Weekly Report

Name: Utkarsh Patel
Roll No: 18EC30048
Summary of lectures given in **Week-5** (Sept 30, Oct 01)

Topics Covered

- **Basic Definitions**
  - **Bayes Rule**: It gives the expression for conditional probability P(B|A) = P(A & B)/P(A) It's also useful for Hypothesis H and evidence E as P(H | E) = (P(E|H) . P(H))/P(E)
  - **Random Variable (X)**: It is basically the set of values of outcomes of the random experiment we are performing.
  - **Expectation**: It is the weighted mean of the values of random variables where weights are their respective probabilities.
  - **Variance**: In simple words it is the expectation of the sum of the squares of difference of random variable values and the mean (expectation of random variable).

$$\text{var}(X) = E((X - E[X]^2))$$

- **Distributions**
  It is a set of all possible random variable along with their probability of occurrence for the random experiment under consideration. There are many types of distributions which is divided by the type of random variable. It represents
  - discrete random variable
  - continuous random variable

- **Discrete Random Variables**
  - **Bernoulli Distribution:** It is the experiment where the outcome is binary i.e. success or failure. The random variable X denotes the outcome Success or Failure.
  - **Binomial Distribution:** It is a sequence of n binomial distributions. The random variable X denotes the number of success out these n outcomes.
  - **Poisson Distribution:** It is used to model the number of events occurring within a given time interval.

- **Continuous Random Variables**
  - **Normal Distribution (Gaussian Distribution**): It is the most useful distribution in practice, it describes how the values of a variable are distributed. We observe a peak near the mean (μ) and the congestion near the peak is controlled by the variance (σ).

- **Accuracy of a Hypothesis**
  Difficulties for this calculation are resolved by distinguishing between the true error Ed(h) of a model and the estimated or sample error Es(h) where Es(h) is the fraction of instances where there is a difference of observation between hypothesis and target function and Ed(h) is simply the probability of random variable X where $f(X) \neq h(X)$

  - **Statistical theory:** Ed(h) lies in close proximity of Es(h). With 95% probability Ed(h) lies in

$$E_s(h) \pm 1.96\sqrt{\frac{E_s(h)(1-E_s(h))}{n}}$$

    which is due to Es(h) follows Bernoulli distribution and the numerator in the square root represents the variance of distribution of Es(h).

  - **Confidence Interval:** In particular N% confidence interval means the interval in the Normal distribution graph where the area covered by the considered interval is N% of the total area (around the mean of distribution) which is same as N% probability. The interval having N% probability is $[M - Z_N\sigma,\ M + Z_N\sigma]$

  - **Central Limit Theorem:** It basically means if we have sufficient large samples with mean μ and variance standard deviation σ, then the distribution of the sample means will be approximately normally distributed.