

Inductive learning

CS 6375 ASSIGNMENT-1



September 9, 2018

fall -2018

THE UNIVERSITY OF TEXAS AT DALLAS

SAIPRAVEEN VABBILISETTY – SXV165130

LIKITHA KOMMINENI -

1. Need to fill formulae I will fill the code
2. False Positive is 10 % and False Negative is 20 %
3. a. Most specific hypothesis (S) based on a training data:

Pros:

* This type of hypothesis covers the observed positive training examples.
* We get more defined hypothetical space

Cons:

* if reduced any further, there are chances to miss out positive training examples, and hence leading to inconsistency
* if novel data that is never seen before is observed by a learner, it assumes it to be negative

b. Most general hypothesis (G) based on a training data are:

Pros:

* covers the observed positive training examples and covers as much of the remaining feature space without including any negative training examples
* Includes all the positive examples

Cons:

* if enlarged any further, there are chances to include negative training examples, and hence leading to inconsistency
* If a negative data already observed, learner assumes it to be positive

1. **Consistent Hypothesis:** A hypothesis h is consistent with a set of training examples D if and only if h(x) = c(x) for each example (x, c(x)) in D.

**Version Space:** The version space, denoted , with respect to hypothesis space H and training examples D, is the subset of hypotheses from H consistent with the training examples in D.

1. The most generic hypothesis has “**?**” value for each attribute.
2. A. The total number of instances possible are 16. Since 4 attributes, the total number of Boolean combinations possible are 2^4.

B. For conjunctive hypothesis, there are 4 possible choices for each attribute ?, T, F,ᴓ. So possible hypothesis is 2^16.

C. There can be 81 such combinational hypothesis. Four attributes and 3 choices for each attribute.

D. The number of ways for selecting two attributes is 4C2 which is 6. Selecting the root and leaf node is 2 ways. So total combinations are 12 decision trees.

E. The total number of ways for labelling is 2^4 that is 16.

1. s0 = (ɸ, ɸ, ɸ, ɸ, ɸ)

Since first training data is positive, S is

s1 = (1, 1, 0, 1, 1)

Since second training data is negative we ignore that and make no changes in S boundary.

s2 = (1, 1, 0, 1, 1)

Third training data is positive, and we update the S boundary.

s3 = (1, 1, ?, 1, ?)

Since fourth training data is negative we ignore that and make no changes in S boundary.

s4 = (1, 1, ?, 1, ?)

Fifth training data is positive, and we update the S boundary.

S5 = (1, 1, ?, 1, ?)

1. DNF f = ( All Data ʌ D2) ∨ (~All Data ʌ D1)
2. Likitha Complete cheyyava

s0 = ((ɸ, ɸ, ɸ, ɸ), ( ɸ, ɸ, ɸ, ɸ))

G0 = ((?,?,?,?),(?,?,?,?))

S1 = ((ug, se, l, hs); (gr, cs,h,hs))

G1 = ((?,?,?,?),(?,?,?,?))

S2 = ((ug, se, ?, ?), (gr, cs,h,hs))

G2 = ((?,?,?,?),(?,?,?,?))

S3 = ((ug, se, ?, ?), (gr, cs,h,hs))

G3 = (((ug,?,?,?),(?,?,?,?)), ((?,?,?,?),(?,?,?,hs)))

S4 = ((ug, se, ?, ?), (gr, ?, h, ?))

G4 = ((ug,?,?,?),(?,?,?,?))

b. Six hypothesis total consistent hypotheses are returned after running the Candidate Elimination algorithm.

Two hypotheses are consistent with the given data point