CS522 Advanced Database Systems
Classification: Ensemble Methods

Chengyu Sun
California State University, Los Angeles

Ensemble Methods

- Use a number of *base* classifiers, and make a predication based on the predications of the base classifiers

Ensemble Classifier Example

- Binary classification
- 3 classifiers, each with 30% error rate
- Classify by majority vote
- Error rate of the ensemble classifier??

Build An Ensemble Classifier

- Approach 1: use several different classification methods, and train each with a different training set
- Approach 2: use the one classification method and one training set

Get K Classifiers Out Of One …

- By manipulating the training set
  - Use a different subset of the training set to train each classifier
  - E.g. Bagging and Boosting
- By manipulating the input features
  - Use a different subset of the attributes to train each classifier
  - E.g. Random Forest

…Get K Classifiers Out Of One

- By manipulating the class labels
  - E.g. ECOC.
- By manipulating the learning algorithm
  - E.g. use of different kernel functions, introducing randomness in attribute selection in decision tree induction

Manipulate the Training Set

- How can we use one training set to train k classifiers?
  - Use the same training set for each classifier??
  - Evenly divide the training set into k subsets??

Bagging

- Use a bootstrap sample for each classifier
- A bootstrap sample D_i
  - Obtained by uniformly samples the training set D with replacement $|D|$ times
  - Contains roughly 63.2% of the original records
    - $1-(1-1/N)^N \rightarrow 1-1/e=0.632$

Bagging Example – Dataset

X	Y
0.1	1
0.2	1
0.3	1
0.4	-1
0.5	-1
0.6	-1
0.7	-1
0.8	1
0.9	1
1.0	1

Bagging Example – Classifier

- Base classifier: decision tree with one level $x \leq k$
  - Best training accuracy possible??
- Ensemble classifier: 10 classifiers, majority vote

Bagging Example – Bagging

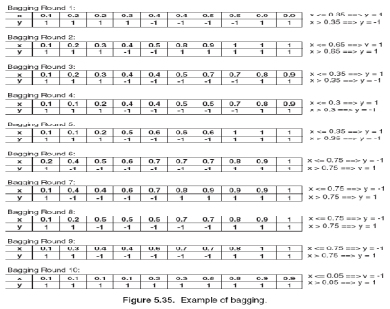

Figure 5.35. Example of bagging.

Bagging Example – Classification

Round	x=0.1	x=0.2	x=0.3	x=0.4	x=0.5	x=0.6	x=0.7	x=0.8	x=0.9	x=1.0
1	1	1	1	-1	-1	-1	-1	-1	-1	-1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	-1	-1	-1	-1	-1	-1	-1
4	1	1	1	-1	-1	-1	-1	-1	-1	-1
5	1	1	1	-1	-1	-1	-1	-1	-1	-1
6	-1	-1	-1	-1	-1	-1	-1	1	1	1
7	-1	-1	-1	-1	-1	-1	-1	1	1	1
8	-1	-1	-1	-1	-1	-1	-1	1	1	1
9	-1	-1	-1	-1	-1	-1	-1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
Sum	2	2	2	-6	-6	-6	-6	2	2	2
Sign	1	1	1	-1	-1	-1	-1	1	1	1
True Class	1	1	1	-1	-1	-1	-1	1	1	1

Figure 5.36. Example of combining classifiers constructed using the bagging approach.

About Bagging

- Reduces the errors associated with random fluctuations in the training data for *unstable classifiers*, e.g. decision trees, rule-based classifiers, ANN
- May degrade the performance of *stable classifiers*, e.g. Bayesian network, SVM, k-NN

Intuition for Boosting

- Sample with weights
  - hard-to-classify records should be chosen more often
- Combine the prediction of the base classifiers with weights
  - Classifiers with lower error rates get more voting power

Boosting – Training

- Initialize the weight of each record
- For `i=1` to `k`
  - Sample with replacement according to the weights
  - Train a classifier M_i
  - Calculate `error(`M_i`)`, assign a weight to M_i based on `error(`M_i`)`
  - Update the weights of the records
    - Increase the weights of the misclassified records
    - Decrease the weights of the correctly classified records

Boosting – Classification

- For each class, sum up the weights of the classifiers that vote for that class
- The class that gets the highest sum is the predicted class

Boosting Implementation

- How the record weights are decided
- How the classifier weights are decided

Adaboost

Error rate of classifier M_i: $error(M_i) = \sum w_j \times err(\mathbf{X_j})$

Update the weights of the correctly classified records: $w_j \times \frac{error(M_i)}{1-error(M_i)}$

Weight of classifier M_i: $\log \frac{1-error(M_i)}{error(M_i)}$

- Initial $w_j = 1/|D|$
- Classifiers with `error(Mi)>0.5` are dropped
- Normalize the weights of all records after updating the weights of the correctly classified records

Adaboost Example

- 5 records
- M_1 classification results:

Record	1	2	3	4	5
Correctly classified	Y	Y	N	N	Y

Weight of M_1?? Updated record weights??

Random Forest

- A number of decision tree classifiers created from the same training set

Create a Random Forest

- *Forest-RI*: randomly select F attributes out of d input attributes usually $F=\log_2 d+1$
- *Forest-RC*: at each node, create F new attributes, each is a random linear combination of L randomly selected attributes
- At each node, randomly select one split out of the top F best splits

Some Empirical Comparison of Ensemble Methods

- See Table 5.5 in Introduction to Data Mining by Tan, Steinbach, and Kumar

Readings

- Textbook 6.14