Class Notes - Basics of ML & NLP
Date: Feb 17, 2018
Notes taken by: Shishir Srivastava

In case of any queries/corrections,
please reach me at : replytoshishir@gmail.com

Presentation on PCA will be shared by the presenters.

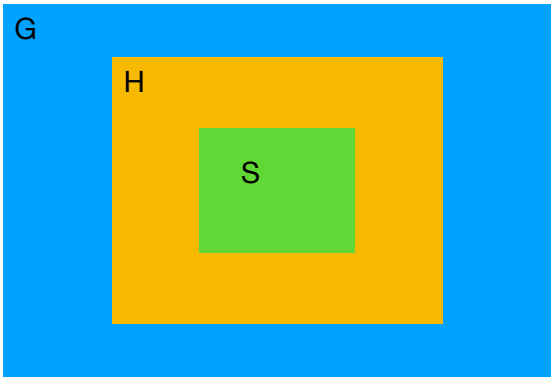
Machine learning can be defined with the help of 3 variables:
- Experience, E
- Decision task, T
- Performance mesare, P

If performance measure, P increases with experience,E, we call it machine learning.

The three major steps under machine learning includes:
- Representation, R
- Evaluation, E
- Optimisation, O

What is Noise? - errors
Noise can be because of following reasons:
1. Human error in
    1. Labeling
    2. Measuring
2. Hidden variables

<u>Observables and unobservable parameters</u>



G - the area after which negative cases starts

H- Hypothesis

S - area that covers only positives

<u>False positives</u>: this is decided to be positive by the system, whereas this should actually be negative. Since the positives lies in area S, but these are negatives, hence G will move in.

False negatives: decided to be negative, but actually this is positive. Since the negative lies outside the S area, but since this is actually positive S will expand.

X^t : Input data
r^t : Output (in training set)
$r^t \mid X^t$: Predict r^t given X^t

function, $f^t = r^t \mid X^t$

Representation are as follows:
$X^t = [\, x^t_1 + x^t_2 \,]$

$$r^t = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Evaluation: Finding S and G, get an optimal hypothesis, H

Regression representations:
$X^t = [x_1, x_2, x_3 \ldots x_n]$
$r^t = f(x^t, \beta^t)$
β^t is the evaluation

Optimisation techniques: the optimisation techniques are helpful in finding optimum solutions or unconstrained maxima and minima. One of the optimisation technique could be *"Least mean square error method"*. This is to evaluate value of β^t

Introduction to Bayes theorem:

$$P(A \mid B) = \frac{P(A \mid B) \,.\, P(B)}{P(A)}$$

P(A|B) is conditional probability.
Given sample, X^t, we need to find r.
$r \mid X^t$ is for one sample, calculating this for all other samples will give multiple values of β
We need to find probability of '*r*' given *'X'* => P(r|X)
Our intention is to maximise the P(r|X).

We use conditional probability to explain and can be expressed very well with the bayesian theorem, hence we call it "*Bayesian decision making*"

Terminologies:
P(B) : Prior
P(A|B) : Likelyhood
P(A) : Evidence