Caret: Select Most Dissimilar Models For Evaluation

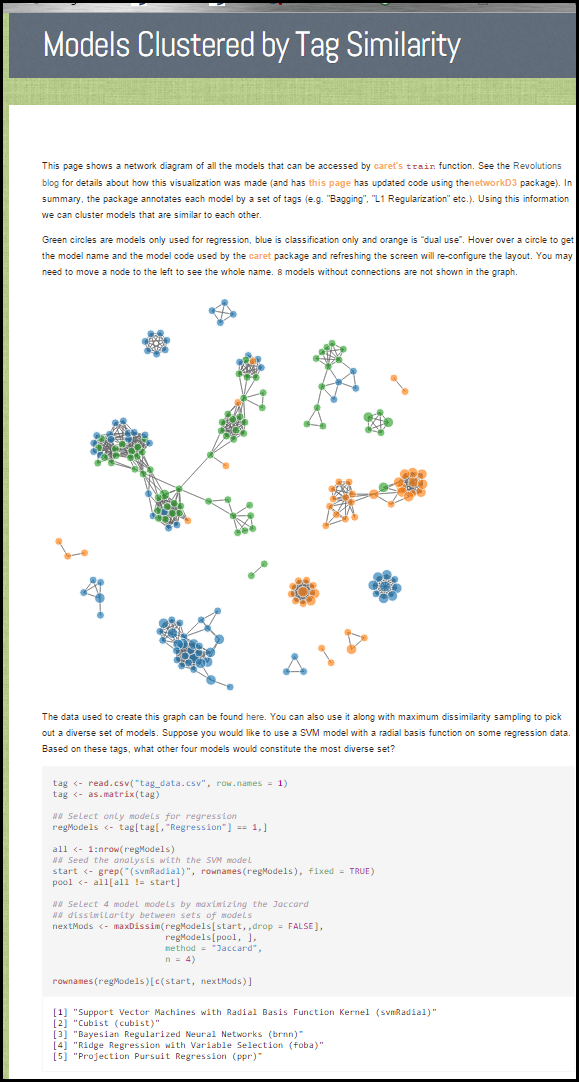
Neal Cariello

8 Jan 2016

From:

<http://topepo.github.io/caret/similarity.html>

Code on web page to select dissimilar models given. Download tag\_data.csv on 8 Jan 2016.



tag <- **read.csv**("tag\_data.csv", row.names = 1)

tag <- **as.matrix**(tag)

*## Select only models for regression*

regModels <- tag[tag[,"Regression"] == 1,]

all <- 1:**nrow**(regModels)

*## Seed the analysis with the SVM model*

start <- **grep**("(svmRadial)", **rownames**(regModels), fixed = TRUE)

pool <- all[all != start]

*## Select 4 model models by maximizing the Jaccard*

*## dissimilarity between sets of models*

nextMods <- **maxDissim**(regModels[start,,drop = FALSE],

regModels[pool, ],

method = "Jaccard",

n = 4)

**rownames**(regModels)[**c**(start, nextMods)]