3.2 Classification Algorithms

Linear Classifier

Uses a linear function to divide each set into two subsets

- For 3+ classes: fit N-1 lines (N classes)

Nearest Neighbor Classifier

If the nearest instance to the previously unseen instance is A, then the class is also A, otherwise it is B.

- the decision surface of a NN classifier (also called Theissen regions) is implicit and divide the space into regions "belonging" to each instance and corresponding class.

KNN Algorithm

Generalization of the nearest neighbor algorithm

- Find the nearest K (typically chosen to be an odd number) instances
- Each one represents a vote

Sensitivity to Irrelevant Attributes

- Using more attributes can sometimes lead to getting the wrong classification for an instance, when using less gave the right one;
- Nevertheless, using some subsets can also lead to wrong classifications.

Advantages

- Simple to implement
- Handles correlated features
- Defined for any distance measure
- Handles streaming data trivially

Disadvantages

- Very sensitive to irrelevant features
- Slow classification time for large datasets
- Work best for real valued datasets

Decision Tree Classifier

Avoid Overfitting

- **Prepruning**: halt tree construction early
  - do not split a node if this would result in the goodness measure falling below a threshold
- **Postpruning**: get a sequence of progressively pruned tree; use a set of data different from the training data to decide which is the "best pruned tree"

Advantages

- Easy to understand
- Easy to generate rules

Disadvantages

- Overfitting
- Does not handle correlated features very well (classified by rectangular partitioning)
- Can be quite large (pruning is necessary)

Bayes Classifier

Use Bayes theorem, which says:

- The probability of an instance being in a certain class is the probability of that instance being generated by that class times the probability of occurence of that class, divided by the probability of that instance existing.
- Assuming independent distributions, the probability of a certain class generating the instance is the multiplication of the probability of that class generating the observed value for each of the features of that instance.

Advantages

- Fast to train (single scan) and to classify
- Not sensitive to irrelevant features
- Handles real and discrete data
- Handles streaming data well

Disadvantages

- Assumes independence of features

Neural Network Learning

- Set of neurons (input/output units) connected with weights
- Supervised learning adjusts weights to ensure outputs to given inputs are the expected ones
- Predicting: feed input values; collect outputs

Advantages

- **Universal**: fit any continuous function
- **Versatile**: output may be one or more discrete and real values
- **Online**: application and learning are intertwined
- **Robust** to errors and noise data
- **Fast** application to new examples
- **Parallel**

Disadvantages

- **Slow** training
- **Low usability**: empirical parameter tuning; network topology and learning rate
- **Low Interpretability**: understand the weights
- **Low Adaptability**: not easy to incorporate domain knowledge

Support Vector Machines (SVM)

- Linear learning machines with maximisation of margin (better separation between classes);
  - Hyperplane farthest from both classes has less risk of overfitting.
- Higher robustness to the curse of dimensionality;
- For non-linear functions, map attributes to space where linear discrimination is possible.

SVM for Regression:

- Minimize the tube "around" the data;
- Instead of maximizing the distance to the closest examples from each class.

[< Go back]