Classification

Elena Baralis
Politecnico di Torino

Associative classification

- The classification model is defined by means of Association rules
  - The rule head is the class label
  - Rules are selected based on
    - Support, confidence and correlation thresholds
    - Database coverage: procedure to cover the training set by using the extracted rules, sorted according to proper criteria
- Strength
  - Easy to interpret model
  - It achieves higher classification accuracy than Decision Trees
    - It considers the correlation among more attributes at the same time
  - Efficient in classification
  - Unaffected by missing data
  - Good scalability in the training set size
- Weakness
  - Rule generation may be slow
  - Low scalability in the number of attributes

Bayesian classification

- Based on computing probabilities
- Strength
  - The model can be incrementally updated
  - Efficient in classification
  - Quite easy to interpret model
- Weakness
  - The generation of the reference model is computationally unfeasible
    - Simplified hypothesis: naïve hypothesis
  - Naïve hypothesis significantly reduces the accuracy

Elena Baralis
Politecnico di Torino

Bayesian classification

- Use of the Bayes theorem
  - P(C|X) = probability that X=<x_1,…,x_k> belongs to class C
  - X is labeled with the class that maximizes P(C|X)
  - Bayes theorem
    P(C|X) = P(X|C)·P(C) / P(X)
    - P(X) constant for all class labels C
    - P(C) a-priori probability of C
    - P(X|C) cannot be completely computed
- Naïve hypothesis
  - It assumes statistical independence among attributes x_1,…,x_k
  - It is not always verified
    - It may affect the quality of the model
- Bayesian networks
  - They allow specifying a subset of dependencies among attributes

Elena Baralis
Politecnico di Torino

Bayesian classification

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

From Han, Kamber,"Data mining; Concepts and Techniques", Morgan Kaufmann 2002

Bayesian classification

outlook	
P(sunny|p) = 2/9	**P(sunny|n) = 3/5**
P(overcast|p) = 4/9	**P(overcast|n) = 0**
P(rain|p) = 3/9	**P(rain|n) = 2/5**
temperature	
P(hot|p) = 2/9	**P(hot|n) = 2/5**
P(mild|p) = 4/9	**P(mild|n) = 2/5**
P(cool|p) = 3/9	**P(cool|n) = 1/5**
humidity	
P(high|p) = 3/9	**P(high|n) = 4/5**
P(normal|p) = 6/9	**P(normal|n) = 2/5**
windy	
P(true|p) = 3/9	**P(true|n) = 3/5**
P(false|p) = 6/9	**P(false|n) = 2/5**

P(p) = 9/14
P(n) = 5/14

Data object to classify
X = <rain, hot, high, false>

P(X|p)·P(p) =
P(rain|p)·P(hot|p)·P(high|p)·P(false|p)·
P(p) = 3/9·2/9·3/9·6/9·9/14 = 0.010582

P(X|n)·P(n) =
P(rain|n)·P(hot|n)·P(high|n)·P(false|n)·
P(n) = 2/5·2/5·4/5·2/5·5/14 = 0.018286

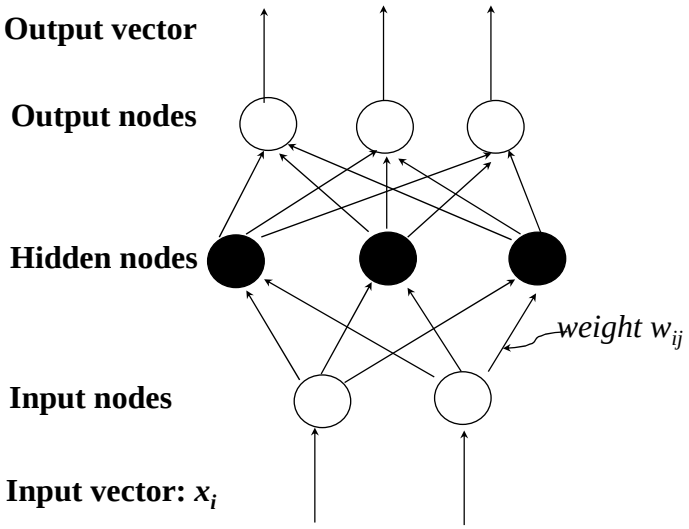
From Han, Kamber,"Data mining; Concepts and Techniques", Morgan Kaufmann 2002

Neural Networks

- Inspired to the structure of the human brain
  - Neurons as elaboration units
  - Synapses as connection network
- Strength
  - High accuracy
  - High tolerance to noise and outliers
  - Efficient in classification
  - It supports both discrete and continuous output
- Weakness
  - Long training time
  - No interpretable model
  - Hard introducing knowledge from the application domain

Elena Baralis
Politecnico di Torino

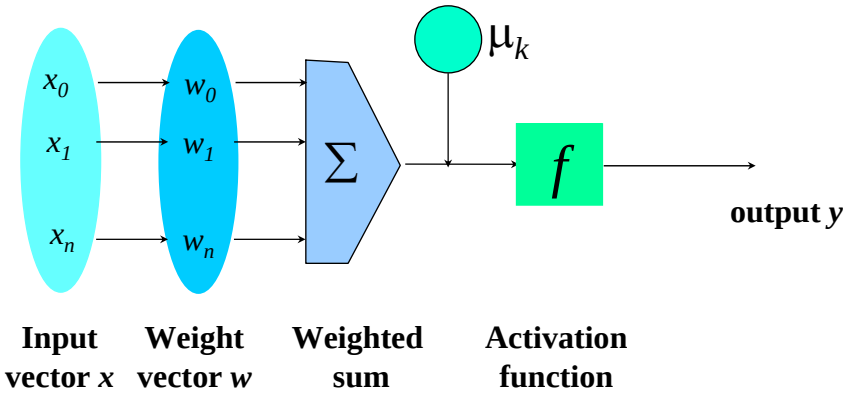
Structure of a Neural Network



From Han, Kamber,"Data mining; Concepts and Techniques", Morgan Kaufmann 2002

Elena Baralis
Politecnico di Torino

Structure of a neuron



From Han, Kamber,"Data mining; Concepts and Techniques", Morgan Kaufmann 2002

Construction of the Neural Network

- Objective
  - Definition of an optimal set of weights and offset
- Basic algorithm
  - Initially assign random values to weights and offset
  - Process instances in the training set one at a time
    - For each neuron, compute the result when applying weights, offset and activation function for the instance
    - Forward propagation until the output is computed
    - Compare the computed output with the expected output, and error evaluation
    - Backpropagation of the error, by updating weights and offset
  - The process ends when
    - % of accuracy above a given threshold
    - % of error below a given threshold
    - The maximum number of epoches is reached