ML LAB 1 | Find S Algorithm | VTU

By Prajwal B Mani

Find S Algorithm

- Find-S algorithm is a basic concept learning algorithm in machine learning
- Find-S algorithm finds the most specific hypothesis that fits all the positive examples
- We have to note here that the algorithm considers only those positive training example.
- Find-S algorithm starts with the most specific hypothesis and generalizes this hypothesis each time it fails to classify an observed positive training data.
- Hence, Find-S algorithm moves from the most specific hypothesis to the most general hypothesis.

Important Representation :

- **?** indicates that any value is acceptable for the attribute.
- **0** indicates that no value is acceptable.
- The most general hypothesis is represented by: **{?, ?, ?, ?, ?, ?}**
- The most specific hypothesis is represented by : **{0, 0, 0, 0, 0, 0}**
- Explicitly specify the attributes or the extract the attributes from the datasets.

Steps involved in Find-S :

- Start with the most specific hypothesis.

  h = {0, 0, 0, 0, 0, 0}

- Take the next example and if it is negative, then no changes occur to the hypothesis.
- If the example is positive and we find that our initial hypothesis is too specific then we update our current hypothesis to general condition.
- Keep repeating the above steps till all the training examples are complete.
- After we have completed all the training examples we will have the final hypothesis when can used to classify the new examples.

Dataset

	A	B	C	D	E	F	G
1	sunny	warm	normal	strong	warm	same	Yes
2	sunny	warm	high	strong	warm	same	Yes
3	rainy	cold	high	strong	warm	change	No
4	sunny	warm	high	strong	cool	change	Yes
5							

Steps for our dataset

Step 1: Initial hypothesis

h = {0,0,0,0,0,0}

Step 2: We see that our initial hypothesis is more specific and we have to generalize it for this example. Hence, the hypothesis becomes

h = {sunny, warm, normal, strong, warm, same}

Step 3: Consider the next sample we will compare each attribute with the initial data and if any mismatch is found we replace that particular attribute with general case (” **?** ”). After doing the process the hypothesis becomes :

h = {sunny, warm, ‘?’, strong, warm, same}

Step 4: Ignore next sample because it has negative outcome

Step 5: In the next sample we will take it has a example because it has positive value

h = {sunny, warm, ‘?’, strong, ‘?’, ‘?’}

Find S Algorithm

1. Load Data set
2. Initialize **h** to the most specific hypothesis in **H**
3. For each positive training instance **x**

    For each attribute constraint **a** in **h**

        If the constraint **a** in **h** is satisfied by **x** then do nothing

        else replace **a** in **h** by the next more general constraint that is satisfied by **x**
4. Output hypothesis **h**