Lesson 15

Random Forest and AdaBoost Classifiers

Random Forest (RF) **Classifier**

- Uses all four categories of predictor variables.
- Ensemble learning method for classification, so has high computational overheads
- Applies to classification as well as regression

RF Classifier

- Constructs a multitude of decision trees
- Outputs the modes of the classes by individual trees
- A tree learning algorithm modified that selects, such that each candidate splits in the learning process and uses a random subset of features for learning (to take decisions)

Random Forest (RF) **Classifier**

- Effectively programs the conditional relationships and non-linear, kernel or other complex functions
- Applications are when the data points are less than 10 M
- Parallel execution with shared nothing architecture

RF Classifier

- Let I (T) = Training input features or values
- O [I (T)] = Output exemplary features or values for **I**
- RF selects a bootstrap sample, randomly from input and output examples (T).

Training

- Decision or regression tree D trains on the sample s = 1 to S of ds (Is, Os).
- Training algorithm uses bootstrap aggregating method
- Puts tree learners in bags (containers)

Training

- Training dataset input vectors and corresponding output variables bootstrap the sample repeatedly, and fits with the decision tree
- The prediction variable can either be estimated by vote (frequency of correct decisions or regression) or by using the averaging formula, in following equation

Averaging Formula Predictive Function

- Predictive Variable $P(c_1, c_2, \ldots, c_m) = \sum P_i (c_1, c_2, .., c_m)$ for best decider D values for the c1, c2, .., m from i = 1 to S randomly selected observations.
- D(S) = Maximum vote frequency (s)/Total sample features or
- $D\,(\mathbf{S}) = (S)^{-1} \sum_{s=1}^{S} d_s\,(\mathbf{y})$

AdaBoost Classifier

- Initially assumes uniform weight of training examples, final classifier is linear combination of individual single feature classifier
- Constructs a “strong” learner as a linear combination of weak learners (Boosting).

Initial Weights of Training Example in AdaBoost Classifier

- Weight means importance of an example with respect to other examples
- Assume a trainer algorithm using a sample (example) in trained vector **T.** The Ts: ε {− 1, 1}
- Means training example, initial Ts: ε weight is − 1 or + 1 to start with

AdaBoost Classifier

- Ts is a week classifier, as it cannot generate a predictor variable and classify
- Each individual classifier is considered sequentially
- When its algorithm increases its weight, then other weights correspondingly reduce.

Consider Training Data: (Car-Model, AI, Age) for Buyers

- Refer Lesson 14
- (JXJ, 142, 39), (JXF, 80, 49), (H, 48, 43), (Z, 28, 46), (JXJ, 138, 44), (JXF, 82, 45), (H, 52, 40), (Z, 24, 34),
- (JXJ, 140, 44), (JXF, 70, 36), (H, 58, 38), (Z, 28, 46), (JXJ, 162, 36), (JXF, 86, 52), (H, 43, 33), (Z, 23, 36), (JXJ, 132, 44), (JXF, 90, 46), (H, 48, 42), (Z, 18, 26),...

Predictor Variables

- JXJ, JXF, H and Z are predictor variables
- Assume that Age = 42, the prediction by individual classifier using age can be for buying any car model from the training data set, :ε (s) weight will decrease for this sample.

Predictor Variables Weights increase

- But for sample, Annual income = 15.0 Million Rupees, decidedly training example predicts buying the JXJ. $T_s : \varepsilon(s)$ will increase towards 1 for this sample.

Increase the weights of variables

- Now those Te, which are misclassified due to need greater importance
- Assume weight of Ts is ε (s). Finally, the classifier adapts and forms linear combinations of those Te whose weights were increased.

Decider D(S)

- The equation for Decider D(S) using the linear combination of features is given by:
- $D\,(\mathbf{S}) = K \sum_{s=1}^{S} \varepsilon\,(s).d_s\,(\mathbf{y})$
- $K = \sum_{s=1}^{S} \varepsilon\,(s)$ = Sum of the weights of
- individual samples

Summary

We learnt:

- Random Forest Classifier
- Ensemble learning method using multitude of decision trees fitted using randomly selected training datasets
-

Summary

We learnt:

- AdaBoost Classifier
- Constructing (adapting) a “strong” classifier as a linear combination of weak single feature classifiers (boosting)
-

End of Lesson 15 on

Random Forest and AdaBoost Classifiers