h2omlgraph ice — Produce individual conditional expectation plot+

+This command includes features that are part of StataNow.

Description	Quick start	Menu	Syntax
Options	Remarks and examples	References	Also see

Description

h2omlgraph ice plots the individual conditional expectation (ICE) curves after h2oml *gbm* and h2oml *rf*. For regression, the ICE values correspond to predictions for an individual observation as values of a predictor of interest vary. For classification, the ICE values correspond to the predicted probabilities for an individual observation as values of a predictor of interest vary. Rather than plotting the ICE curve for every observation, h2omlgraph ice plots ICE curves at the boundaries of the deciles of the predictor of interest. The graph produced by h2omlgraph ice is useful for evaluating the partial effect of a predictor on the response and how that effect differs across deciles of the predictor. It is also useful for determining whether interaction effects exist between the variable of interest and other predictors.

The ICE plots are similar to the partial density plot (PDP), but the PDP estimates the average predictions for the entire dataset and can be considered as the average of the ICE curves for all observations.

Quick start

Plot the ICE for predictor x1

```
h2omlgraph ice x1
```

As above, but do not show histogram in the plot

```
h2omlgraph ice x1, nohistogram
```

Plot the ICE after the multiclass classification for the class no and using H2O frame myframe

```
h2omlgraph ice x1, target(no) frame(myframe)
```

Menu

Statistics > H2O machine learning

Syntax

`h2omlgraph ice` *predictor* [, *options*]

options	Description
Main	
* `target(`*class*`)`	specify the target class of the response after multiclass classification
`maxlevels(#)`	specify the maximum number of levels for categorical predictors; default is `maxlevels(30)`
`savedata(`*filename*[, `replace`]`)`	save plot data to *filename*
Plot options	
`nohistogram`	do not plot histogram of the predictor
`histopts(`*bar_opts*`)`	affect rendition of the histogram
`line#opts(`*line_options*`)`	affect rendition of the ICE curve for quantile #
`nopdline`	do not plot partial dependence curve
`pdlineopts(`*line_options*`)`	affect rendition of partial dependence curve
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***
`train`	specify that the ICE be reported using training results
`valid`	specify that the ICE be reported using validation results
`test`	specify that the ICE be computed using testing frame
`test(`*framename*`)`	specify that the ICE be computed using data in testing frame *framename*
`frame(`*framename*`)`	specify that the ICE be computed using data in H2O frame *framename*
`framelabel(`*string*`)`	label frame as *string* in the output

*`target()` is required after multiclass classification.

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

Options

Main

`target(`*class*`)` specifies for which class of the response variable the ICE should be plotted. `target()` is required after multiclass classification with `h2oml gbmulticlass` or `h2oml rfmulticlass`.

`maxlevels(#)` specifies the maximum number of levels of the specified categorical predictor to be included in the ICE estimation. The default is `maxlevels(30)`.

`savedata(`*filename*[, `replace`]`)` saves the plot data to a Stata data file (`.dta` file). `replace` specifies that *filename* be overwritten if it exists.

Plot options

`nohistogram` removes the histogram of the predictor. By default, the histogram is included.

`histopts(`*bar_opts*`)` affects rendition of the histogram; see [G-2] **graph twoway bar**.

`line#opts(`*line_options*`)` affects the rendition of the ICE curve for decile #. See [G-3] ***line_options***.

nopdline removes the line for the partial dependence curve. The partial dependence curve is included by default.

pdlineopts(*line_options*) affects rendition of the partial dependence curve; see [G-3] ***line_options***.

twoway_options are any of the options documented in [G-3] ***twoway_options***, excluding by(). These include options for titling the graph (see [G-3] ***title_options***) and options for saving the graph to disk (see [G-3] ***saving_option***).

The following options are available with h2omlgraph ice but are not shown in the dialog box:

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

train specifies that ICE be reported using training results. This is the default when validation is not performed during estimation and when a postestimation frame has not been set with h2omlpostestframe.

valid specifies that ICE 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*.

test specifies that ICE 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 by using 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 ICE 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, it is more computationally efficient and convenient to specify the testing frame by using h2omlpostestframe instead of specifying test(*framename*) with individual postestimation commands.

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

framelabel(*string*) specifies the label to be used for the frame in the output.
stata.com

Remarks and examples

We assume you have read the *Interpretation and explanation* in [H2OML] **Intro**.

Remarks are presented under the following headings:

Introduction

The PDP, introduced in [H2OML] **h2omlgraph pdp**, graphs the average predictions across the values of a predictor of interest and is useful for understanding the average or partial effect of the predictor on the response. However, when there is an interaction effect among predictors, the PDP cannot fully capture the effect. In fact, there may be no average effect shown by a flat curve in the PDP, while there are substantial effects at various levels of the predictor, but the effects are in opposite directions and cancel each other out when averaged in the PDP. The ICE plots improve upon the PDPs by visualizing the relationship between the response and the predictor for individual observations (Goldstein et al. 2015).

Formally, let $f(\mathbf{X}_S, \mathbf{X}_C)$ be our machine learning model, $\mathbf{X}_S$ be the predictor whose effect we wish to study, and $\mathbf{X}_C$ be all other predictors in our model.

To obtain ICE values for all observations $i = 1, 2, \ldots, n$, the values of predictors $\mathbf{X}_C$ are fixed to their observed values of $\mathbf{x}_{Ci}$. Then the values of $\mathbf{X}_S$ are iteratively set to the observed value $\mathbf{x}_{Sj}$ for observations $j = 1, 2, \ldots, n$ to obtain predictions $\widehat{f}(\mathbf{x}_{Sj}, \mathbf{x}_{Ci})$. Thus, for each observation i in the dataset, we obtain n predicted values. These correspond to predictions where $\mathbf{X}_S$ is set to its observed value in observations $j = 1, \ldots, n$, while the remaining predictors $\mathbf{X}_C$ are held at their observed values for the same observation.

The ICE curve for observation i plots the resulting predicted values on the y axis and the predictor of interest $\mathbf{X}_S$ on the x axis. In practice, if the number of observations n is large, displaying a graph with curves for each observation becomes difficult to read. Therefore, it is recommended to consider using only deciles or quantiles of the data. `h2omlgraph ice` plots ICE curves for deciles of the predictor of interest. By default, it also plots the partial dependence curve for comparison with the ICE curves.

Examples of ICE curves

In this section, we demonstrate the advantage of `h2omlgraph ice` when an interaction effect is present among predictors. As with most explainable machine learning methods, caution is advised when using those results for decision making. For examples where explainable machine learning methods fail, see example 2 of [H2OML] **h2omlgraph varimp**, Krishna et al. (2022), Lakkaraju and Bastani (2020), and Slack et al. (2020).

The examples are presented under the following headings:

Example 1: Capturing an interaction effect through ICE
Example 2: Finding regions of interactions
Example 3: ICE plot for multinomial classification

▷ Example 1: Capturing an interaction effect through ICE

This example is borrowed from Goldstein et al. (2015). We consider the following data-generation process with an interaction: $Y = 0.2X_1 + 5X_2 + \varepsilon$ if $X_3 \geq 0$ and $Y = 0.2X_1 - 5X_2 + \varepsilon$ otherwise. Here $X_1, X_2, X_3 \sim U(-1, 1)$ and $\varepsilon \sim N(0, 1)$.

We start by opening the simulated `interaction.dta` dataset in Stata and then putting it 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 [H2OML] **H2O setup**.

```
. use https://www.stata-press.com/data/r18/interaction
(Fictional interaction data)
. h2o init
  (output omitted)
. _h2oframe put, into(interaction)
Progress (%): 0 100
. _h2oframe change interaction
```

For illustration purposes, we use h2oml rfregress to perform random forest regression with default values for hyperparameters. We then store the estimation results by using the h2omlest store command.

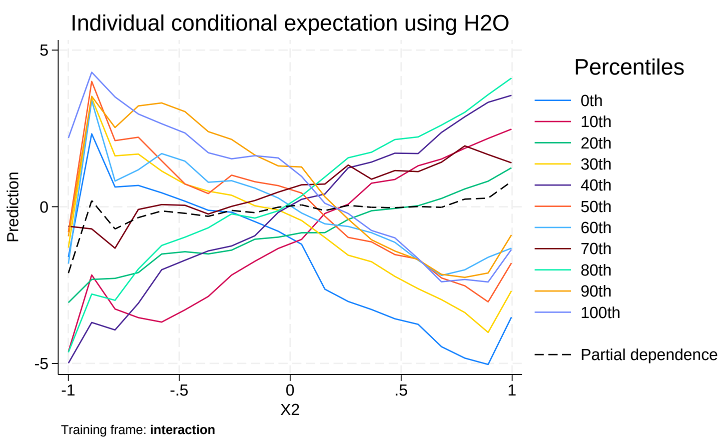
```
. h2oml rfregress Y X1 X2 X3, h2orseed(19)
Progress (%): 0 54.0 100
Random forest regression using H2O
Response: Y
Frame:                                      Number of observations:
  Training: interaction                                Training =    500
Model parameters
Number of trees        =   50
              actual   =   50
Tree depth:                                 Pred. sampling value =     -1
           Input max   =   20               Sampling rate        =   .632
                 min   =   16               No. of bins cat.     =  1,024
                 avg   = 18.8               No. of bins root     =  1,024
                 max   =   20               No. of bins cont.    =     20
Min. obs. leaf split   =    1               Min. split thresh.   = .00001
Metric summary
```

Metric	Training
Deviance	2.876126
MSE	2.876126
RMSE	1.695915
RMSLE	.
MAE	1.29916
R-squared	.6973235

```
. h2omlest store rf_inter
```

Next we plot ICE curves for X2 by using the h2omlgraph ice command.

```
. h2omlgraph ice X2
```



Here the dashed black line represents the partial dependence, and the other 11 lines correspond to ICE computed at the boundaries of the deciles X2—the 0th, 10th, ..., 100th percentiles of the observed values of X2 in the dataset. The partial dependence suggests no partial effect of X2 on the response, because the curve is mostly flat over the range of X2 values. This aggregate effect close to zero is actually the result of the individual effects canceling each other out. Some of them are positive (the ICE lines that increase with X2), and some of them negative (the ICE lines that decrease with X2).

In contrast to the PDP, the ICE curves provide a more comprehensive representation of the relationship between X2 and the response. Moreover, an interaction effect can be inferred from the ICE plots, because depending on the region of the X2 predictor space, ICE is either increasing or decreasing.

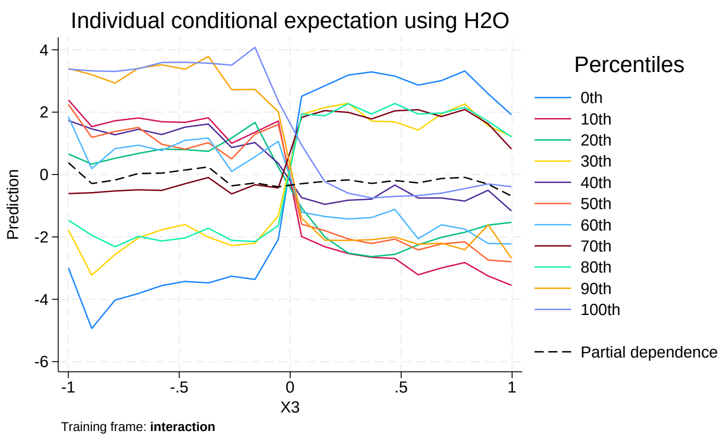
◁

▷ Example 2: Finding regions of interactions

In example 1, we showed that the ICE plots suggest some interaction effects among predictors. In this example, we are interested in detecting the regions where those interactions occur. For details, see Goldstein et al. (2015, sect. 4.2).

We now visualize ICE plots for the predictor X3.

```
. h2omlgraph ice X3
Progress (%): 0 10 20 30 40 50 60 70 80 90 100
```



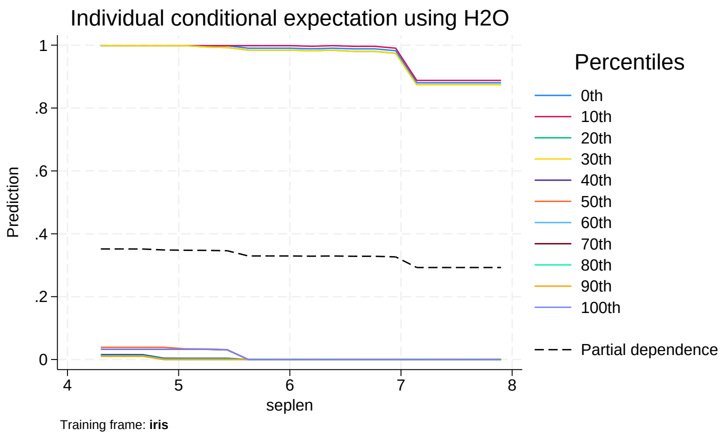
As in example 1, PDP suggests no effect of X3 on the response. However, the nonparallel ICE curves show the effect of X3 changes for each of the plotted percentiles near the neighborhood of $X3 = 0$. This indicates an interaction of X3 with another variable at this point, and we know this to be true based on the data-generating process for our simulated data.

◁

▷ Example 3: ICE plot for multinomial classification

In example 5 of [H2OML] **h2omlgraph pdp**, we showed how to implement and interpret PDP after multiclass classification. In this example, we continue from example 5 and plot ICE curves. Note that, compared with `h2omlgraph pdp`, the `target()` option of `h2omlgraph ice` supports only one class of the response variable. Here we plot ICE for the `Setosa` class in `iris`.

```
. h2omlgraph ice seplen, target(Setosa)
Progress (%): 0 10 20 30 40 50 60 70 80 90 100
```



For observations below the 50th percentile of `seplen`, the probability of predicting `Setosa` is around 1 when `seplen` < 7 and goes down afterward. For observations in the higher percentiles of `seplen`, the probability of predicting `Setosa` is close to 0. PDP, the dashed black line, is an average of ICE curves for all observations.

◁

References

Goldstein, A., A. Kapelner, J. Bleich, and E. Pitkin. 2015. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics* 24: 44–65. https://doi.org/10.1080/10618600.2014.907095.

Krishna, S., T. Han, A. Gu, S. Wu, S. Jabbari, and H. Lakkaraju. 2022. The disagreement problem in explainable machine learning: A practitioner's perspective. arXiv:2202.01602 [cs.LG], https://doi.org/10.48550/arXiv.2202.01602.

Lakkaraju, H., and O. Bastani. 2020. "“How do I fool you?”: Manipulating user trust via misleading black box explanations". In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 79–85. New York: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375833.

Slack, D., S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju. 2020. "Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods". In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 180–186. New York: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375830.

Also see

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

[H2OML] **h2omlgraph pdp** — Produce partial dependence plot+