Deep Learning Package in R
An Introduction to H2O

Yue Jiang

Hong Kong University of Science and Technology

April 28, 2017

Overview

Outline

Introduction

- H2O is an open-source math and in-memory prediction engine for big data science developed by Oxdata.
- Computes parallel distributed machine learning algorithms within various cluster environments.
- https://www.h2o.ai/resources/
- generalized linear models, gradient boosting machines, random forests, and neural networks (deep learning)

Introduction

- H2O flow: notebook-style open-source user interface, similar to Jupyter Notebook
- H2O on Python
- H2O on R

Outline

R Implement

- Installation: install directly from CRAN.
  https://cran.r-project.org/web/packages/h2o/index.html
- Initialization: h2o.init(...) function
  - Set up and connect to a H2O cluster from R
  - By default, H2O starts a cluster on your local machine using all available threads
  - All computations are performed (in highly optimized Java code) in the H2O cluster
  - Can also provide IP address and port number to connect to a remote cluster
- h2o data frame: as.h2o(...) function
- deep learning: h2o.deeplearning(...) function

Outline

Financial Prediction Example

- Original data: open/close/high/low/volume data of a stock
- Input
  - last n one-minute pseudo-log-returns
  - last n standard deviation of prices
  - current time (hour and minute)
- 5 hidden layers: 8, 6, 4, 3, 1
- Activation function: Tanh
- Output: Next one-minute pseudo-log-return
- R code example

Thank you!