Bookmarks

- Unit 1: An Introduction to Analytics
- Entrance Survey
- Unit 2: Linear Regression
- Unit 3: Logistic Regression
- Unit 4: Trees
- Unit 5: Text Analytics
- Unit 6: Clustering
- Kaggle Competition
- Unit 7: Visualization
- Unit 8: Linear Optimization
- Exit Survey
- Unit 9: Integer Optimization

Final Exam > Final Exam > FORECASTING INTEREST RATE HIKES BY THE U.S. FEDERAL RESERVE

Bookmark

# FORECASTING INTEREST RATE HIKES BY THE U.S. FEDERAL RESERVE

The federal funds rate is the key interest rate that the U.S. Federal Reserve uses to influence economic growth. The Federal Open Market Committee meets regularly to decide whether to increase, decrease, or maintain the target interest rate. Their choice has important ramifications that cascade through the economy, so the announcement of the interest rates is eagerly awaited each month.

In this problem, we will use analytics to try to predict when the Federal Reserve will raise interest rates. We will look at monthly economic and political data dating back to the mid-1960's. In this analysis, the dependent variable will be the binary outcome variable **RaisedFedFunds**, which takes value 1 if the federal funds rate was increased that month and 0 if it was lowered or stayed the same. For each month, the file federalFundsRate.csv contains the following independent variables:

- **Date**: The date the change was announced.
- **Chairman**: The name of the Federal Reserve Chairman at the time the change was announced.
- **PreviousRate**: The federal funds rate in the prior month.
- **Streak**: The current streak of raising or not raising the rate, e.g. +8 indicates the rate has been increased 8 months in a row, whereas -3 indicates the rate has been lowered or stayed the same for 3 months in a row.
- **GDP**: The U.S. Gross Domestic Product, in Billions of Chained 2009 US Dollars.
- **Unemployment**: The unemployment rate in the U.S.
- **CPI**: The Consumer Price Index, an indicator of inflation, in the U.S.
- **HomeownershipRate**: The rate of homeownership in the U.S.
- **DebtAsPctGDP**: The U.S. national debt as a percentage of GDP

▾ Final Exam

**Final Exam**
Final Exam due Jul 05, 2016 at 00:00 UTC

- **DemocraticPres**: Whether the sitting U.S. President is a Democrat (DemocraticPres=1) or a Republican (DemocraticPres=0)
- **MonthsUntilElection**: The number of remaining months until the next U.S. presidential election.

## Problem 1 - Loading the Data

(1/1 point)
Use the read.csv function to load the contents of federalFundsRate.csv into a data frame called fedFunds, using stringsAsFactors=FALSE. What proportion of months did the Fed raise the interest rate?

0.5025641 ✓ **Answer:** 0.50

**EXPLANATION**

This can be computed with the table, mean, or summary functions.

*You have used 1 of 2 submissions*

## Problem 2 - The Longest-Serving Fed Chair

(1/1 point)
Which Federal Reserve Chair has presided over the most interest rate decisions?

○ Ben Bernanke

○ Arthur Burns

● Alan Greenspan ✓

○ William M. Martin

○ G. William Miller

○ Paul Volcker

○ Janet Yellen

**EXPLANATION**

This can be determined with the table or summary functions.

*You have used 1 of 1 submissions*

## Problem 3 - Converting Variables to Factors

(1/1 point)

Convert the following variables to factors using the as.factor function:

- Chairman

- DemocraticPres

- RaisedFedFunds

Which of the following methods requires the dependent variable be stored as a factor variable when training a model for classification?

○ Logistic regression (glm)

○ CART (rpart)

● Random forest (randomForest) ✓

**EXPLANATION**

We convert the outcome variable to a factor for the randomForest() method.

*You have used 1 of 1 submissions*

## Problem 4 - Splitting into a Training and Testing Set

(1/1 point)

Obtain a random training/testing set split with:

set.seed(201)

library(caTools)

spl = sample.split(fedFunds$RaisedFedFunds, 0.7)

Split months into a training data frame called "training" using the observations for which spl is TRUE and a testing data frame called "testing" using the observations for which spl is FALSE.

**EXPLANATION**

Use the subset function to put the TRUE observations in the training set, and the FALSE observations in the test set.

Why do we use the sample.split() function to split into a training and testing set?

○ It is the most convenient way to randomly split the data

○ It balances the independent variables between the training and testing sets

● It balances the dependent variable between the training and testing sets ✓

*You have used 1 of 1 submissions*

## Problem 5 - Training a Logistic Regression Model

(1/1 point)

Train a logistic regression model using independent variables "PreviousRate", "Streak", "Unemployment", "HomeownershipRate", "DemocraticPres", and "MonthsUntilElection", using the training set to obtain the model.

Which of the following characteristics is the most statistically significant associated with an increased chance of the federal funds rate being raised?

□ A higher federal funds rate in the previous month.

☑ A longer consecutive streak of months in which the federal funds rate was raised. ✓

□ A higher unemployment rate.

□ A higher rate of homeownership.

□ The current president being a Democrat.

□ A longer amount of time until the next presidential election.

✓

**EXPLANATION**

The model can be trained with the glm function (remember the argument family="binomial") and summarized with the summary function.

*You have used 1 of 2 submissions*

## Problem 6 - Predicting Using a Logistic Regression Model

(1/1 point)

Imagine you are an analyst at a bank and your manager has asked you to predict whether the federal funds rate will be raised next month. You know that the rate has been lowered for 3 straight months (Streak = -3) and that the previous month's rate was 1.7%. The unemployment rate is 5.1% and the homeownership rate is 65.3%. The current U.S. president is a Republican and the next election will be held in 18 months. According to the logistic regression model you built in Problem 5, what is the predicted probability that the interest rate will be raised?

0.3464 ✓ **Answer:** 0.3464

**EXPLANATION**

The observation has PreviousRate=1.7, Streak=-3, Unemployment=5.1, DemocraticPres=0, MonthsUntilElection=18. Therefore, the prediction has logistic function value 9.121012 + 1.7*(-0.003427) - 3* 0.157658 + 5.1*(-0.047449) + 65.3*(-0.136451) + 0*0.347829 + 18*(-0.006931) = -0.6347861. Then you need to plug this into the logistic response function to get the predicted probability.

*You have used 1 of 2 submissions*

## Problem 7 - Interpreting Model Coefficients

(1/1 point)

What is the meaning of the coefficient labeled "DemocraticPres1" in the logistic regression summary output?

○ When the president is Democratic, the odds of the federal funds rate increasing are 34.8% higher than in an otherise identical month (i.e. identical among the variables in the model).

○ When the president is Democratic, the odds of the federal funds rate increasing are 34.8% higher than in an average month in the dataset.

● When the president is Democratic, the odds of the federal funds rate increasing are 41.6% higher than in an otherise identical month (i.e. identical among the variables in the model). ✓

○ When the president is Democratic, the odds of the federal funds rate increasing are 41.6% higher than in an average month in the dataset.

**EXPLANATION**

The coefficients of the model are the log odds associated with that variable; so we see that the odds of being sold are exp(0.347829)=1.41599 those of an otherwise identical month. This means the month is predicted to have 41.6% higher odds of being sold.

*You have used 1 of 1 submissions*

## Problem 8 - Obtaining Test Set Predictions

(2/2 points)

Using your logistic regression model, obtain predictions on the test set. Then, using a probability threshold of 0.5, create a confusion matrix for the test set.

We would like to compare the predictions obtained by the logistic regression model and those obtained by a naive baseline model. Remember that the naive baseline model we use in this class always predicts the most frequent outcome in the training set for all observations in the test set.

What is the number of test set observations where the prediction from the logistic regression model is different than the prediction from the baseline model?

91

✓ **Answer:** 91

**EXPLANATION**

Obtain test-set predictions with the predict function, remembering to pass type="response". Using table, you can see that there are 91 test-set predictions with probability less than 0.5.

*You have used 1 of 2 submissions*

## Problem 9 - Computing Test-Set AUC

(2/2 points)
What is the test-set AUC of the logistic regression model?

0.704023

✓ **Answer:** 0.704023

**EXPLANATION**

The test-set AUC can be obtained by loading the ROCR package, and then using the prediction and performance functions.

*You have used 1 of 2 submissions*

## Problem 10 - Interpreting AUC

(1/1 point)
What is the meaning of the AUC?

● The proportion of the time the model can differentiate between a randomly selected month during which the federal funds were raised and a randomly selected month during which the federal funds were not raised ✓

○ The proportion of the time the model correctly identifies whether or not federal funds were raised

**EXPLANATION**

The AUC is the proportion of time the model can differentiate between a randomly selected true positive and true negative.

*You have used 1 of 1 submissions*

## Problem 11 - ROC Curves

(1/1 point)

Which logistic regression threshold is associated with the upper-right corner of the ROC plot (true positive rate 1 and false positive rate 1)?

● 0 ✓

○ 0.5

○ 1

**EXPLANATION**

A model with threshold 0 predicts 1 for all observations, yielding a 100% true positive rate and a 100% false positive rate.

*You have used 1 of 1 submissions*

## Problem 12 - ROC Curves

(1/1 point)

Plot the colorized ROC curve for the logistic regression model's performance on the test set.

At roughly which logistic regression cutoff does the model achieve a true positive rate of 85% and a false positive rate of 60%?

○ 0

○ 0.16

● 0.37 ✓

○ 0.52

○ 0.66

○ 0.87

**EXPLANATION**

You can plot the colorized curve by using the plot function, and adding the argument colorize=TRUE.

From the colorized curve, we can see that the light turqoise color, corresponding to cutoff 0.37, is associated with a true positive rate of about 0.85 and false positive rate of about 0.6.

*You have used 1 of 1 submissions*

## Problem 13 - Cross-Validation to Select Parameters

(1/1 point)

Which of the following best describes how 10-fold cross-validation works when selecting between 2 different parameter values?

○ 2 models are trained on subsets of the training set and evaluated on the testing set

○ 10 models are trained on subsets of the training set and evaluated on the testing set

○ 20 models are trained on subsets of the training set and evaluated on the testing set

○ 2 models are trained on subsets of the training set and evaluated on a portion of the training set

○ 10 models are trained on subsets of the training set and evaluated on a portion of the training set

● 20 models are trained on subsets of the training set and evaluated on a portion of the training set ✓

**EXPLANATION**

In 10-fold cross validation, the model with each parameter setting will be trained on 10 90% subsets of the training set. Hence, a total of 20 models will be trained. The models are evaluated in each case on the last 10% of the training set (not on the testing set).

*You have used 1 of 1 submissions*

## Problem 14 - Cross-Validation for a CART Model

(1/1 point)

Set the random seed to 201 (even though you have already done so earlier in the problem). Then use the caret package and the train function to perform 10-fold cross validation with the training data set to select the best cp value for a CART model that predicts the dependent variable

"RaisedFedFunds" using the independent variables "PreviousRate," "Streak," "Unemployment," "HomeownershipRate," "DemocraticPres," and "MonthsUntilElection." Select the cp value from a grid consisting of the 50 values 0.001, 0.002, ..., 0.05.

What cp value maximizes the cross-validation accuracy?

0.016 ✓ **Answer:** 0.016

**EXPLANATION**

The cross-validation can be run by first setting the grid of cp values with the expand.grid function and setting the number of folds with the trainControl function. Then you want to use the train function to run the cross-validation.

From the output of the train function, parameter value 0.016 yields the highest cross-validation accuracy.

*You have used 1 of 2 submissions*

## Problem 15 - Train CART Model

(1/1 point)

Build and plot the CART model trained with the parameter identified in Problem 14, again predicting the dependent variable using "PreviousRate", "Streak", "Unemployment", "HomeownershipRate", "DemocraticPres", and "MonthsUntilElection". What variable is used as the first (upper-most) split in the tree?

○ PreviousRate

● Streak ✓

○ Unemployment

○ HomeownershipRate

○ DemocraticPres

○ MonthsUntilElection

**EXPLANATION**

The CART model can be trained and plotted by first loading the "rpart" and "rpart.plot" packages, and then using the rpart function to build the model and the prp function to plot the tree.

*You have used 1 of 1 submissions*

## Problem 16 - Predicting Using a CART Model

(1/1 point)

If you were to use the CART model you created in Problem 15 to answer the question asked of the analyst in Problem 6, what would you predict for next month?

Remember: The rate has been lowered for 3 straight months (Streak = -3). The previous month's rate was 1.7%. The unemployment rate is 5.1%. The homeownership rate is 65.3%. The current U.S. president is a Republican and the next election will be held in 18 months.

○ The Fed will raise the federal funds rate.

● The Fed will not raise the federal funds rate. ✓

EXPLANATION

Once the tree is plotted using the prp function, you can follow the splits to find that, because Streak is less than 2.5 and Streak is less than -1.5, the model predicts RaisedFedFunds=0.

*You have used 1 of 1 submissions*

## Problem 17 - Test-Set Accuracy for CART Model

(2/2 points)

Using the CART model you created in Problem 15, obtain predictions on the test set (using the parameter type="class" with the predict function). Then, create a confusion matrix for the test set.

What is the accuracy of your CART model?

0.64 ✓ **Answer:** 0.64

EXPLANATION

The test set predictions can be obtained using the predict function. The confusion matrix can be obtained using the table function.

From the table, the accuracy is (64+48)/(64+48+23+40) = 0.64.

*You have used 1 of 2 submissions*

© All Rights Reserved

© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY OPENedX

Download on the App Store

ANDROID APP ON Google play