

Análisis predictivo en R

(II)

SISTEMAS INTELIGENTES PARA LA GESTIÓN DE LA EMPRESA
CURSO 2016-2017

Bagging vs Boosting

Bagging

- Crear clasificadores simples e **independientes** con muestras del conjunto de entrenamiento

Boosting

- Crear clasificadores **dependientes** de forma iterativa utilizando una función de optimización global

Combinar la salidas utilizando:

1. Voto por mayoría
2. Voto ponderado

(* asignación de probabilidades, definición de umbrales, etc.)

Bagging

Bagging con caret

<http://topepo.github.io/caret/train-models-by-tag.html#bagging>

- Random Forests

- http://topepo.github.io/caret/train-models-by-tag.html#Random_Forest

- <https://www.linkedin.com/pulse/different-random-forest-packages-r-madhur-modi>

- `party`: method = `'cforest'`
 - `ranger`: method = `'ranger'`
 - `randomForest`: method = `'rf'`
 - `extraTrees`: method = `'extraTrees'`
 - `randomForest, inTrees`: method = `'rfRules'`
 - `randomForest`: method = `'parRF'`
- Bagged CART
 - `ipred`: method = `'treebag'`
- Model Averaged Neural Network
 - `nnet`: method = `'avNNET'`

Boosting

Boosting con caret

<http://topepo.github.io/caret/train-models-by-tag.html#boosting>

- AdaBoost Classification Trees
 - `fastAdaboost`: method = `'adaboost'`
 - `adabag`: method = `'AdaBoost.M1'`
 - `adabag`: method = `'AdaBag'`
 - `ada`: method = `'ada'`
- Boosted Tree
 - `party`, `mboost`: method = `'blackboost'`
 - `bst`: method = `'bstTree'`
 - `C50`: method = `'C5.0'`
 - `C50`: method = `'C5.0Cost'`
- eXtreme Gradient Boosting
 - `xgboost`: method = `'xgbTree'`

xgboost

Documentación

- General: <http://xgboost.readthedocs.io/en/latest/>
- Fundamentos teóricos: <http://xgboost.readthedocs.io/en/latest/model.html>
- Tutoriales: <https://github.com/dmlc/xgboost/tree/master/demo>
- Parámetros: <http://xgboost.readthedocs.io/en/latest//parameter.html>

Dataset - carcrash

Información sobre personas involucradas en accidentes de tráfico

<http://datamarket.azure.com/dataset/bigml/carcrashusa2011>

Objetivo: Predecir el “tipo” de un accidente

Variables (14):

- Age
- Alcohol_Results
- Atmospheric_Condition
- Crash_Date
- Drug_Involvement
- Fatalities_in_crash
- First_Harmful_Event
- Gender
- ID
- Injury_Severity
- Person_Type
- Race
- Roadway
- State