

19Z601- Machine Learning

Presented by
Ms.Anisha.C.D
Assistant Professor
CSE

Algorithm Vs Model

ALGORITHM	MODEL
Algorithm is set of rules or procedure to follow, Mathematical technique or equation which is not definite (no actual values.)	Model are build using algorithm, which are definite or concrete with equation having actual values based on the formulation from data.
Equation : Logistic function : $Y = (1 / (1 + e^{-x}))$ Suppose : X1 and X2 variables, then $Y = (1 / (1 + e^{-(b_0 + b_1 * X_1 + b_2 * X_2)}))$	$Y = 1 / (1 + e^{-(0.1 + 0.2 * X_1 + 0.3 * X_2)})$

Machine Learning System Design

1. Choosing the Experience

Data Acquisition



Data Preprocessing



Splitting of Training and Testing Set

Train Data



2. Choosing the Target Function, 3. Choosing the Representation of Target Function, 4. Choosing the Function Approximation Algorithm

Choosing Machine Learning Model



Model Training (Fitting the function)



Model Optimization (Tuning)

Test Data



Model Testing and Evaluation

Concept Learning

- Concept Learning is a learning strategy of acquiring abstract knowledge or inferring a general concept.
- It is a process of abstraction and generalization from the data.

Find S Algorithm

- **Input : Positive Instances in the training dataset**

- **Output : Hypothesis 'h'**

1. Initialize 'h' to the most specific hypothesis

$$h = \langle \varphi, \varphi, \varphi, \varphi, \varphi, \varphi \rangle$$

2. Generalize the initial hypothesis for the first positive hypothesis [Since 'h' is more specific]

3. For each subsequent instances :

- 3.1 If it is a positive instance,

- 3.1.1 Check for each attribute value in the instance with the hypothesis 'h'.

- If the attribute value is the same as the hypothesis value, then do nothing

- Else if the attribute value is different than the hypothesis value, change it to '?' in 'h'

- 3.2 Else if it is a negative instance,

- Ignore it.

Find S Algorithm – Example Training Dataset

Consider the training dataset of 4 instances. It contains the details of the Performance of the film and their likelihood of the film to be blockbuster or not after release. Apply the Find S-Algorithm.

Story	Screenplay	Retro / Nostalgic Music	Cast and Crew	Positive Response from Critics	Film Blockbuster
Good	Good	Present	Famous	8	Blockbuster
Good	Good	Present	Debut	8	Blockbuster
Not Bad	Good	Not Present	Famous	7	Not Blockbuster
Good	Not Good	Present	Debut	8	Blockbuster

Find S Algorithm – Solution

- Step 1 : Initially 'h' to the most specific hypothesis. There are six attributes , so for each attribute, initially fill ' φ ' in the initial hypothesis 'h'

$$h = \langle \varphi \ \varphi \ \varphi \ \varphi \ \varphi \ \varphi \rangle$$

Find S Algorithm – Solution

- Step 2 : Generalize the initial hypothesis for the first positive instance. I1 is a positive instance, so generalize the most specific hypothesis 'h' to include this positive instance. Hence

I1 = < Good Good Present Famous 8 >

h = < Good Good Present Famous 8 >

Find S Algorithm – Solution

- Step 3 : Scan the next instance I2, since I2 is positive instance. Generalize 'h' to include positive instance I2. For each of the non matching attribute value in 'h' put a '?' to include this positive instance.

I2 = < Good Good Present **Debut** 8 >

h = < Good Good Present ? 8 >

Find S Algorithm – Solution

- Step 4 : Scan the next instance I3, since I3 is negative instance. Ignore it. Hence the hypothesis remains the same without any change after scanning.

I3 = < Not Good Not Present **Debut** 8 > - Negative Instance

Find S Algorithm – Solution

- Step 4: Scan the next instance I4, since I4 is positive instance. Generalize 'h' to include positive instance I4. For each of the non matching attribute value in 'h' put a '?' to include this positive instance.

I4 = < Good Not Good Present Debut 8 >

h = < Good ? Present ? 8 >

Find S Algorithm – Solution

- **Inferences:**
 - **Story : Must be “Good”**
 - **Screenplay : can be any value , generalized to “?”**
 - **Retro/Nostalgic Music : Must be “Present”**
 - **Cast and Crew : can be any value , generalized to “?”**
 - **Positive Response : 8**

Activity

- Form a group of 4, create a dataset of 4 instances that consists of features along with the likelihood of the longevity of friendship .
- Label : Longevity of Friendship : Yes/No
- Features : Minimum 4 , add four instances (provides values against each feature).

Homework

- Consider the training dataset of 4 instances. It contains the details of the performance of students and their likelihood of getting a job offer or not in their final semester. Apply the Find-S Algorithm.

CGPA	Interactiveness	Practical Knowledge	Communication Skills	Logical Thinking	Interest	Job Offer
≥ 9	Yes	Excellent	Good	Fast	Yes	Yes
≥ 9	Yes	Good	Good	Fast	Yes	Yes
≥ 8	No	Good	Good	Fast	No	No
≥ 9	Yes	Good	Good	Slow	No	Yes