Report

Group ID : B2

Lab Group no : 04

ID : 14.02.04.081
ID : 14.02.04.088

February 10, 2018

1 Problem:

There are many kind of smartphones in the market. If a buyer choose a phone that he wants to buy but doesn't know the price, then he should be predicting the price of the phone based on the specification of the phone. He can do the prediction by looking up the prices of the other phones.

2 About Dataset:

In the dataset we have 4 features column and a price column based on the features. For the features we have used CPU benchmark result, User Experience result, 3D rendering benchmark and the manufacturer brand of the phone. For each phone we have got different prices. We are gonna be train our machine with these data and try predict the price of a phone based on these features.

3 Description of The Models:

There are in total 4 attributes in each case of the dataset. They are:

1. **Linear Regression**: linear regression is a linear approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.[1] (This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.)

2. **Random Forest Regression**: Random forests or random decision forests[1][2] are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

3. **Gradient Boosting Regression**: Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

4. **Decision Tree Regression**: Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree

structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.

5. **Extra Trees Regression**: The Extra-Tree method (standing for extremely randomized trees) was proposed in [GEW06], with the main objective of further randomizing tree building in the context of numerical input features, where the choice of the optimal cut-point is responsible for a large proportion of the variance of the induced tree.

4 Comparison of the Performance Scores:

Model	**Performance score**
Linear Regression	0.620507248003
Random Forest Regression	0.816205176239
Gradient Boosting Regression	0.650105890973
Decision Tree Regression	0.7254657019
Extra Trees Regression	0.850135450504