h2omlestat threshmetric — Display threshold-based metrics for binary classification

Description

`h2omlestat threshmetric` reports threshold-based performance metrics after binary classification performed by `h2oml gbbinclass` or `h2oml rfbinclass`. Threshold-based metrics are functions of predicted classes, which are determined by comparing predicted probabilities with a threshold value. Observations with predicted probabilities greater than the threshold are predicted to be in the "positive" class, and observations with predicted probabilities below the threshold are predicted to be in the "negative" class. The elements of the confusion matrix—the numbers of true positives, false positives, true negatives, and false negatives—are threshold-based metrics and are components of a variety of additional threshold-based metrics that are reported by `h2omlestat threshmetric`. Each of these metrics has a different threshold value.

`h2omlestat threshmetric` reports the optimized (minimum or maximum) value of each metric and the corresponding threshold that produces that optimized metric. Alternatively, the metrics can be reported for one or more selected threshold values.

Quick start

Display threshold-based metrics

```
h2omlestat threshmetric
```

Same as above, but report metrics based on a validation set

```
h2omlestat threshmetric, valid
```

Same as above, but report metrics corresponding to threshold values of 0.4, 0.5, 0.6, 0.7, and 0.8

```
h2omlestat threshmetric, valid thresholds(0.4(0.1)0.8)
```

Menu

Statistics > H2O machine learning

Syntax

```
h2omlestat threshmetric [ , options ]
```

options	Description
Main	
thresholds(*numlist*)	specify the thresholds for which to compute the metrics; by default, the threshold that optimizes each metric is reported
Table options	
all	report metrics for all stored threshold values
index	display threshold index
title(*string*)	specify the title to be displayed above the table
train	specify that performance metrics be reported using training results
valid	specify that performance metrics be reported using validation results
cv	specify that performance metrics be reported using cross-validation results
test	specify that performance metrics be computed using the testing frame
test(*framename*)	specify that performance metrics be computed using data in testing frame *framename*
frame(*framename*)	specify that performance metrics be computed using data in H2O frame *framename*
framelabel(*string*)	label frame as *string* in the output

collect is allowed; see **[U] 11.1.10 Prefix commands**.

train, valid, cv, test, test(), frame(), and framelabel() do not appear in the dialog box.

Options

Main

thresholds(*numlist*) specifies the list of threshold values in *numlist*. All values in *numlist* must be between 0 and 1. Observations with predicted probabilities greater than the specified threshold are classified as "positive", and the remaining observations are classified as "negative". The threshold-based metrics are calculated based on these classifications. By default, the threshold values that optimize (maximize or minimize) each metric are reported.

The list of threshold values for which threshold-based metrics are computed corresponds to the predicted probabilities of the positive class (the predicted class is the largest numeric value, such as 1 in a 0/1 coded variable, or the second label in lexicographical order). If a value specified in *numlist* is not in the list of predicted probabilities, the metric based on the closest threshold value is reported. thresholds() is not allowed with all.

Table options

all returns all stored threshold values and metrics. The default is to report the optimized (maximum or minimum) values for each metric. all is not allowed with thresholds().

index displays the index number of the threshold. By default, the index column is suppressed.

title(*string*) specifies the title to be displayed above the table.

The following options are available with h2omlestat threshmetric but are not shown in the dialog box:

train, valid, cv, test, test(), and frame() specify the H2O frame for which performance metrics are reported. Only one of train, valid, cv, test, test(), or frame() is allowed.

train specifies that performance metrics be reported using training results. This is the default when neither validation nor cross-validation is performed during estimation and when a postestimation frame has not been set with h2omlpostestframe.

valid specifies that performance metrics be reported using validation results. This is the default when validation is performed during estimation and when a postestimation frame has not been set with h2omlpostestframe. valid may be specified only when the validframe() option is specified with h2oml *gbm* or h2oml *rf*.

cv specifies that performance metrics be reported using cross-validation results. This is the default when cross-validation is performed during estimation and when a postestimation frame has not been set with h2omlpostestframe. cv may be specified only when the cv or cv() option is specified with h2oml *gbm* or h2oml *rf*.

test specifies that performance metrics be computed on the testing frame specified with h2omlpostestframe. This is the default when a testing frame is specified with h2omlpostestframe. test may be specified only after a testing frame is set with h2omlpostestframe. test is necessary only when a subsequent h2omlpostestframe command is used to set a default postestimation frame other than the testing frame.

test(*framename*) specifies that performance metrics be computed using data in testing frame *framename* and is rarely used. This option is most useful when running a single postestimation command on the named frame. If multiple postestimation commands are to be run on the same test frame, h2omlpostestframe provides a more convenient and computationally efficient process for doing this.

frame(*framename*) specifies that performance metrics be computed using the data in H2O frame *framename*.

framelabel(*string*) specifies the label to be used for the frame in the output. This option is not allowed with the cv option.

Remarks and examples

Binary classification divides observations into two classes, typically labeled as "positive" and "negative". In H2O, the positive class corresponds to the class that contains 1, True, or the second label in lexicographical order. A binary classifier classifies all observations as either positive or negative by comparing the predicted probability for each observation with a threshold value. Observations greater than the threshold are classified as positive, and the remaining observations are classified as negative. This results in two types of correct or true classification, true positive and true negative, and two types of incorrect or false classification, false positive and false negative. These four metrics are reported in the confusion matrix produced by the h2omlestat confmatrix command. The h2omlestat threshmetric command reports these metrics as well as other performance metrics that are derived from the elements of a confusion matrix.

By default, h2omlestat threshmetric reports the optimized (minimum or maximum) value of each metric and the corresponding threshold value that produces the optimized metric. You can also evaluate how different threshold values affect each metric by specifying one or more threshold values in the

thresholds() option. When you specify the thresholds() option, metrics may not be reported for the exact threshold values you have selected. In H2O, the available thresholds are limited to the list of predicted probabilities of the positive class. Threshold-based metrics are reported for the threshold corresponding to the closest available predicted probability.

The table below provides definitions of the available threshold-based metrics. See *Metrics for classification* in [H2OML] ***metric_option*** for additional information.

Metric	Formula
true positive (tp)	number of correct predictions of the positive class
true negative (tn)	number of correct predictions of the negative class
false positive (fp)	number of incorrect predictions of the positive class
false negative (fn)	number of incorrect predictions of the negative class
true-positive rate (tpr), recall	$\frac{\text{tp}}{\text{tp}+\text{fn}}$
true-negative rate (tnr)	$\frac{\text{tn}}{\text{tn}+\text{fp}}$
false-positive rate (fpr)	$\frac{\text{fp}}{\text{tn}+\text{fp}}$
false-negative rate (fnr)	$\frac{\text{fn}}{\text{tp}+\text{fn}}$
accuracy	$\frac{\text{tp}+\text{tn}}{\text{tp}+\text{tn}+\text{fp}+\text{fn}}$
mean per class accuracy	$\frac{\text{tpr}+\text{tnr}}{2}$
min. per class accuracy	minimum of $\{\text{tpr}, \text{tnr}\}$
specificity	$\frac{\text{tn}}{\text{tn}+\text{fp}}$
precision	$\frac{\text{tp}}{\text{tp}+\text{fp}}$
F_β score, for $\beta = \{1, 0.5, 2\}$	$(1+\beta^2)\frac{\text{precision}\times\text{recall}}{\beta^2(\text{precision}+\text{recall})}$
Matthews correlation coefficient	$\frac{\text{tp}\times\text{tn}-\text{fp}\times\text{fn}}{\sqrt{(\text{tp}+\text{fp})(\text{tp}+\text{fn})(\text{tn}+\text{fp})(\text{tn}+\text{fn})}}$

▷ Example 1: Report threshold-based metrics

Below, we illustrate the use of h2omlestat threshmetric after h2oml gbbinclass.

We start by opening the 1978 automobile data (auto.dta) in Stata and then putting the data into an H2O frame. Recall that h2o init initiates an H2O cluster, _h2oframe put loads the current Stata dataset into an H2O frame, and _h2oframe change makes the specified frame the current H2O frame. For details, see *Prepare your data for H2O machine learning in Stata* in [H2OML] **h2oml** and see [H2OML] **H2O setup**.

We use the _h2oframe split command to randomly split the auto frame into a training frame (70% of observations) and a testing frame (30% of observations), which we name train and test, respectively. We also change the current frame to train.

```
. use https://www.stata-press.com/data/r19/auto
(1978 automobile data)
. h2o init
  (output omitted)
. _h2oframe put, into(auto)
Progress (%): 0 100
. _h2oframe split auto, into(train test) split(0.7 0.3) rseed(19)
. _h2oframe change train
```

Next we perform gradient boosting binary classification with default values.

```
. h2oml gbbinclass foreign price mpg weight length, h2orseed(19)
Progress (%): 0 100
Gradient boosting binary classification using H2O
Response: foreign
Loss:     Bernoulli
Frame:                                  Number of observations:
  Training: train                                 Training =     57
Model parameters
Number of trees      =  50              Learning rate       =     .1
              actual =  50              Learning rate decay =      1
Tree depth:                             Pred. sampling rate =      1
           Input max =   5              Sampling rate       =      1
                 min =   2              No. of bins cat.    =  1,024
                 avg = 2.9              No. of bins root    =  1,024
                 max =   4              No. of bins cont.   =     20
Min. obs. leaf split =  10              Min. split thresh.  = .00001
Metric summary
```

Metric	Training
Log loss	.1057473
Mean class error	.0125
AUC	.9948529
AUCPR	.9870295
Gini coefficient	.9897059
MSE	.0255994
RMSE	.1599981

```
. h2omlest store mygbm
```

To report threshold-based metrics, we use the `h2omlestat threshmetric` command.

```
. h2omlestat threshmetric
Maximum or minimum metrics using H2O
Training frame: train
```

Metric	Max/Min	Threshold	
F1	.9714	.6608	
F2	.9884	.6608	
F0.5	.9551	.6608	
Accuracy	.9825	.6608	
Precision	1	.9694	
Recall	1	.6608	
Specificity	1	.9694	
Min. class accuracy	.975	.6608	
Mean class accuracy	.9875	.6608	
True negatives	40	.9694	
False negatives	0	.6608	+
True positives	17	.6608	
False positives	0	.9694	+
True-negative rate	1	.9694	
False-negative rate	0	.6608	+
True-positive rate	1	.6608	
False-positive rate	0	.9694	+
MCC	.9596	.6608	

```
+ identifies minimum metrics.
```

By default, because we did not use validation or cross-validation, `h2omlestat threshmetric` reports training results. The reported table has three columns. The first column provides the names of the classification metrics. The second and third columns report the optimal value of each metric (maximum or minimum) and the threshold value that achieves the optimum. The reported optimal value of the metric is the minimum for the false-negative rate, false-positive rate, false negatives, and false positives metrics and is the maximum for all other metrics.

We can use the `thresholds()` option to obtain the reported metrics for a different threshold value or values. For example, to report metrics for a threshold of 0.5, we type

```
. h2omlestat threshmetric, thresholds(0.5)
Metrics for specific threshold using H2O
Training frame: train
```

Threshold	
Input	.5
Computed	.4477
Metric	
F1	.9444
F2	.977
F0.5	.914
Accuracy	.9649
Precision	.8947
Recall	1
Specificity	.95
Min. class accuracy	.95
Mean class accuracy	.975
True negatives	38
False negatives	0
True positives	17
False positives	2
True-negative rate	.95
False-negative rate	0
True-positive rate	1
False-positive rate	.05
MCC	.922

We see that, even though we specified `thresholds(0.5)`, H2O returned results for a threshold of 0.4477, which is the closest available threshold (those found among the stored predicted probabilities).

◁

Example 2: Threshold-based metrics using testing frame

Above, we reported metrics for the training frame. If we wish to report those metrics on the new testing data frame, then we can take one of two approaches.

In the first approach, we specify the `test()` option with the name of our testing frame.

```
. h2omlest restore mygbm
(results mygbm are active now)
. h2omlestat threshmetric, test(test)
Maximum or minimum metrics using H2O
Testing frame: test
```

Metric	Max/Min	Threshold	
F1	.8333	.4477	
F2	.9259	.4477	
F0.5	.8824	.8916	
Accuracy	.8824	.8916	
Precision	1	.9694	
Recall	1	.4477	
Specificity	1	.9694	
Min. class accuracy	.8333	.4477	
Mean class accuracy	.9167	.4477	
True negatives	12	.9694	
False negatives	0	.4477	+
True positives	5	.4477	
False positives	0	.9694	+
True-negative rate	1	.9694	
False-negative rate	0	.4477	+
True-positive rate	1	.4477	
False-positive rate	0	.9694	+
MCC	.7715	.4477	

```
+ identifies minimum metrics.
```

In the second approach, which we recommend, we use the h2omlpostestframe command to specify test as the default testing frame to be used by this and other postestimation commands.

```
. h2omlpostestframe test
(testing frame test is now active for h2oml postestimation)
. h2omlestat threshmetric
Maximum or minimum metrics using H2O
Testing frame: test
```

Metric	Max/Min	Threshold	
F1	.8333	.4477	
F2	.9259	.4477	
F0.5	.8824	.8916	
Accuracy	.8824	.8916	
Precision	1	.9694	
Recall	1	.4477	
Specificity	1	.9694	
Min. class accuracy	.8333	.4477	
Mean class accuracy	.9167	.4477	
True negatives	12	.9694	
False negatives	0	.4477	+
True positives	5	.4477	
False positives	0	.9694	+
True-negative rate	1	.9694	
False-negative rate	0	.4477	+
True-positive rate	1	.4477	
False-positive rate	0	.9694	+
MCC	.7715	.4477	

```
+ identifies minimum metrics.
```

Stored results

h2omlestat threshmetric stores the following in r():

Macros

r(thresholds)	specified thresholds
r(thresholds_a)	actual thresholds

Matrices

r(threshmetric)	classification performance metrics

Also see

[H2OML] **h2oml** — Introduction to commands for Stata integration with H2O machine learning