L1—wtr

TianruiWang

April 2019

1 Introduction to Machine Learning

Brief Introduction to Machine Learning

Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is *infeasible to develop an algorithm of specific instructions for performing the task.*
classification for discrete data
regression for consinuous data

Basic Terminology

Attribute = Attribute value such as color shape weight
Samples data set such as melon1 melon2 and so on
Feature vector such as a melon with green color , big shape , high weight

Hold-out

We divide the whole sample set into two parts, training set and testing set.
Method 1 Cross validation
1 Divide the dataset into k subsets(mutually exclusive)
2 Each time, the combination of (k - 1) subsets is used as the training set, the last one as the testing set
3 K-fold cross validation
4 LOO (leave one out) (in which k == m¡number of samples¿)
5 K = 10 (usually)
Method 2 Bootstrapping
1 Dataset D (m samples) -¿ generate D'
2 For i in range (m)
3 Randomly select a sample s, and place its copy into D'
4 End
5 About 36.8 percent of samples don't appear in D'

Bias-variance

Expectation of the learning algorithm
f*$(x) = E_D[f(x;D)]$
$Variance$
$var(x) = E_D[(f(x;D) - f^*(x))^2]$

$Noise$
$e^2 = E_D[(y_d - y)^2]$
$Bias(difference\ between\ expected\ output\ and\ actual\ label)$
$bias^2 = (f^*(x) - y)^2$
$Conclusion$
$E(f;D) = bias^2 + var(x) + e^2$

$DecisionTree$

A decision tree = root node + internal node + leaf node
Node set   leaf node -¿ Each node corresponds to an attribute test
Leaf Node set -¿ Each node corresponds to an output

Input: Training data set D =¡x1,y1¿,¡x2,y2¿,......,¡xn,yn¿
Attribute set A =a1,a2,a3,......

```
Function : Generate$_T ree(D, A)$
$generatenode$
if$(allsamplesinDareinthesameclassC)$then
$labelnodeasCnode$return
end if
if$A\ is\ empty$  OR $all\ samples\ in\ D\ have\ same\ attributes\ in\ A$ then
$label\ node\ as\ a\ leaf\ node\ ,classify\ it\ as\ a\ class\ with\ the\ most\ samples\ in\ D$
end if
$select\ the\ best\ attribute\ in\ A\ (how\ to\ select\ it\ will\ be\ introduced\ later\ )$
for$every$a$^*$ do create a branch for node ; make Dv represent for sample set where $a^* = a_v^*$
if $D_v$ is empty then
label it as leaf node ; label the class which receives most samples return
else Generate$_T ree($D$_v$  , A- $a_n$ )
end if
end for
```

Information Entropy

We use this to evaluate the gain of information , in order to decide which attribute fits best.
Gain(D,a)= Ent(D) - $\sum_{v=1}^{V} Ent(D^v)\ |D_v|/|D|$
$Among\ those\ attributes\ ,we\ chose\ the\ attribute\ with\ the\ largest\ Gain(D,a)$.

Pruning

Pruning is mainly divided into two kinds, prepruning and postpruning.
The purpose is to improve the generalizing ability of decision trees.
We need to find a balance between low learning rate and high learning rate.
If the learning rate is too low, the model is not effective enough.
If the learning rate is too high, the model will mistaken some special attributes as general attributes,