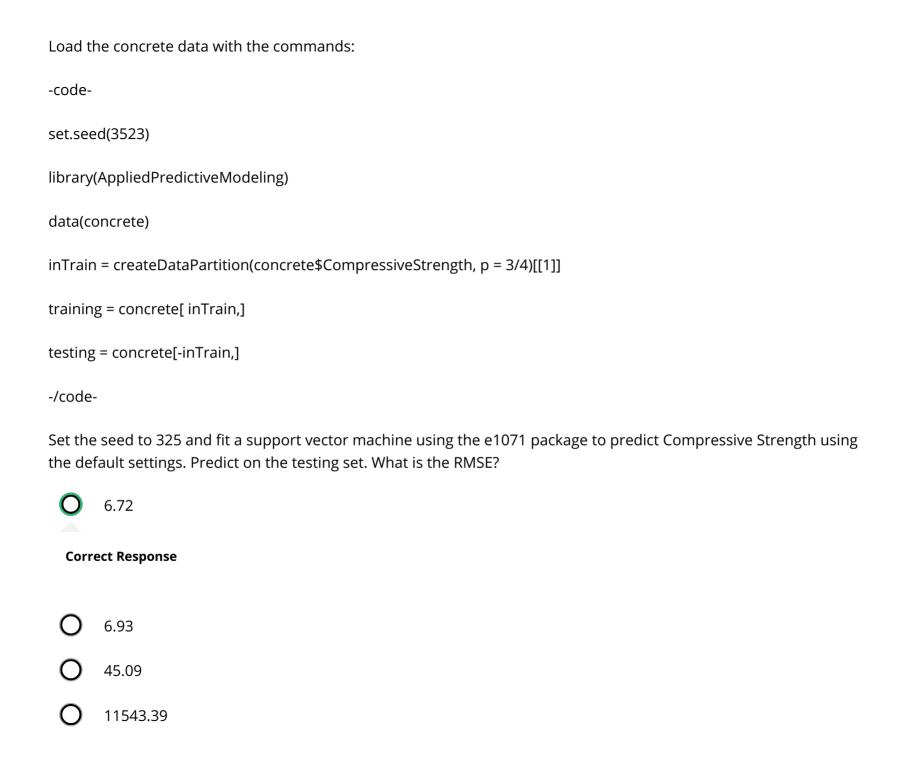
0 100%

92%



1/1 points

5.



r r