Multiclass Classification

Prof.Mingkui Tan

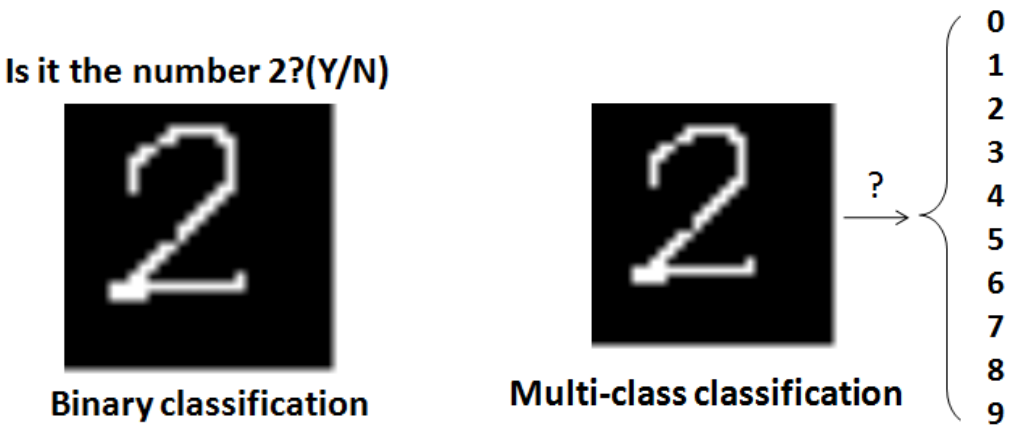
South China University of Technology
Southern Artificial Intelligence Laboratory(SAIL)

October 17, 2017

Content

Three common classification problems

- Binary classification
- Multi-class classification
- Multi-label classification



Multi-class classification

Multi-class classification is the common classification problem, which classifies instances into one of the more than two classes.

Dataset

- MNIST
- Cifar-10 and Cifar-100
- ImageNet
- ...

Transformation to binary classification

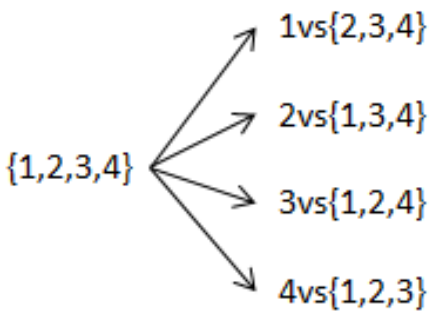
The strategies reduces the problem of multi-class classification to multiple binary classification problems.

- One-vs.-rest
- One-vs.-one
- Decision Directed Acyclic Graph, DDAG

One-vs.-rest method

Train:

- For each class
  - Train a binary classifier with the samples of that class as positive samples and others as negatives.



One-vs.-rest

General strategies

One-vs.-rest

Predict:

- For each binary classifier
  - Produce a real-valued confidence score
- Predict the label with the highest confidence score

$$\hat{y} = \underset{k \epsilon \{1 \ldots k\}}{argmax} \; f_k(x)$$

One-vs.-rest method

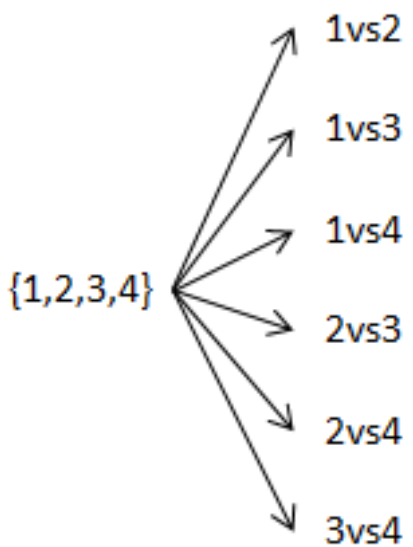
Advantage vs Disadvantage

- Advantage
  - Trains K binary classifiers for a K-way multi-class problem.
- Disadvantage
  - The distributions of the binary classifications are unbalanced. (The set of negatives is much larger than the set of positives.)
  - The scale of the confidence values may differ between the binary classifiers.

One-vs.-one method

Train:

- For each pair of classes
  - Train a binary classifier to discriminate between them.



One-vs.-one

One-vs.-one method

Predict:

- For each binary classifier
  - Contrast the two categories and do a voting.

```
A=B=C=D=0
A vs B-classifier: if A win, A=A+1; otherwise, B=B+1;
A vs C-classifier: if A win, A=A+1; otherwise, C=C+1;
...
C vs D-classifier: if C win, C=C+1; otherwise, D=D+1;
```

- Predict the label with the maximum number of votes wins.

$$\hat{y}=\texttt{Max(A,B,C,D)}$$

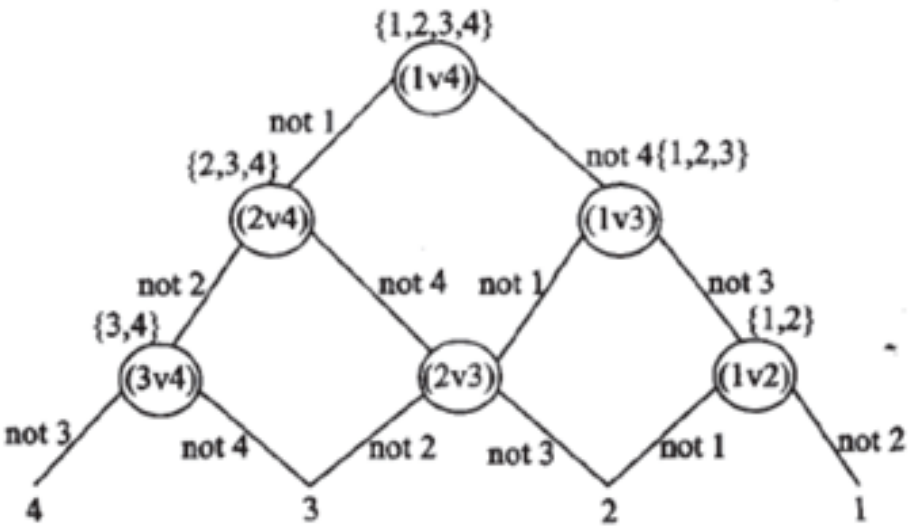
One-vs.-one method

Advantage vs Disadvantage

- Advantage
  - The distributions of the binary classifications are balanced.
- Disadvantage
  - Train $K(K-1)/2$ binary classifiers for a K-way multi-class problem, which has high computed complexity.
  - Suffer from ambiguities when receive the same number of votes.

Decision Directed Acyclic Graph

Compared to One-vs.-one method, it uses a rooted binary directed acyclic graph which has internal nodes and leaves.



Decision Directed Acyclic Graph

Decision Directed Acyclic Graph

Predict:

- Start at the root node.
- Before reaching a leaf node:
  - Evaluate the binary decision function.
  - Move to either left or right depending on the output value.

Compared to One-vs.-one method

The correlation between each of the binary classifications brings the cost of the prediction down.

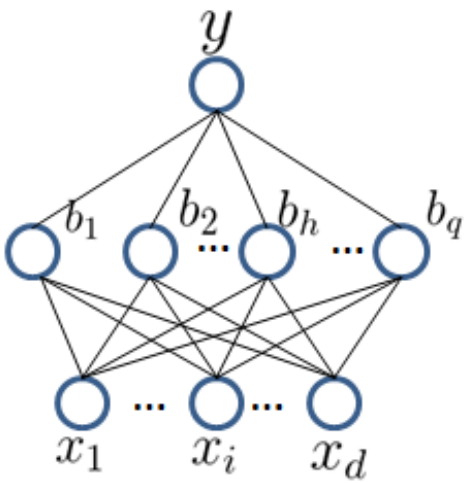
Extension from binary classification

The strategies extends the existing binary classifiers to solve multi-class classification problems.

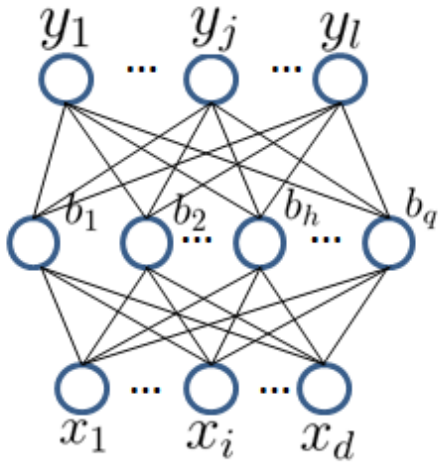
- Neural networks
- Decision trees
- K-nearest neighbours
- Softmax function

Neural networks

Instead of just having one neuron in the output layer, the network could have N binary neurons leading to multi-class classification.



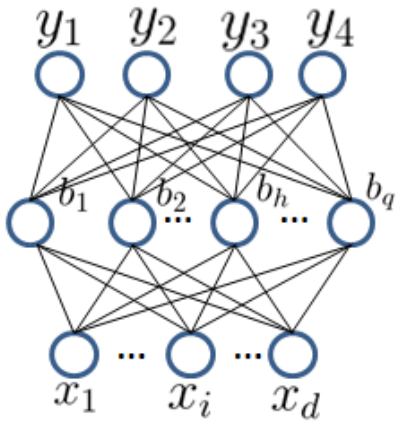
binary classification



multi-class classification

Neural networks

Each output neuron is designated to identify a given class. N=K



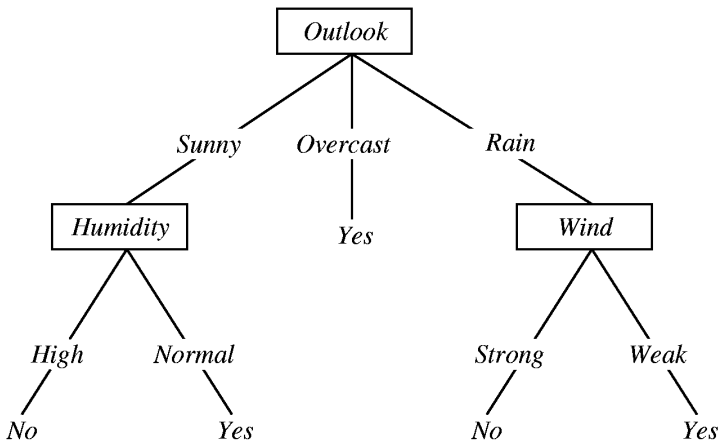
K \ N	y_1	y_2	y_3	y_4
Class 1	1	0	0	0
Class 2	0	1	0	0
Class 3	0	0	1	0
Class 4	0	0	0	1

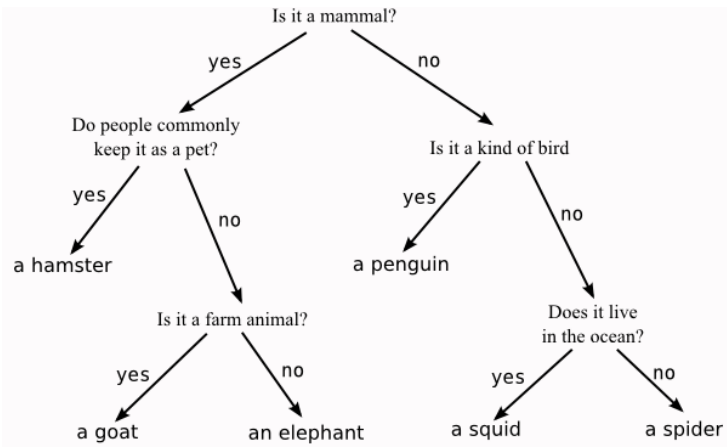
One-per-class coding

Train: The loss function $\mathbf{E} = \sum_{j=1}^{l} \mathbf{E}_j$
Predict: The neuron with the maximum output is considered as the class of the example.

Decision trees

The decision tree tries to split the training data based on the values of the features to produce a good generalization.

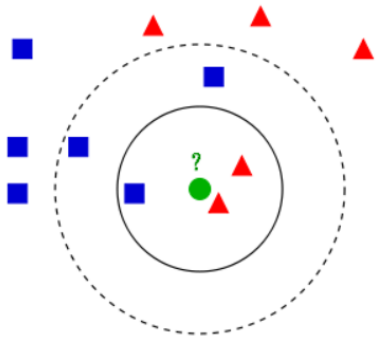


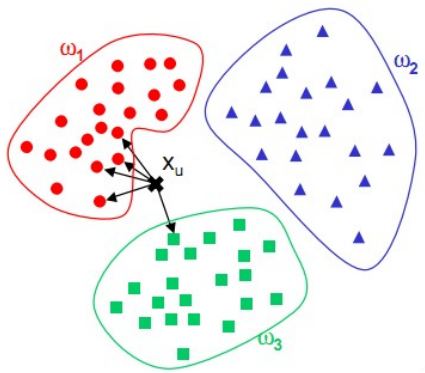


- The algorithm can naturally handle binary and multi-class classification.

K-nearest neighbours

- Calculate the distances
  Calculate the distances between the test object and each object in the training set.
- Find the neighbors
  Get the K nearest training objects as neighbors.
- Vote on labels
  Classify the test object based on the most frequent class of the neighbors.





K-nearest neighbours

Advantage vs Disadvantage

- Advantage
  - The method is a non-parametric classification algorithms.
  - The algorithm can naturally handle binary and multi-class classification.
- Disadvantage
  - The computational and memory requirements are high.
  - Finding good representations and distance measures between objects is hard.

Hierarchical classification

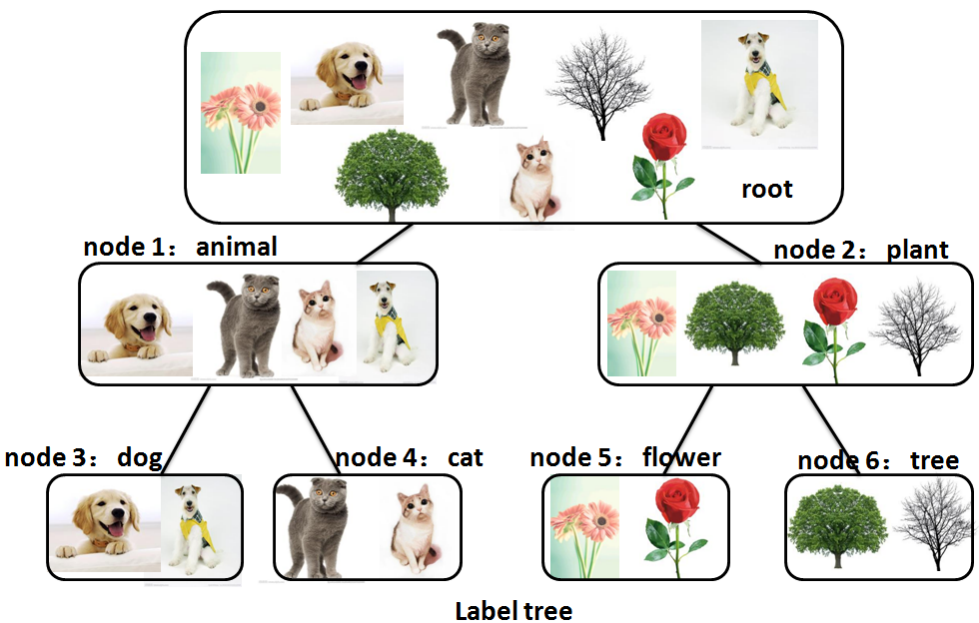
The strategies tackles the multi-class classification problem by dividing the output space i.e. into a tree.

- Label tree

Label tree

Train:

- Before the leaf nodes contain only a single class
  - Each parent node are divided into a number of clusters, one for each child node.
- At each node, a simple classifier is trained to discriminate between the different child class clusters.



Label tree

Label tree

Predict:

- Start from the root node
  - Travel to a leaf node which is associated with a label

Advantage vs Disadvantage

- Advantage
  - The tree method brings the cost of the prediction down.
- Disadvantage
  - Finding good clustering method is important.

THANK YOU!