

Supplementary material for article” Extraction of Liana Stems from Terrestrial LiDAR Data using Geometric Features”

Tao Han, G. Arturo Sánchez-Azofeifa*

Centre for Earth Observation Sciences, Department of Earth and Atmospheric Sciences,
University of Alberta, Edmonton, Alberta, Canada.

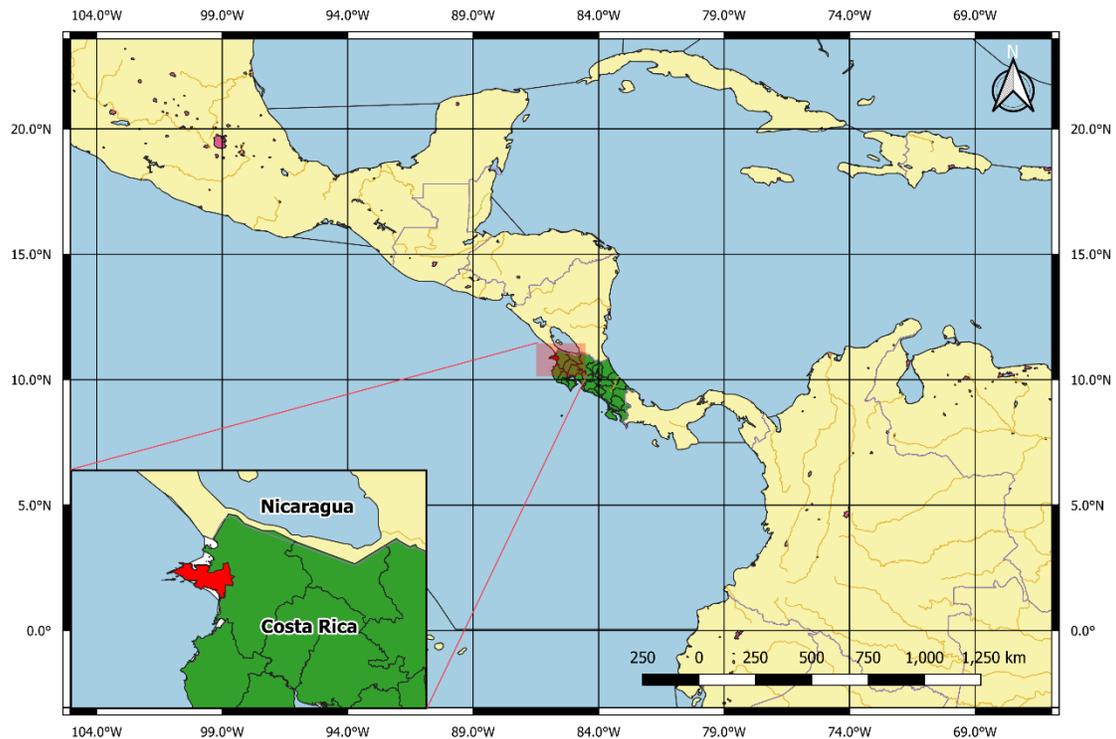


Figure S1. Study area, the red area indicates the Santa Rosa National Park – Environmental Monitoring Supersite (SRNP-EMSS), Guanacaste, Costa Rica.

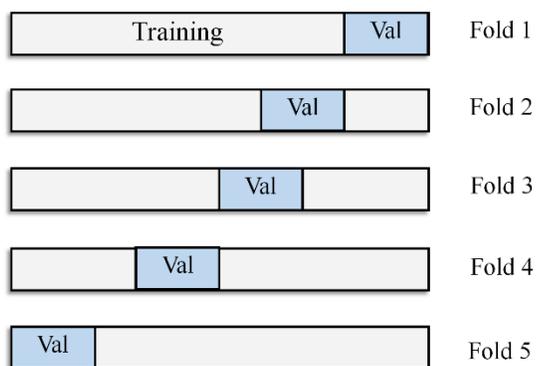
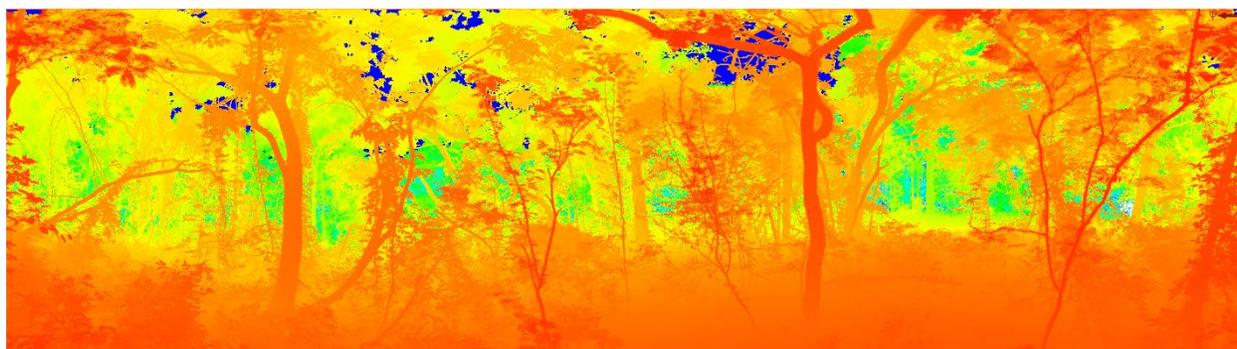
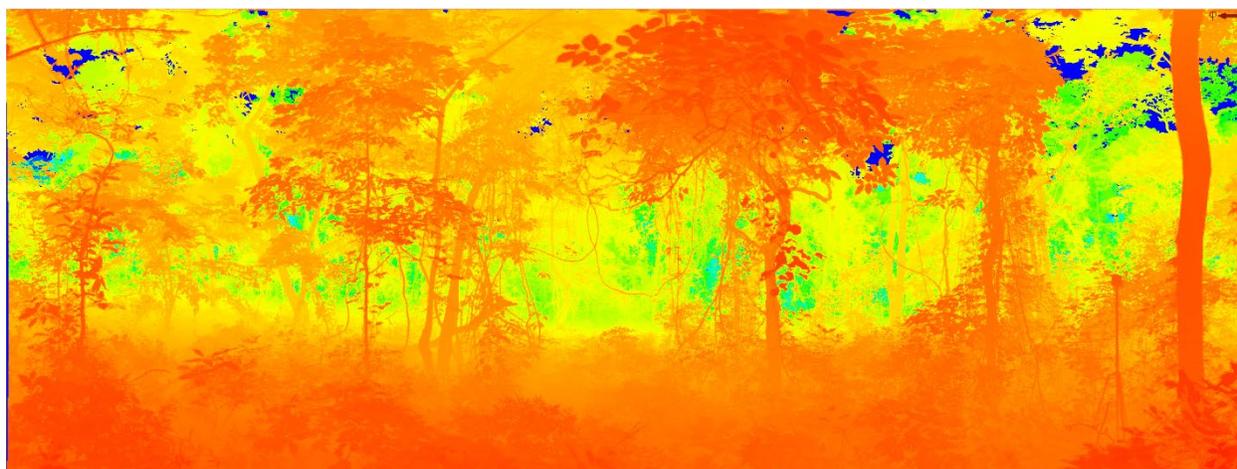


Figure S2. Five-fold spatial cross-validation approach for our data.

Environment 1



Environment 2



Environment 3

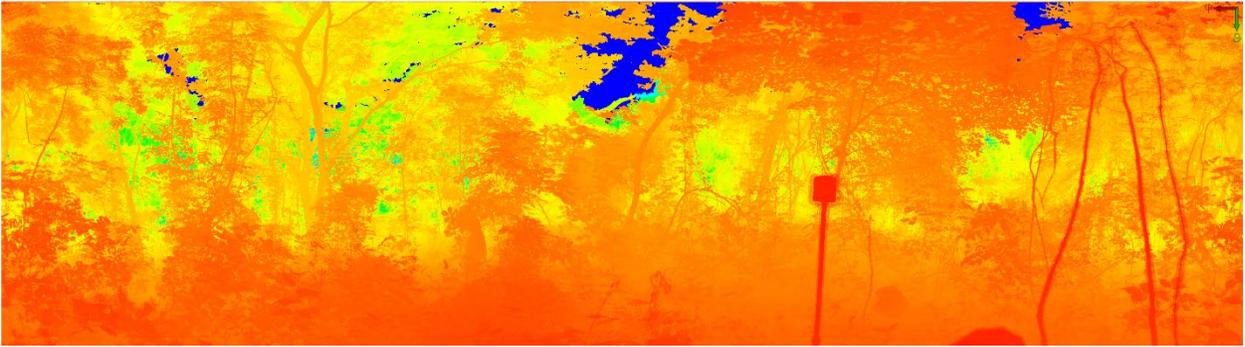


Figure S3. The examples of local environment of the plot where collected the point clouds



Figure S4. The actual photos where we scan our trees.

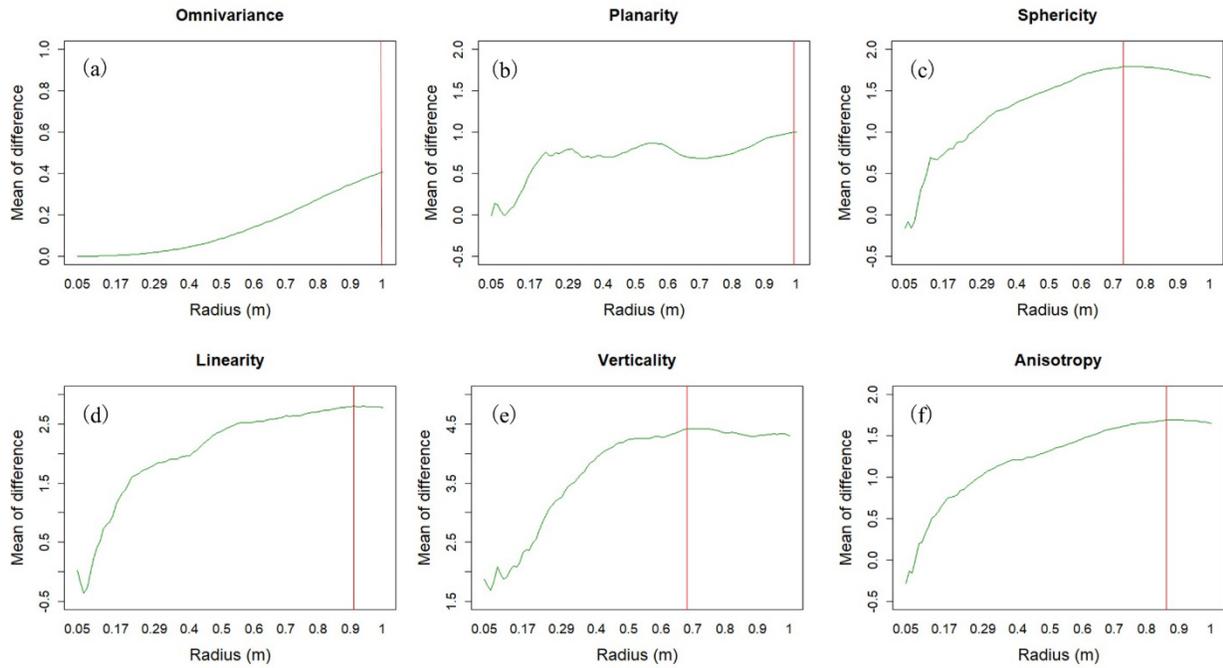


Figure S5. The mean difference of Omnivariance, Planarity, Sphericity, Linearity, Verticality, and Anisotropy between liana and tree for Tree 2 across the various radius (y-axis indicate the mean difference between liana and tree, and x-axis shows the value of radius from 0.05 to 1 m; red vertical line means the largest difference of the feature).

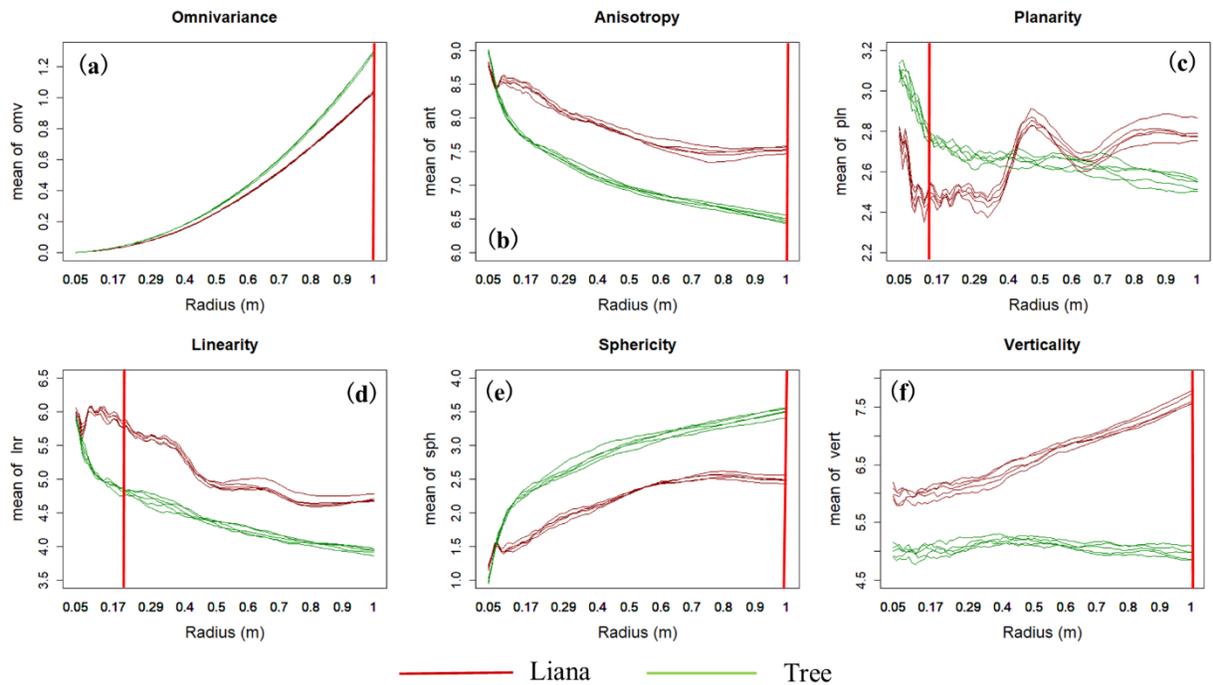


Figure S6. The changing trends of Omnivariance, Anisotropy, Planarity, Linearity, Sphericity and Verticality for Tree 1 across the various radius (y-axis indicates the mean value of each feature, and the x-axis shows the value of radius from 0.05 to 1 m; Red line: Tree; Green line: Liana); Red vertical line indicates the optimal radius used in this study (see Table 3).

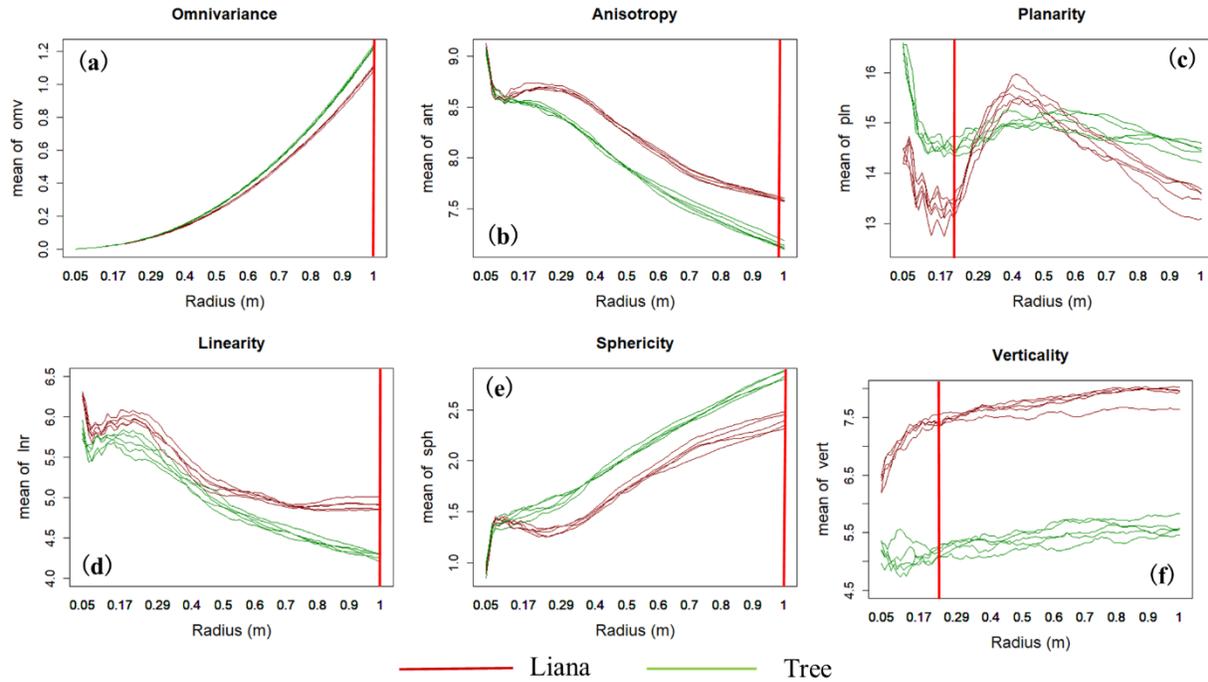


Figure S7. The changing trends of Omnivariance, Anisotropy, Planarity, Linearity, Sphericity and Verticality for Tree 3 across the various radius (y-axis indicates the mean value of each feature, and the x-axis shows the value of radius from 0.05 to 1 m; Red line: Tree; Green line: Liana); Red vertical line indicates the optimal radius used in this study (see Table 3).

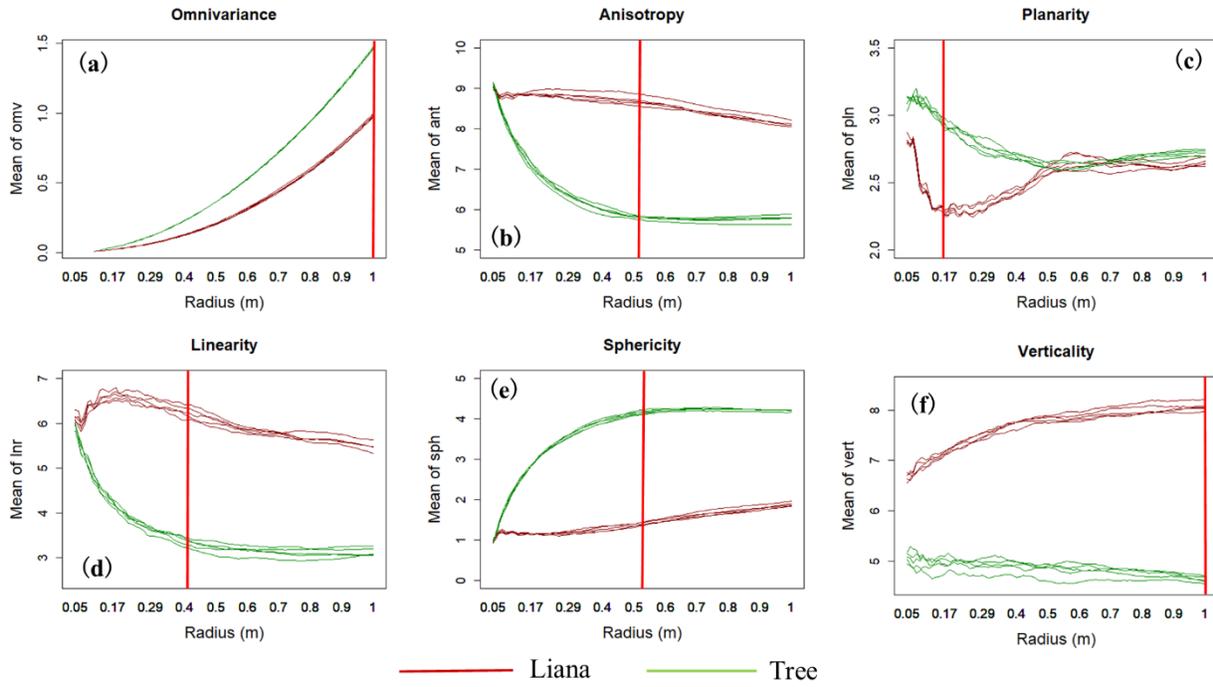


Figure S8. The changing trends of Omnivariance, Anisotropy, Planarity, Linearity, Sphericity and Verticality for Tree 4 across the various radius (y-axis indicates the mean value of each feature, and the x-axis shows the value of radius from 0.05 to 1 m; Red line: Tree; Green line: Liana); Red vertical line indicates the optimal radius used in this study (see Table 3).

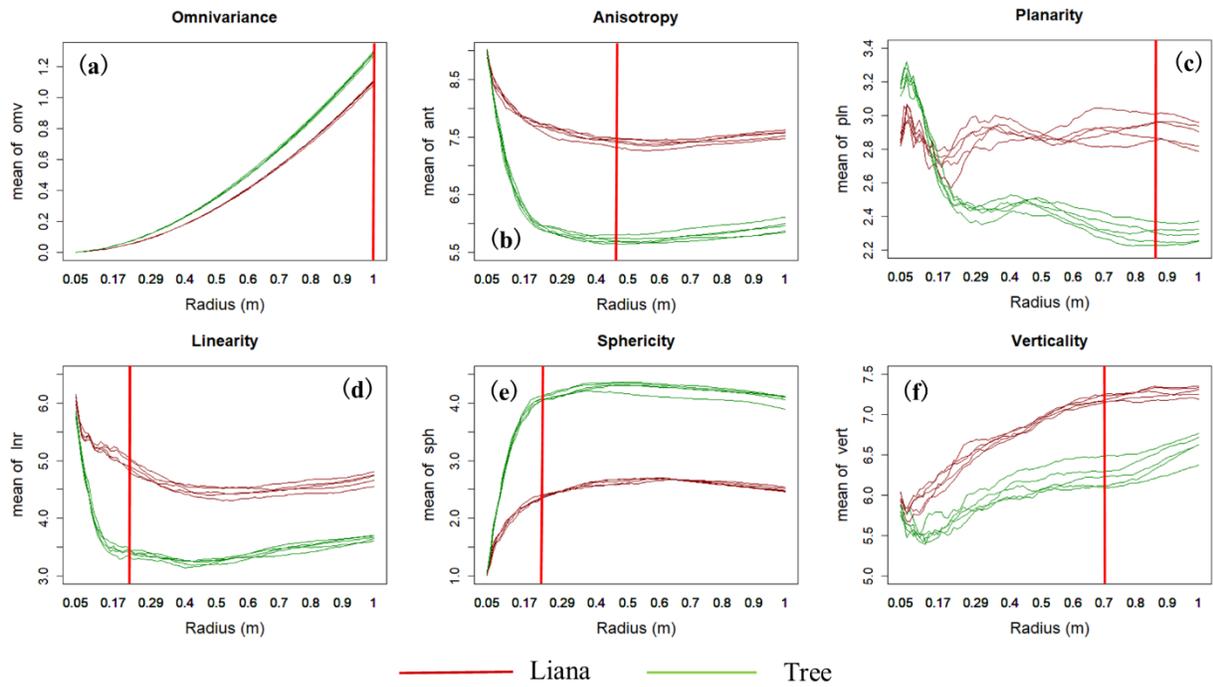


Table S2. The number of clusters generated by DBSCAN on our data.

Tree ID	Liana	Tree	Total
Tree 1	966	11712	12678
Tree 2	75	1568	1643
Tree 3	4006	12293	16299
Tree 4	1768	16124	17892
Tree 5	169	2019	2188

Table S3. Feature importance of Random Forest (RF) and eXtreme Gradient Boosting (XGBoosting) models.

Features	Random Forest		XGBoosting
	MeanDecreaseAccuracy	MeanDecreaseGini	Gain
Sphericity	25.19	40939.35	0.41
Verticality	45.55	32676.12	0.28
Linearity	14.17	1801.34	0.05
Omnivariance	10.32	4514.18	0.12
Anisotropy	9.16	7222.51	0.09
Planarity	3.01	247.83	0.05

Table S3 indicates the feature importance of RF and XGBoosting models. There are two importance indices for the RF model: (1) MeanDecreaseAccuracy, which illustrates how much accuracy the model loses when excluding each variable; (2) MeanDecreaseGini, which measures how each variable contributes to the purity of the nodes and leaves in RF model. The higher value of MeanDecreaseAccuracy and MeanDecreaseGini, the higher importance of the variable in the model. Concerning the XGBoosting model, Gain represents the relative contribution of each feature to the model, which can be obtained by calculating each feature's contribution for each tree in the model. A higher Gain value means the feature is more important to making a prediction. According to Table S3, Sphericity and Verticality are the most important features in our RF and XGBoosting models. By contrast, the least important feature in the RF model is Planarity, while that in the XGBoosting model are Planarity and Linearity.

Hyperparameter optimization of Random Forest and XGBoosting algorithms

A: Random Forest:

"ntree", lower = 10, upper = 200, the number of decision tree in the forest.

"mtry", lower = 1, upper = 6, the number of features in each node.

"nodesize",lower = 1,upper = 5, the minimum number of the data allowed in a leaf.

"maxnodes",lower = 5,upper = 20, the maximum number of leaves.

Combination for Random Forest:

ntree	84
mtry	5
nodezsize	3
maxnode	20

5-fold cross-validation mean classification error for RF: 0.188.

B: XGBoosting:

"eta",lower = 0,upper = 1, learning rate.

"gamma",lower = 0,upper = 5, the minimum number of splitting.

"max_depth",lower = 1,upper = 5, the maximum depth of a decision tree.

"min_child_weight",lower = 1,upper = 10, the minimum total weight of all the data in a leaf node

"subsample",lower = 0.5,upper = 1, the proportion of the data to be randomly sampled.

"colsample_bytree",lower = 0.5,upper = 1, the fraction of predictors sampled by each tree.

"nrounds",lower = 10,upper = 200, the sequentially built trees to be used.

"eval_metric",values = c("merror", "auc"), the loss function used in the model.

Combination for XGBoosting:

eta	0.643
gamma	3.98
max_depth	5
min_child_weight	4.31
subsample	0.864
colsample_bytree	0.952
nround	188
eval_metric	auc

5-fold cross-validation mean classification error for XGBoosting: 0.151.