Machine Learning (CS60050) – Weekly Report

Kaushal Banthia (19CS10039)

Week 1: 11th – 13th August, 2021

Topics Covered:

- Introduction to Machine Learning & Examples
- Different Types of Machine Learning
- Concept Learning
- Ordering on Hypothesis
- Find - S Algorithm

Summary (Topic Wise):

- Introduction to Machine Learning & Examples

  - Paradigms of Computing – Computing by Algorithms and Computing by Learning
  - Sometimes Computing by Algorithms may not have any solution for the problem at hand (like spam detection in emails). For such cases, we turn to Computing by Learning (Automatic Learning by Machines).
  - Underlying Assumptions for Computing by Learning
    - Existence of a process that generates the data
    - Existence of certain patterns in the data
    - We can construct an approximate model of the process from the examples that are available to us.
  - What does it mean to learn: A model is said to learn if we have some task T at hand and the model improves its performance on it by having experience E and the performance evaluating measure being P.

- Different types of Machine Learning

  - Supervised Learning: Learning with Labelled Data (Regression & Classification)
  - Unsupervised Learning: Learning without Labelled Data, i.e., learning only from the input data (Clustering).
  - Reinforcement Learning: Learn a sequence of actions to try to deliver the best possible result.

- Concept Learning: Try to derive a Boolean function from the set of available training examples. ($Find\ h, such\ that\ h = c$)

  - The hypothesis (h) can be represented as a conjunction of the constraints over the training examples, having three forms (A certain value, don't care value, and no value allowed).
  - This hypothesis is represented as a vector. For the EnjoySport example, a valid hypothesis could be $< Sunny, Warm, ?, Strong, \varnothing, Same >$.
  - The hypothesis would be a n – dimensional vector, where, n is the number of attributes in the training examples.
  - We have the target concept (c), which is the Boolean function that generates the training data.

- We have the hypothesis space (H), which contains all the valid hypotheses.
- Our performance measure would be:
  P: Hypothesis h in H, such that, $h(x) = c(x) \; \forall \; x \; \in D$ (the set of training data).

- Ordering on Hypothesis

  - h is more general than h' ($h \geq_g h'$) if, for each instance x, $h'(x) = 1$, implies that $h(x) = 1$.
  - The most general hypothesis is $<?, ?, ?, \ldots., ?, ?>$
  - The most specific hypothesis is $< \varnothing, \varnothing, \varnothing, \ldots., \varnothing, \varnothing >$
  - We can consider the learning problem as a search problem, over the entire hypothesis space (H). We can do this using different approaches:
    - Search based on ordering of hypotheses.
    - Search based on finding all possible hypotheses, using a good representation of the hypothesis space. (The choice of hypothesis space reduces the number of hypotheses).

- Find – S Algorithm

```
1. Initialize h to the most specific hypothesis in H < ∅, ∅, ∅, . . . . ,∅, ∅ >
2. for each positive training instance x
   3. for each attribute constraint ai in h:
      4. if the constraint a_i in h is satisfied by x
         5. do nothing
      6. else
         7. Replace a_i in h by the next more general constraint, that is
   satisfied by x.
      8. end if
   9. end for
10. end for
11. return the final hypothesis h
```

There are certain assumptions for this algorithm to work:

- There is a hypothesis h in the describing target function c.
- There are no errors in the Training Examples.
- Everything except the positive examples is negative (The classification is binary).

There are some problems with this algorithm too. It can't tell if the training data is inconsistent, since it throws away the negative examples and it also can't tell if the concept has been learnt or not.

Concepts Challenging to Comprehend: None yet.

Interesting and Exciting Concepts: Concept Learning and the representation of the various hypothesis in a simple format.

Concepts not understood: None yet.

A novel idea: The Find – S Algorithm could have been improved, if it could take into consideration, examples that are at the same level in the ordering of the hypotheses, like <Sunny, Warm, Strong, ?> and <Rainy, Warm, Strong, ?>.