R

Which versions of R are compatible with H2O?

Currently, the only version of R that is known to NOT work well with H2O is R version 3.1.0 (codename “Spring Dance”). If you are using this version, we recommend upgrading the R version before using H2O.

What R packages are required to use H2O?

The following packages are required:

- `methods`
- `statmod`
- `stats`
- `graphics`
- `RCurl`
- `jsonlite`
- `tools`
- `utils`

Some of these packages have dependencies; for example, `bitops` is required, but it is a dependency of the `RCurl` package, so `bitops` is automatically included when `RCurl` is installed.

If you are encountering errors related to missing R packages when using H2O, refer to the following list for a complete list of all R packages, including dependencies:

- `statmod`
- `bitops`
- `RCurl`
- `jsonlite`
- `methods`
- `stats`
- `graphics`
- `tools`
- `utils`
- `stringi`
- `magrittr`
- `colorspace`

- `stringr`
- `RColorBrewer`
- `dichromat`
- `munsell`
- `labeling`
- `plyr`
- `digest`
- `gtable`
- `reshape2`
- `scales`
- `proto`
- `ggplot2`
- `h2oEnsemble`
- `gtools`
- `gdata`
- `caTools`
- `gplots`
- `chron`
- `ROCR`
- `data.table`
- `cvAUC`

Finally, if you are running R on Linux, then you must install `libcurl`, which allows H2O to communicate with R.

How can I install the H2O R package if I am having permissions problems?

This issue typically occurs for Linux users when the R software was installed by a root user. For more information, refer to the following link.

To specify the installation location for the R packages, create a file that contains the `R_LIBS_USER` environment variable:

```
echo R_LIBS_USER=\"~/.Rlibrary\" > ~/.Renviron
```

Confirm the file was created successfully using `cat`:

```
$ cat ~/.Renviron
```

You should see the following output:

```
R_LIBS_USER="~/.Rlibrary"
```

Create a new directory for the environment variable:

```
$ mkdir ~/.Rlibrary
```

Start R and enter the following:

```
.libPaths()
```

Look for the following output to confirm the changes:

```
[1] "<Your home directory>/.Rlibrary"
[2] "/Library/Frameworks/R.framework/Versions/3.1/Resources/library"
```

I received the following error after launching H2O in RStudio and using `h2o.init`. What should I do to resolve this error?

```
Error in h2o.init() :
Version mismatch! H2O is running version 3.30.0.6 but R package is version 3.28.1.3
```

This error is due to a version mismatch between the H2O R package and the running H2O instance. Make sure you are using the latest version of both files by downloading H2O from the downloads page and installing the latest version and that you have removed any previous H2O R package versions by running:

```
if ("package:h2o" %in% search()) { detach("package:h2o", unload=TRUE) }
if ("h2o" %in% rownames(installed.packages())) { remove.packages("h2o") }
```

Make sure to install the dependencies for the H2O R package as well:

```
if (! ("methods" %in% rownames(installed.packages()))) { install.packages("methods") }
if (! ("statmod" %in% rownames(installed.packages()))) { install.packages("statmod") }
if (! ("stats" %in% rownames(installed.packages()))) { install.packages("stats") }
if (! ("graphics" %in% rownames(installed.packages()))) { install.packages("graphics") }
if (! ("RCurl" %in% rownames(installed.packages()))) { install.packages("RCurl") }
if (! ("jsonlite" %in% rownames(installed.packages()))) { install.packages("jsonlite") }
if (! ("tools" %in% rownames(installed.packages()))) { install.packages("tools") }
if (! ("utils" %in% rownames(installed.packages()))) { install.packages("utils") }
```

Finally, install the latest stable version of the H2O package for R:

```
install.packages("h2o", type = "source", repos = (c("http://h2o-
release.s3.amazonaws.com/h2o/latest_stable_R")))
library(h2o)
h2o.init()
```

If your R version is older than the H2O R package, upgrade your R version using `update.packages(checkBuilt = TRUE, ask = FALSE)`.

I received the following error message after launching H2O in RStudio and using `h2o.init`. What should I do to resolve this error?

```
Server error - server 127.0.0.1 is unreachable at this moment.
Please retry the request or contact your administrator.
```

This error occurs when the proxy is set in your R environment. The resolution is to unset that so that you can access localhost from within R. Run the following to unset the proxy:

```
Sys.unsetenv("http_proxy")
Sys.unsetenv("https_proxy")
Sys.unsetenv("http_proxy_user")
Sys.unsetenv("https_proxy_user")
```

I received the following error message after trying to run some code. What should I do?

```
> fit <- h2o.deeplearning(x = 2:4, y = 1, training_frame = train)

|=============================================================================================================
 100%
Error in model$training_metrics$MSE :
  $ operator not defined for this S4 class
In addition: Warning message:
Not all shim outputs are fully supported, please see ?h2o.shim for more information
```

Remove the `h2o.shim(enable = TRUE)` line and try running the code again. Note that the `h2o.shim` is only a way to notify users of previous versions of H2O about changes to the H2O R package - it will not revise your code, but provides suggested replacements for deprecated commands and parameters.

How do I extract the model weights from a model I've created using H2O in R? I've enabled `extract_model_weights_and_biases`, but the output refers to a file I can't open in R.

For an example of how to extract weights and biases from a model, refer to the following repo location on GitHub.

How do I extract the run time of my model as output?

For the following example:

```
rf <- h2o.randomForest(x = c("x1", "x2", "x3", "w"), y = "y", training_frame = train)
```

Use `rf@model$run_time` to determine the value of the `run_time` variable.

What is the best way to do group summarizations? For example, getting sums of specific columns grouped by a categorical column.

We strongly recommend using `h2o.group_by` for this function instead of `h2o.ddply`, as shown in the following example:

```
newframe <- h2o.group_by(h2oframe, by = "footwear_category", nrow("email_event_click_ct"),
sum("email_event_click_ct"), mean("email_event_click_ct"), sd("email_event_click_ct"), gb.control =
list(col.names=c("count", "total_email_event_click_ct", "avg_email_event_click_ct",
"std_email_event_click_ct")))
```

Using `gb.control` is optional; here it is included so the column names are user-configurable.

The `by` option can take a list of columns if you want to group by more than one column to compute the summary as shown in the following example:

```
newframe <- h2o.group_by(h2oframe, by = c("footwear_category","age_group"),
nrow("email_event_click_ct"), sum("email_event_click_ct"), mean("email_event_click_ct"),
sd("email_event_click_ct"), gb.control = list( col.names = c("count", "total_email_event_click_ct",
"avg_email_event_click_ct", "std_email_event_click_ct")))
```

I'm using Linux and I want to run H2O in R. Are there any dependencies I need to install?

Yes, make sure to install `libcurl`, which allows H2O to communicate with R. We also recommend disabling SElinux and any firewalls, at least initially until you have confirmed H2O can initialize.

- On Ubuntu, run: `apt-get install libcurl4-openssl-dev`
- On CentOS, run: `yum install libcurl-devel`

How do I change variable/header names on an H2O frame in R?

There are two ways to change header names. To specify the headers during parsing, import the headers in R and then specify the header as the column name when the actual data frame is imported:

```
header <- h2o.importFile(path = pathToHeader)
data   <- h2o.importFile(path = pathToData, col.names = header)
data
```

You can also use the `names()` function:

```
header <- c("user", "specified", "column", "names")
data   <- h2o.importFile(path = pathToData)
names(data) <- header
```

To replace specific column names, you can also use a `sub/gsub` in R:

```
header <- c("user", "specified", "column", "names")
## I want to replace "user" column with "computer"
data   <- h2o.importFile(path = pathToData)
names(data) <- sub(pattern = "user", replacement = "computer", x = names(header))
```

My R terminal crashed. How can I re-access my H2O frame?

Launch H2O and use your web browser to access the web UI, Flow, at `localhost:54321`. Click the **Data** menu, then click **List All Frames**. Copy the frame ID, then run `h2o.ls()` in R to list all the frames, or use the frame ID in the following code (replacing `YOUR_FRAME_ID` with the frame ID):

```
library(h2o)
h2o.init(startH2O = FALSE, strict_version_check = TRUE)
data_frame <- h2o.getFrame(frame_id = "YOUR_FRAME_ID")
```

How do I remove rows containing NAs in an H2OFrame?

To remove NAs from rows:

```
  a  b  c  d  e
1 0 NA NA NA NA
2 0  2  2  2  2
3 0 NA NA NA NA
4 0 NA NA  1  2
5 0 NA NA NA NA
6 0  1  2  3  2
```

Removing rows 1, 3, 4, 5 to get:

```
  a  b  c  d  e
2 0  2  2  2  2
6 0  1  2  3  2
```

Use `na.omit(myFrame)`, where `myFrame` represents the name of the frame you are editing.

I installed H2O in R using OS X and updated all the dependencies, but the error below message displays: What should I do?

Error message:

`Error in .h2o.doSafeREST(h2oRestApiVersion = h2oRestApiVersion, Unexpected CURL error: Empty reply from server`.

This error message displays if the `JAVA_HOME` environment variable is not set correctly. The `JAVA_HOME` variable is likely points to Apple Java version 6 instead of Oracle Java version 8.

If you are running OS X 10.7 or earlier, enter the following in Terminal:

`export JAVA_HOME=/Library/Internet\ Plug-Ins/JavaAppletPlugin.plugin/Contents/Home`

If you are running OS X 10.8 or later, modify the launchd.plist by entering the following in Terminal:

```
cat << EOF | sudo tee /Library/LaunchDaemons/setenv.JAVA_HOME.plist
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN" "http://www.apple.com/DTDs/PropertyList-
1.0.dtd">
  <plist version="1.0">
  <dict>
  <key>Label</key>
  <string>setenv.JAVA_HOME</string>
  <key>ProgramArguments</key>
  <array>
    <string>/bin/launchctl</string>
    <string>setenv</string>
    <string>JAVA_HOME</string>
    <string>/Library/Internet Plug-Ins/JavaAppletPlugin.plugin/Contents/Home</string>
  </array>
  <key>RunAtLoad</key>
  <true/>
  <key>ServiceIPC</key>
  <false/>
</dict>
</plist>
EOF
```

How does the `col.names` argument work in `group_by`?

You need to add the `col.names` inside the `gb.control` list. Refer to the following example:

```
newframe <- h2o.group_by(dd, by = "footwear_category", nrow("email_event_click_ct"),
sum("email_event_click_ct"), mean("email_event_click_ct"),
    sd("email_event_click_ct"), gb.control = list(col.names = c("count",
"total_email_event_click_ct", "avg_email_event_click_ct", "std_email_event_click_ct")))
newframe$avg_email_event_click_ct2 = newframe$total_email_event_click_ct / newframe$count
```

How are the results of `h2o.predict` displayed?

The order of the rows in the results for `h2o.predict` is the same as the order in which the data was loaded, even if some rows fail (for example, due to missing values or unseen factor levels). To bind a per-row identifier, use `cbind`.

How do I view all the variable importances for a model?

By default, H2O returns the top five and lowest five variable importances. To view all the variable importances, use the following:

```
model <- h2o.getModel(model_id = "my_H2O_modelID")

varimp <- as.data.frame(h2o.varimp(model))
```

How do I add random noise to a column in an H2O frame?

To add random noise to a column in an H2O frame, refer to the following example:

```
h2o.init()

fr <- as.h2o(iris)

  |======================================================================| 100%

random_column <- h2o.runif(fr)

new_fr <- h2o.cbind(fr, random_column)

new_fr
```