Is Learning Feasible?

Prof. Aritra Hazra

Thursday 07 January 2021

Summary of class

1 Question: Can we learn?.To be more specific,learning is feasible or not?

Answer: Let us understand with a related experiment For example: Let us take a vessel of balls with two colors(green,red)
Prob[getting a red marble]=μ
Prob[getting a green marble]=$1 - \mu$
Take sample of size **N**
Prob[getting a red marble from sample **N**]= λ.

1.1 Does value of λ predict the value of μ ?

Answer:There are two possibilities: (a) No!(possibility) Sample can be mostly green (b) Yes!(probability) value of λ can be likely close to value of μ. [possible versus probable]

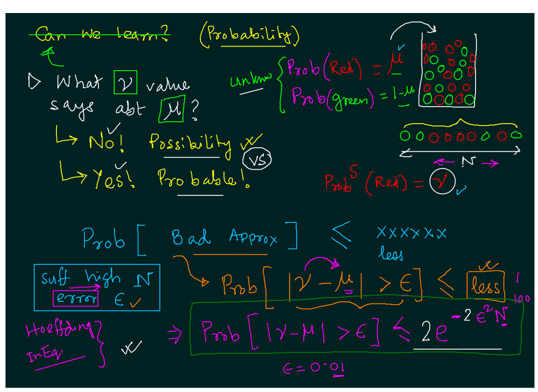

Figure 1: Related experiment

Let us take the case of possibility.In this sample N, the value of λ is probably close to μ.
P[Bad Aprrox]=P[$|\mu$ - $\lambda| >\in$]. Larger the value of sample set N, the error e can stoop down to 0.01.
$HoeffdingEquation$:
$P[|\mu$-$\lambda| >\in] \leq 2e^{-2\epsilon^2 N}$

We use **Hoeffding Inequality** to prove that learning is feasible.This equation provides an upper bound on the probability that the sum of bounded independent random variables deviates from its expected value by more than a certain amount(< 0.01).

Incidentally, we have a very handy tool for this: **the Hoeffding Inequality**. This equation depends on error$\in$ and the sample set of size N.

Probably approximately correct (PAC) learning theory helps analyze whether and under what condition

1.2 What is the connection to learning?

Answer: The Probability of getting red from a bin λ is unknown to us.The unknown function $f : X \mapsto Y$.

Let us go back to the related experiment(Figure: 2).There will be two hypothesis:(a)First hypothesis: we will pick the red marble(i.e,$h^R = f(x)$) (b) Second hypothesis will pick the green marble(i.e,h$^G \neq$ f(x)).
While predicting the value of unknown μ, we form the hypothesis set $[h_1, h_2, h_3, h_4......]$ denoted by $E_{in}(x)$=λ.
Hypothesis space is the set of all the possible legal hypothesis. This is the set from which the machine learning algorithm would determine the best possible (only one) which would best describe the target function or the outputs.

We can reduce the H.I.Equation to:

P[$|E_{in}(h) - E_{out}(h)| >\epsilon] \leq 2e^{-2\epsilon^2 N}$
The hypothesis set consists of all the possible hypothesis or $E_{in}(h)$ ["in sample"] which can lead to the outcome of predicting the value of unknown λ by diluting the other term of Hoeffding Inequation.
In the next and last section we will understand how we can reach using the hypothesis set to the abstract function g $\approx$ f with approximately zero error.

1.3 Producing the final hypothesis

Let us go back to the basic of designing machine learning model consisting of training data(provided by the probability distribution),Hypothesis

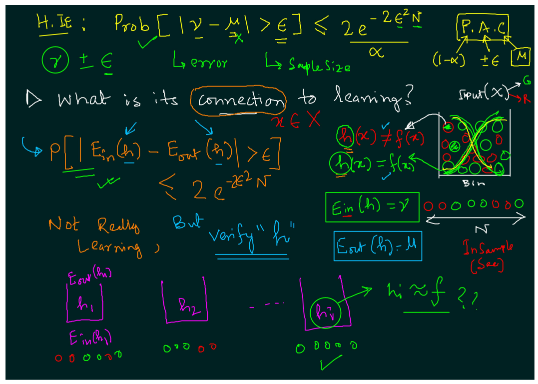

Figure 2: Hypothesis set

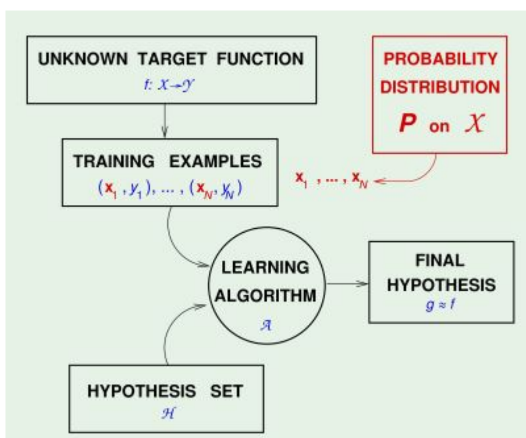

Figure 3: Machine Learning model

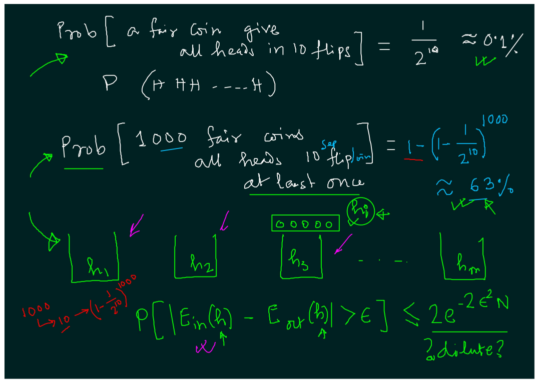

Figure 4: Examples of hypothesis set

set,Learning Algorithm.(shown in figure 3)

The concept of hypothesis can been explained by two simple examples of coins.[The coin analogy]given in the figure 4

Example 1: a simple probability question on fair coin
and Example 2: a probability which arises different set of hypothetical sample set to predict the correct output abstract function f $\approx$ g(The given/known output).
Therefore formulating into an equation:

P$[|E_{in}(h) - E_{out}(g = f)| > \epsilon]$ where $E_{in}(g)$ has h$_1, h_2, h_3, h_4.....h_M$
represented as hypothesis set H
which is expandable to summation of all the hypothesis set present in the Hypothesis set H i.e P$[|E_{in}(h_1) - E_{out}(h_1)| > \epsilon] \leq (P[|E_{in}(h_2) - E_{out}(h_2)| > \epsilon])$
$+($P$[|E_{in}(h_3) - E_{out}(h_3)| > \epsilon])+$........$+P[|E_{in}(h_m) - E_{out}(h_m)| > \epsilon] \leq$ 2M$e^{-2\epsilon^2 N}$
where M is no of hypothesis set in H

This value of M is the upper bound but not a generalised value.(shown in figure 5)

1.4 Conclusion

We can conclude few points from today's lecture

- The thing we learn is that the error is bounded or can be diluted on sufficient higher value of sample set N.
- Given the hypothesis set,the value of M is the upper bound and not a generalised.

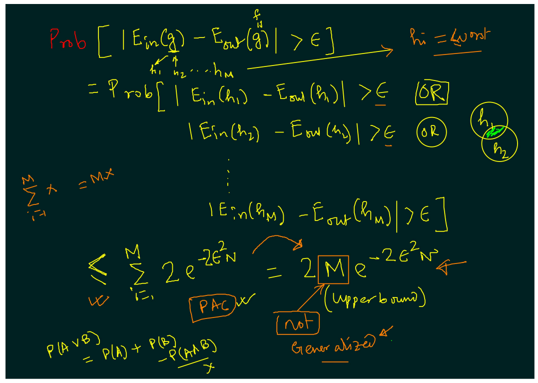

Figure 5: Final procedure of hypothesis

- Today's class enlightens us with the fact that learning is feasible and can be termed as Probably Approximate Correct Learning.

Prepared by Tanurima Halder(Roll No: 20CS92P04)