Class

Tree Based Models

Topic

Tree Based Ensembles: Adaboost and Gradient Boosting

Adaboost

Boosting – Another ensemble technique built using decision trees as base learners

Boosted trees works differently than Bagged trees and Random forests

	Boosted Trees	Bagged Trees	Random Forests
Used to build an ensemble	Data re-weighing strategy	Bootstrapped samples	
Depth of tree	Not large (2 to 3 levels)	As many required	

Adaboost is a popular technique

Data Re-weighing Strategy

Classification task using a data set

Wherever the model makes a mistake, that row is given more importance

Data re-weighing

Data re-weighing

X	Y
...	1
...	0
...	1

→ Tree Model T1 →

X	Y	Y'
...	1	0
...	0	1
...	1	1

→ Tree Model T2 →

X	Y	Y'
...	1	1
...	0	1
...	1	1

→ Tree Model Tn

Row 1 and 2 are given more weight

Row 2 is given more weight

Tree models will be shallow

Final model is a combination of these trees (T1, T2,... Tn)

Data Re-weighing Strategy

Classification task using a data set

Wherever the model makes a mistake, that row is given more importance

Data re-weighing

Data re-weighing

X	Y
...	1
...	0
...	1

Tree Model T1

X	Y	Y'
...	1	0
...	0	1
...	1	1

Tree Model T2

X	Y	Y'
...	1	1
...	0	1
...	1	1

Tree Model Tn

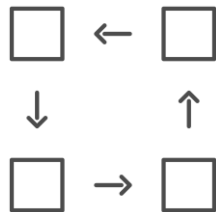Row 1 and 2 are given more weight

Row 2 is given more weight

Re-weighing strategy - each successive tree pays more attention to the parts of the data that preceding trees have failed to correctly predict

Successive trees try to improve the error rate

Gradient Boosting

Gradient Boosting is another popular boosting technique

Regression task using a dataset

Yet, the mechanics of the discussions are valid in a classification task

Gradient boosting is an iterative algorithm

Gradient Boosting

Step 1

Predictions

X	Y
30	32
48	62
19	23
22	25

Data set with 1 predictor and 1 target variable

T1

Simple tree model

X	Y	Y'	Residual
30	32	31	1
48	62	60	2
19	23	24	-1
22	25	26	-1

Error

T2

Error – difference between the actual variable and predicted variable

Step 2

X	Y
30	32
48	62
19	23
22	25

T1+T2

Combination of 2 tree models

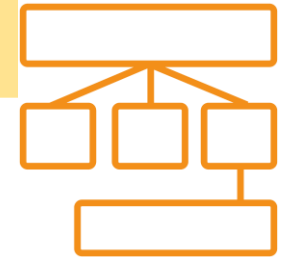| X	Y	Y'	Residual
30	32	92.5	0.9
48	62	61	1
19	23	23.5	-0.5
22	25	25.5	-0.5

T3

Gradient Boosting

Keep on repeating this process quite a few times and eventually end up with an ensemble of trees

Gradient boosting is a general ensemble framework

Boosting trees can use other base learner other than decision tree

Partial Dependence Plot

Not as interpretable as simple models like decision trees or linear models

Ensembles provide a list of important predictors by computing variable importance measures

Able to calculate the variables that are important predictors

What is the direction of impact a given predictor has on the dependent variable?

Is the given predictor positively or negatively impacting the dependant variable?

Partial Dependence Plot

Partial Dependence Plot – helps in understanding relationships between a dependent variable and an independent variable

Help in establishing the direction of impact of a predictor on target variable

Depending on which machine learning framework is being used, partial dependence plots for the ensembles may or may not be supported

Drawbacks

Only bivariate relationships can be understood but unearthing interaction effects can be difficult

Creating partial dependence plot is computationally expensive

Partial Dependence Plot

Partial dependence plots helps identifying the relationship between value of target and the value of a predictor variable after considering the effect of all the other variables

Y axis = values of target variable

Plot 1 and 2 - Value of target variable doesn't change

Plot 3 - Positive relationship between the target variable and the predictor variable
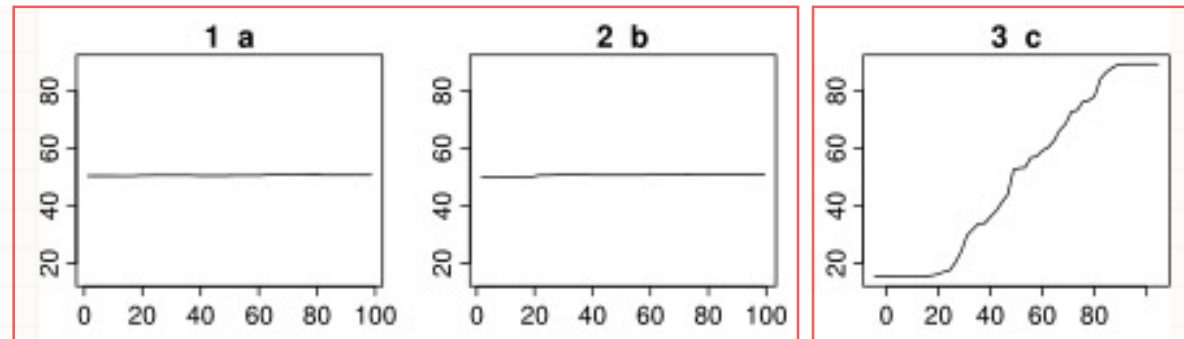
3 partial dependence plots

X axis = values of a predictor

Code Demo

Recap

- Adaboost
- Data re-weighing strategy
- Gradient boosting
- Partial dependence plot
- Code demo