29.11.2021 MACHINE LEARNING SACHIN RAGHUL T

OBSERVATION - 13 2019103573

Aim :-

To implement the ensemble classifier.

Adaboost algorithm :-

* Initialise all weights to $1/N$ where N is the number of datapoints.

* while $0 < \epsilon_t < \frac{1}{2}$ (and $t < T$, some number of iterations):

    - Train classifier on $\{S, w^{(t)}\}$ getting hypotheses $h_t(x_n)$ for datapoints x_n

    - compute training error $\epsilon_t = \sum_{n=1}^{N} w_n(t) I(y_n \neq h_t(x_n))$

    - set $\alpha_t = \log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$

    - Update weights using:

      $$w_n^{(t+1)} = w_n^{(t)} \exp(\alpha_t I(y_n \neq h_t(x_n))/Z_t,$$

      where Z_t is normalisation constant.

* output $f(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$

Bagging classification :-

classifier generation :

* Let N be the size of the training set for each of iteration.

* Sample N instances with replacement from the original training set apply the learningalgorithm to the sample.

* Store the resulting classifier.

classification :-

* For each of the t classifier:

  predict class of instance using classifier.

* return class that was predicted most often.

Basic Random Forest Training Algorithm :-

* For each of N trees.

  - Creates a new bootstrap sample of training set
  - Use this bootstrap sample to train a decision tree.
  - At each node of the decision tree, randomly select m features and compute the information gain (or Gini impurity) only on that set of features, selecting the optional one.

  - repeat until the tree is complete.