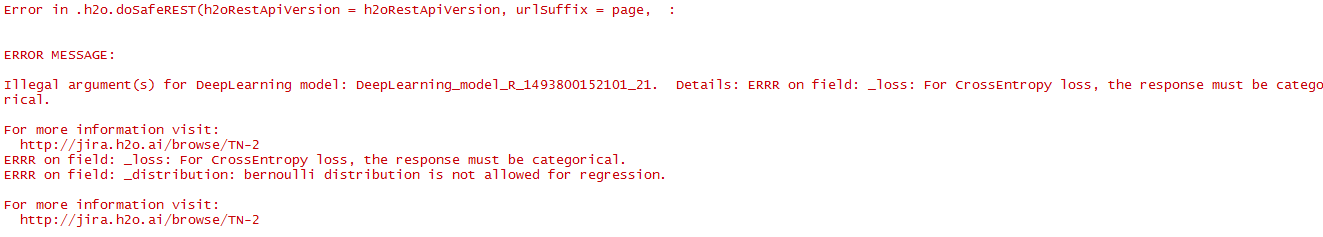
问题1

Q：在用洪鼎的代码进行分类问题的时候遇到了如下问题：



查找网页http://jira.h2o.ai/browse/TN-2

得到如下解释：

Certain combinations of distributions/families and response types don't make sense together. For example, if you specify the model to be built using a "multinomial" distribution, then the response must be categorical (i.e, "cat", "dog", "mouse), and not numerical (i.e., -19...15).

All supervised H2O models (GLM/GBM/DRF/DL) require the user to specify a "correct" response type for the given set of parameters. If not explicit model type parameter is available, as for DRF, then the type of problem is inferred from the response type. For classification problems, for example, the response must be converted explicitly to a categorical column. Note that simply specifying an integer column with N unique values is not sufficient, as it could also be a numerical column with very few non-integer "outliers" that are absent in the training data (and during test time, there might be a new non-integer number, etc.). To avoid these kinds of issues, H2O uses the response column type as the highest-priority indicator for the problem type.

The only exception to this rule is GLM, which will accept any strictly binary response for the "binomial" family.

分布/家庭和响应类型的某些组合在一起是没有意义的。例如，如果您使用“多项式”分布指定要构建的模型，则响应必须是分类的（即“猫”，“狗”，“鼠标”），而不是数字（即-19 ...） 15）。

所有监督的H2O模型（GLM / GBM / DRF / DL）要求用户为给定的一组参数指定“正确”的响应类型。如果没有明确的模型类型参数可用，对于DRF，则从响应类型推断问题的类型。对于分类问题，例如，响应必须明确转换为分类列。注意，简单地指定具有N个唯一值的整数列是不够的，因为它也可以是训练数据中不存在非常少的非整数“离群值”的数值列（并且在测试时间期间，可能存在新的非整数等）。为了避免这些问题，H2O使用响应列类型作为问题类型的最高优先级指标。

此规则的唯一例外是GLM，它将接受“二项式”族的任何严格的二进制响应。

A：将分类模型中y的0和1值分别改为A/B，程序可以跑通。

问题2

Q：在进行格点搜索的时候出现这样的错误：



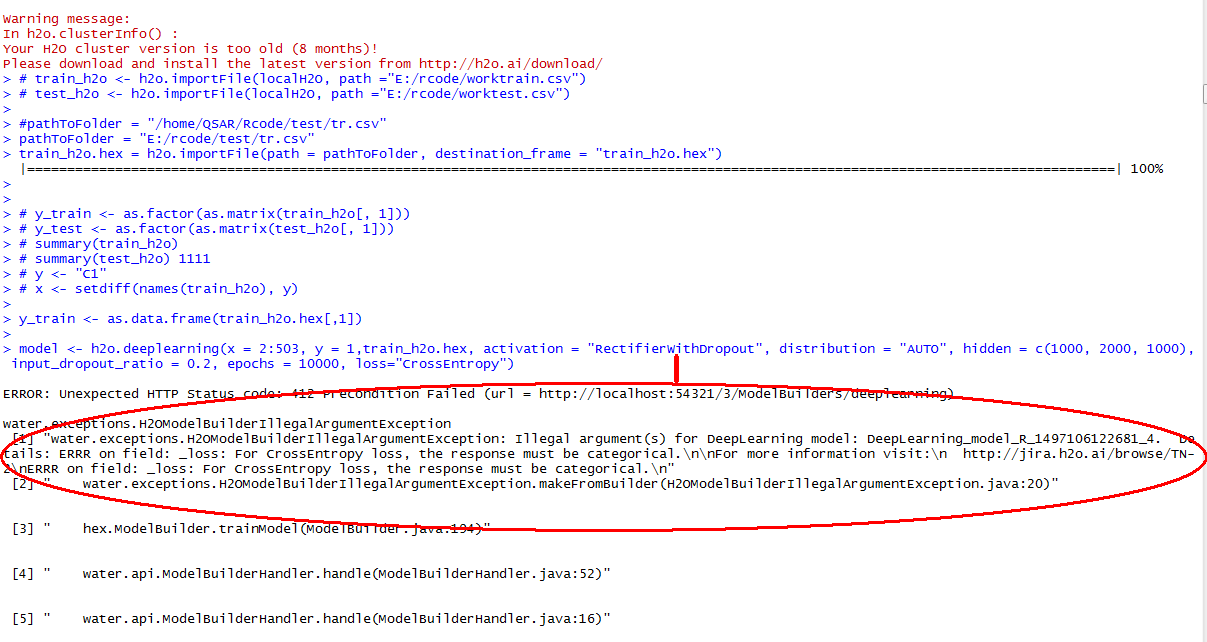
找到以下原因：

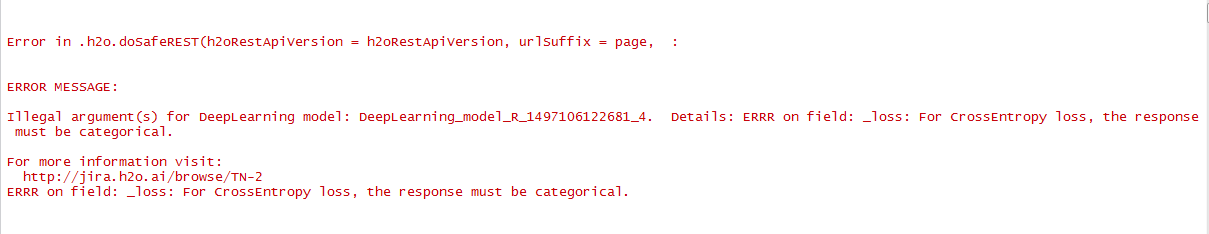


A：格点搜索要用服务器才能跑，要用deepwater。

问题3

回归模型的缺失函数loss不能用交叉熵分布“CrossEntropy”会报错，如下：





进行loss参数的调整，可用的缺失函数有：

Absolute、Automatic、Quatratic、Quantile。

R中可用的缺失函数有：

