Weekly Report

Name: Utkarsh Patel
Roll No: 18EC30048
Summary of lectures given in Week-2 (Sept 9, Sept 10 and Sept 11)

Topics Covered

- **Learning as a search problem**
  Given a set of training examples D, we try to search for hypotheses $h \in H$, such that $\forall x \in D, h(x) = c(x)$.
  Here, we assume that there is no error in D and that $\exists h \in H, h(x) = c(x), \forall x \in D$
  Now there are several algorithms to do so.

  - **Find-S Algorithm** We basically start from $h \leftarrow \langle \emptyset, \emptyset, \dots, \emptyset \rangle$ and traverse through all the positive instance x in D, and try to generalize h so that $h(x) = 1$ is true. Negative examples are neglected. Not a very effective way to approach the problem.
  - **List-then-eliminate Algorithm** This is a brute force approach. Here, concept of *Version Space* is used. $VS_{H,D} = \{h \mid h \in H \text{ and } h \text{ is consistent on } D\}$. Initially, $VS_{H,D} \leftarrow H$ and we iterate through every training instance and every hypothesis in $VS_{H,D}$ and eliminate the hypothesis if it is inconsistent.
  - **Candidate Elimination Algorithm** Improvement to above algorithm. Instead of storing all consistent hypotheses in version space, we only consider the most specific set of consistent hypotheses S and most general set G. Positive instances in D are used to generalize S, and negative examples are used to make G more specific. In the end, all hypotheses in between the two extremes are consistent. The complexity of this algorithm is highly dependent on the order in which training instances are considered, however the final state of S and G remains invariant. It should be noted that computation of S is linear in number of features and number of training instances, however for G, it is exponential in number of training instances. While doing classification on population distribution $Đ$, voting is used from hypotheses in version space. However, this may not be the correct approach.

- **PAC learning model**
  What is the formal definition of a consistent hypothesis? Consider a population distribution $Đ$ with target concept c. Let's learn this target concept over some $D \subset Đ$, and let h be the hypothesis we obtained. Then, h is consistent on $Đ$ if

  $$h(x) = c(x), \forall x \in D \text{ and}$$
  $$error_{Đ}(h) = p(h(x) \neq c(x) \mid x \in Đ) \text{ is very small}$$

  When will a model be called PAC learnable? Consider sets of instances $D_1, D_2, \dots, D_N$ with dimension n, whose population distribution is given by $Đ$, let $C = \{c_1, c_2, \dots, c_N\}$, where c_i is the target concept over D_i. Then, a learner L will be considered as PAC learnable if $\forall c \in C$ and $\forall \varepsilon\ (0 < \varepsilon < 1/2)$, $\exists \delta\ (0 < \delta < 1/2)$, such that L, with probability at least $(1 - \delta)$, outputs hypothesis h such that $error_{Đ}(h) < \varepsilon$. In addition to this, the time complexity must be a polynomial in $1/\varepsilon$, $1/\delta$, n and N.
  - **Sample Complexity of PAC learner** This refers to the minimum number of training instances required such that $\forall h \in VS_{H,D}$, $error_{Đ}(h) < \varepsilon$.
    - **Result** Let m training samples be independently and randomly drawn from distribution $Đ$, and let H be the hypothesis space constructed, then

      $$p\left(VS_{H,D} \text{ is } \varepsilon - exhausted\right) < |H| e^{-\varepsilon m}$$

      Using this result, we have $m \geq \frac{1}{\varepsilon}(ln|H| + ln\frac{1}{\delta})$

- **VC dimensions**
  Let H be our hypothesis space and x be a training instance. Their intersection is defined as

  $$H \cap x = \{h \cap x \mid h \in H\}$$

  We say that x is shattered by H if $|H \cap x| = 2^{|x|}$. The VC dimension of H is the largest cardinality of sets shattered by H. The upper bound is known to be $log_2|H|$.

Ideas Out of Lessons

Let H be set of lines in a plane. Let's assume, we are given some points in the plane, with positive and negative label and we want to separate the two classes. Then, maximum number of points in a plane that H can shatter is 3.