Hyperparameter Tuning

Kaggle Datasets: 1) Insurance fraud detection, 2) Portuguese Bank

SOWMIYA MURUGANANDAM

Hyperparameter Tuning

Process to determine the right combination of "Hyperparameters" to maximize model performance

ADVANTAGES :

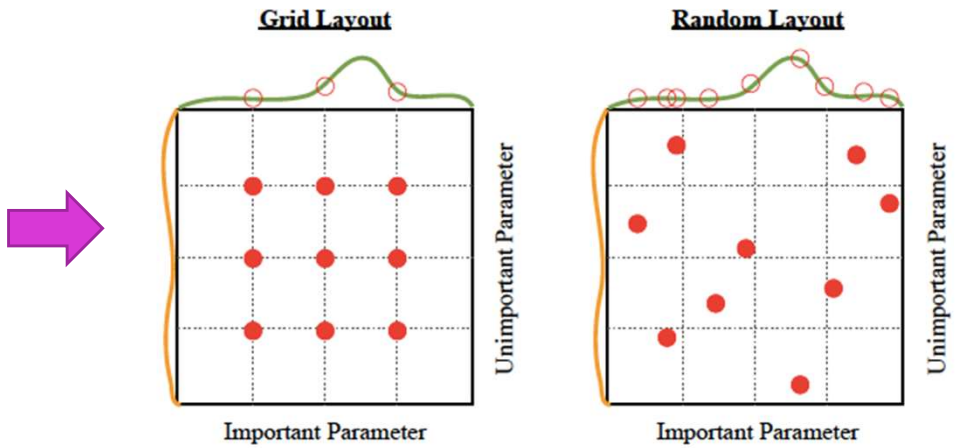
1. IMPROVES ACCURACY
2. REDUCES OVERFITTING AND UNDERFITTING

A FEW METHODS:

1. GRID SEARCH
2. RANDOM SEARCH
3. BAYESIAN OPTIMIZATION

Grid Vs Random Search

- Easy to Implement
- Easily Parallelized
- Computation time is higher



Source

- Easy to Implement
- Easily Parallelized
- Computation time is better than Grid search

Insurance Fraud Detection

Methods used to build the model

1. Decision Tree Classifiers
2. Random Forest Classifier

Methods used for Hyperparameter Tuning

1. Random Search
2. Grid Search
3. Cross Validation

Best Classifier Criterion obtained from Hyperparameter tuning		
(1) Decision Tree Classifier		
Hyperparameter Tuning method	Criterion	Accuracy Score
Random Search Algorithm	min_samples_leaf': 50, 'max_depth': 17, 'criterion': 'entropy'	0.893482
Grid Search Algorithm	criterion': 'entropy', 'max_depth': 7, 'min_samples_leaf': 30	0.893482
(2) Random Forest Classifier		
Random Search Algorithm	n_estimators': 40, 'min_samples_leaf': 20, 'max_features': 30, 'max_depth': 7	0.937993
Grid Search Algorithm	max_depth': 9, 'max_features': 30, 'min_samples_leaf': 20, 'n_estimators': 30	0.933349

Portuguese Bank Target Marketing

Methods used to build the model

1. Decision Tree Classifiers
2. Random Forest Classifier

Methods used for Hyperparameter Tuning

1. Random Search
2. Grid Search
3. Cross Validation

Best Classifier Criterion obtained from Hyperparameter tuning		
(1) Decision Tree Classifier		
Hyperparameter Tuning method	Criterion	Accuracy
Random Search Algorithm	min_samples_leaf': 70, 'max_depth': 15, 'criterion': 'gini	0.892858
Grid Search Algorithm	'gini', 'max_depth': 5, 'min_samples_leaf': 10	0.889872
(2) Random Forest Classifier		
Random Search Algorithm	n_estimators': 20, 'min_samples_leaf': 90, 'max_features': 10, 'max_depth': 1	0.883015
Grid Search Algorithm	'max_depth': 1, 'max_features': 10, 'min_samples_leaf': 10, 'n_estimators': 20	0.883015

Thank You!