Course Material

COURSE	**MATHEMATICS FOR COMPUTER ENGINEERS**
COURSE CODE	**21MAT31A**
MODULE	**V**
MODULE NAME	**SAMPLING DISTRIBUTION AND OPTIMIZATION TECHNIQUES**
STAFF INCHARGE	**Dr. Thriveni K , Yamuna B**

SAMPLING DISTRIBUTION(S)

Population: A large collection of individuals or attributes or numerical data can be regarded as population or universe. It is an aggregate of objects, animate or inanimate, under study. The population may be finite or infinite.

If the population is large, complete enumeration is not possible most of the times because of the cost involved, time consumed and also in some cases units are destroyed in the course of inspection (e.g., inspection of crackers). So we take help of sampling.

A finite subset of the population is known as sample.

Size of the population N is the number of objects or observations in the population. Population is said to be finite or infinite depending on the size N being finite or infinite.

Size of the sample is denoted by n. Sampling is the process of drawing samples from a given population.
Large sampling: If $n \geq 30$, the sampling is said to be large sampling.
Small sampling: If $n<30$, the sampling is said to be small sampling.

Examples:

1. Population of India is population whereas the population of Karnataka is sample.

2. Cars produced in India are the population whereas the mantis cars produced in India is sample.

The statistical constants of the population such as mean (μ), Standard deviation (σ) etc are called the parameters. Similarly, the constants for the sample drawn from the given population i.e., Mean (x) standard deviation (S) etc. are called statistics.

Random sampling:

The selection of an item from the population in such a way that each has the same chance of being selected is called random sampling.

Suppose we take a sample of size n from the finite population of size N. Random sampling is a technique in which each element has an equal chance of being selected.

Sampling where each member of a population may be chosen more than once is called *sampling with replacement* i.e. here the items are drawn one by one and are put back to the population before the next draw. If N is the size of the finite population and n is the sample size then we have N^n samples.

Sampling where if a member cannot be chosen more than once it is called **sampling without replacement**. Here the items are drawn one by one and are not put back to the population before the next draw. In this case there will be ${}^{N}C_{n}$ samples.

Sampling distribution:

Given a population, suppose we consider a set of samples of a certain size drawn from the population. For each sample, suppose we compute a statistics such as the mean, standard deviation etc., these statistics will vary from the sample to the other sample, suppose we group these statistics according to their frequencies and form a frequency distribution. The frequency distribution so formed is called a sampling distribution.

The standard deviation of sampling distribution is called its **Standard error**.
The reciprocal of the standard errors is called **precision**.

Sampling distribution of means:

Consider a population for which the mean is μ and the standard deviation is $\sigma_{\bar{x}}^2$ suppose we draw a set of samples of a certain size n, from this population and find the mean $\mu_{\bar{x}}$ of each of these population. The frequency distribution of these means is called a sampling distribution of means. Let the mean and the standard deviation of sampling distribution of means be $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}^2$ respectively.

Suppose the population is finite with size N or random sampling without the replacement i.e., the items drawn one by one and are not put back to the population before the next draw. In this case there will be ${}^{N}C_{n}$ samples and we have

$\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}}^2 = c\,\frac{\sigma^2}{n}$

where $c = \frac{N-n}{N-1}$ is called the finite population correction factor.

Note: If N is very large i.e., if the population is infinite or the sampling is finite with replacement then c is closer to 1 as $N \to \infty$

$\therefore\ \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}.$

So, the mean of sampling distribution is equal to population mean and the corresponding standard error is $\frac{\sigma}{\sqrt{n}}$ where σ is the standard deviation of the population.

If the population is distributed normally with mean μ and S.D. σ, then the mean of all positive random samples of size n are also distributed normally with mean μ and S.E $\sigma/\sqrt{n}$.

Central Limit Theorem:

This is a very important theorem regarding the distribution of the mean of a sample if the parent population is non-normal and the sample size is large.

If the variable X has a non-normal distribution with mean μ and standard deviation σ, then the limiting distribution of

$Z=\frac{x-\mu}{\sigma/\sqrt{n}}$, $n \to \infty$, is the standard normal distribution (i.e, with mean 0 and unit S D)

There is no restriction upon the distribution of X except that it has a finite mean and variance. This theorem holds well for a sample of 30 or more which is regarded as large.

Statistical Estimation is the method in which the parameters are estimated with the aid of the corresponding statistics. An estimate of the unknown true or exact value of the parameter or an interval in which the parameter is to be determined on the basis of sample data from the population.

Confidence interval:

Consider sampling distribution of a statistic S. Suppose S follows normal distribution.
Let μ_s and σ_s be the mean and S.D. of the normal distribution.

a. The probability that μ_s lies in the interval ($s-\sigma_s$, $s+\sigma_s$) is 68.26% (i.e., $Z_c = 1$)
b. The Probability that μ_s lies in the interval ($s-2\sigma_s$,s+2) is 95.44% (i.e., $Z_c = 2$)
c. The probability that μ_s lies in the interval ($s-3\sigma_s$, $s+3\sigma_s$) is 99.74% (i.e., $Z_c = 3$).

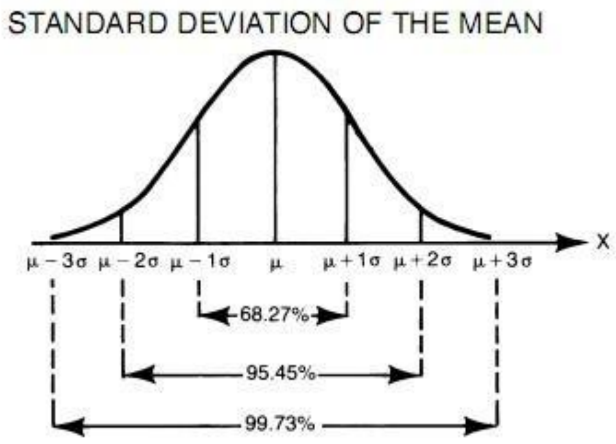


Figure 2

Percent	99.73%	99%	95.45%	95%	90%	80%	68.27%
No. of ± σ's	3.00	2.58	2.00	1.96	1.645	1.28	1.00

Z% Confidence interval:

The Z% confidence interval for μ_s if P { ($s - z_c\sigma_s$, $s+ z_c\sigma_s$) } = z%

i.e. $\frac{Z}{100} = P\{ -z_c\sigma_s \le \mu_s - s \le z_c\sigma_s \}$

$= P\{ |\frac{s-\mu_s}{\sigma_s}| \le Z_c \}$

$= P\{ |Z| \le Z_c \}$, where $Z = \frac{s-\mu_s}{\sigma_s}$

then $\frac{Z}{100} = P\{ -Z_c \le Z \le z_c \}$

$= 2P\{ 0 \le Z \le Z_c \}$

$= 2\phi(Z_c)$

$Z = 2\phi(Z_c) \times 100$

Confidence limit: The interval ($s - z_c\sigma_s$, $s + z_c\sigma_s$) is the Z% confidence interval for μ_s ,then the quantities ($s \pm Z_c\sigma_s$) are called Z% confidence limits. The member Z_c is called the corresponding confidence coefficient or the critical value confidence.

The length of the confidence interval ($s - z_c\sigma_s$, $s + z_c\sigma_s$) ie $2l = 2Z_c\sigma_s$ is called the error in the confidence level.

Table for the confidence coefficients Z_c for various values of Z

Z	Z_c	Z	Z_c
50	.6745	90	1.645
55	.7639	95	1.96
60	.843	95.44	2
65	.9259	96	2.05
68.26	1	97	2.195
70	1.041	98	2.33
75	1.15	99	2.58
80	1.277	99.5	2.81
85	1.445	99.74	3

Example1: A random sample of size N=100 is taken from a population with standard deviation $\sigma = 5.1$. Given that the sample mean is $\overline{X} = 21.6$. Obtain the 95% confidence interval for the population mean μ

$N=100, \sigma = 5.1, \overline{X} = 21.6$

Confidence limits for the population mean are

$$\overline{X} \pm Z_c \frac{s}{\sqrt{N}} = 21.6 \pm Z_c \frac{5.1}{\sqrt{100}}$$

For 95% confidence level, $Z_c = 1.96$

$\therefore \overline{X} \pm Z_c \frac{s}{\sqrt{N}} =$

$= 21.6 \pm (1.96) \frac{5.1}{\sqrt{100}}$

$= 21.6 \pm .9996$

Statistical Decision

Introduction:

For reaching statistical decisions, we start with some assumptions or guesses about the populations involved. Such assumptions / guesses, which may or may not be true, are called Statistical Hypotheses.

In many instances we formulate a statistical hypothesis for the sole purpose of rejecting or nullifying it. Such Hypotheses are called Null hypotheses and are generally denoted by H_0 .

A statistical hypothesis which differs from a given hypothesis is called an alternative hypothesis. A hypothesis alternative to the null hypothesis is generally denoted by H_1 .

Example:

i) Suppose we wish to reach the decision that a certain coin is biased (that is, thecoin shows more heads tails or vice versa). To reach this decision, we start with the hypothesis that the coin is faired (not biased) with the sole purpose of rejecting it (at theend). This hypothesis is a null hypothesis.

ii) Consider the situation where the probability of an event is, say, 1/3, according to some hypothesis. For arriving at some decision, if we make the hypothesis that theprobability is, say, ¼, then the hypothesis we have made is an alternative hypothesis.

Tests of hypothesis and significance:

Procedures which enable us to decide whether to accept or reject a hypothesis or to determine whether observed samples differ significantly from expected results are called tests of hypothesis, tests of significance, or rules of decision**.**

By an error of judgement, suppose we reject a hypothesis, when it should have been accepted. Then we say that an error of Type I have been made. On the other hand, suppose we accept a hypothesis when it should be rejected; in this case, we say that an error of Type II has been made.

Type I error:

In a hypothesis test, a type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H_0 is wrongly rejected.

For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; i.e.

H_0: there is no difference between the two drugs on average.

A type I error would occur if we concluded that the two drugs produced different effects

when in fact there was no difference between them.
The following table gives a summary of possible results of any hypothesis test:

		Decision	
		Reject H0	**Don't reject H0**
Truth	**H0**	Type I Error	Right decision
	H1	Right decision	Type II Error

A type I error is often considered to be more serious, and therefore it is more important to avoid than a type II error. The hypothesis test procedure is therefore adjusted so that there is a guaranteed 'low' probability of rejecting the null hypothesis wrongly; this probability is never 0. This probability of a type I error can be precisely computed as

P(type I error) = significance level $=\alpha$

The exact probability of a type II error is generally unknown.

If we do not reject the null hypothesis, it may still be false (a type II error) as the sample may not be big enough to identify the falseness of the null hypothesis (especially if the truth is very close to hypothesis).

For any given set of data, type I and type II errors are inversely related; the smaller the risk of one, the higher the risk of the other.

A type I error can also be referred to as an error of the first kind.

Type II error:

In a hypothesis test, a type II error occurs when the null hypothesis H0, is not rejected when it is in fact false. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; i.e.

H0: there is no difference between the two drugs on average.

A type II error would occur if it was concluded that the two drugs produced the same effect, i.e. there is no difference between the two drugs on average, when in fact they produced different ones.

A type II error is frequently due to sample sizes being too small.

The probability of a type II error is generally unknown, but is symbolised by β and written

P(type II error) = β

A type II error can also be referred to as an error of the second kind.

Levels of significance:

Suppose that, under a given hypothesis H, the sampling distribution of a statistic S is a normal distribution with mean μ_s and standard deviation σ_S. Then

$z = \frac{S-\mu_S}{\sigma_S}$ ----------- (1) is the standard normal variate associated with S, so that for the distribution of Z the mean is zero and the standard deviation is 1.

Accordingly, for the distribution of z, the z% confidence interval is ($-z_c$,z_c). This means that we can be Z% confident that, if the hypothesis H is true, then the value of z will lie between $-z_c$ and z_c . This is equivalent to saying that there is (100 – z)% chance that the hypothesis H is true but the value of Z lies outside the interval ($-z_c$, z_c). If we reject the hypothesis H on the grounds that the value of Z lies outside the interval ($-z_c$,z_c), we would be making a type I error and the probability of making the error is (100-Z)%. Her, we say that the hypothesis is rejected at a (100 - Z)% level of significance. Thus, a level of significance is the probability level below which we reject a hypothesis. In practice, only two levels of significance are employed: one corresponding to Z = 95and the other corresponding to Z= 99.

The value of the normal variate Z, determinate by using the formula (1) is usually called the z – score of the statistic S. It is this score that determines the "fate” of a hypothesis H and is called the test statistic.

Rule of decision:

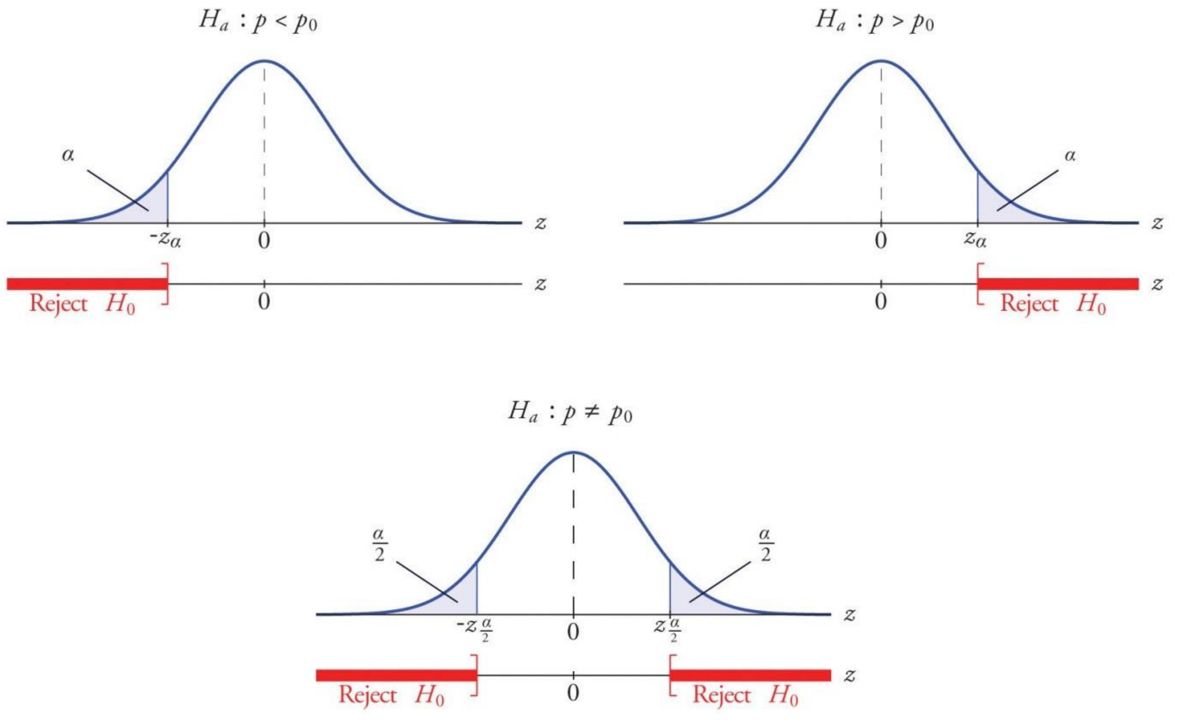
"Reject a hypothesis H at a (100 – Z)% level of significance if the z – score of the statistic S, determined on the basis of H, is outside the interval ($-z_c$, z_c). Do not reject the hypothesis otherwise".

Here the interval ($-z_c$,z_c) is called the interval of test.

Critical Region: The region in which a sample value falling is rejected, is known as critical region.

Normally, the test statistic, we consider follows normal distribution. Let us look into the normal curve.

If level of significance is α% then we have critical region as follows



Student 't' – distribution

Thus far, we considered sampling distributions on the assumption that they are normal or approximately normal. As mentioned before, this assumption is valid when the sample size N is large. For small samples, this assumption is not generally valid. In this section, we consider an important sampling distribution, called the 't-distribution', which is used in studies involving small samples.

Let N be the sample size, $\bar{X}$ and μ be respectively the sample mean and the population mean, and s be the sample standard deviation. Consider the statistic -t defined by

$$t= \left(\frac{\bar{X}-\mu}{s}\right)\sqrt{N-1}$$

Suppose we obtain a frequency distribution of t by computing the value of t for each of a et of samples of size N drawn from a normal or a nearly normal population. The sampling distribution so obtained is called the

Properties of the t Distribution

The t distribution has the following properties:
The mean of the distribution is equal to 0 .
The variance is always greater than 1, although it is close to 1 when there are many degrees of freedom. With infinite degrees of freedom, the t distribution is the same as the standard normal distribution.

Note: Table of values of $t_\beta(\gamma)$at $\beta = 0.01$ and 0.05 levels of significance for values ofγ from 1 to 50.

γ	$t_{0.01}(\gamma)$	$t_{0.05}(\gamma)$
1	63.66	12.71
2	9.92	4.30
3	5.84	3.18
4	4.60	2.78
5	4.03	2.57
6	3.71	2.45
7	3.50	2.36
8	3.36	2.31
9	3.25	2.26
10	3.17	2.23
11	3.11	2.20
12	3.06	2.18
13	3.01	2.16
14	2.98	2.14
15	2.95	2.13
16	2.92	2.12
17	2.90	2.11
18	2.88	2.10
19	2.86	2.09
20	2.84	2.09
21	2.83	2.08
22	2.82	2.07
23	2.81	2.07
24	2.80	2.06
25	2.79	2.06
26	2.78	2.06
27	2.77	2.05
28	2.76	2.05
29	2.76	2.04
30	2.75	2.04

Problems

Q.No		Question
1	a)	Explain the following i) Null hypothesis ii) Alternative hypothesis iii) Type I and type II error iv) Level of significance v) Standard error
	b)	A population has mean 75 and standard deviation 12. a) Random samples of size 121 are taken. Find the mean and standard deviation of the sample. b) How would the answers to part a) change if the size of the samples were 400 instead of121? Given μ=75, σ=12, n=121 $\mu\,\bar{x} = \mu = 75$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{121}} = \frac{12}{11} = 1.09$ n = 400 $\mu\,\bar{x} = \mu = 75$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{20}} = 0.6$ So if the size of the samples is changed from 121 to 400. $\sigma_{\bar{x}}$ decrease from 1.09 to 0.6.

2	a)	A population has mean 5.75 and standard deviation 1.02. a) Random samples of size 81 are taken. Find the mean and standard deviation of the sample. b) How would the answers to part a) change if the size of the samples were 25 instead of 81? $\mu = 5\cdot 75, 0 = 1.02$ $and\ n = 81$ $\mu\bar{x} = \mu = 5\cdot 75$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.02}{9} = 0\cdot 1133$ $\mu\bar{x} = \mu = 5\cdot 75$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.02}{95} = 0\cdot 204$ $\mu\bar{x}\ remains\ same\ but\ the\ \sigma_{\bar{x}}\ changes\ from\ 0.1133\ to\ 0.204$
	b)	The weights of 1500 ball bearings are normally distributed with a mean of 635 gms and S.D of 1.36gms. If 300 random samples of size 36 are drawn from this population, determine the expected mean and S.D of the sampling distribution of means if sampling is done a) with replacement b) without replacement. Here N=1500 μ = 635 σ= 1.36 n= 36 a) Expected Mean $\mu\bar{x} = \mu = 635$ $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} =$ 1.36/6=0.227 b) Expected Mean $\mu\bar{x} = \mu = 635$ $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$ $\sqrt{\frac{(1.36)^2}{36}} \times \sqrt{\frac{1500-36}{1500-1}}$ $= \sqrt{0.05}$ $= 0.224$

3	a)	A population consists of 4 numbers 3, 7, 11, 15. a) Find the mean and S.D. of the sampling distribution of means by considering samplings of size 2 with replacement. b) If n, n denotes respectively the population size and sample size, σ and $\sigma_{\bar{x}}$ respectively denotes population S.D. and S.D. of the sampling distribution of means without replacement $$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$ i. ii. $\mu_{\bar{x}} = \mu$ where $\mu_{\bar{x}}$ is the mean of this distribution and μ is the population mean. a) Population mean μ=1/4(3+7+11+15)=9 Population Variance $\sigma^2 = \frac{1}{4}\{(3-9)^2 + (7-9)^2 + (11-9)^2 + (15-9)^2\}$=20 a) Let us consider the sample of Size 2 with supplement. They are follows (3,3), (3,7), (3,11),(3,15), (7,3), (7,7), (7,11),(7,15), (11,3),(11,7),(11,11),(11,15), (15,3),(15,7),(15,11),(15,15) Sampling means are as follows 3,5,7,9, 5,7,9,11, 7,9,11,13, 9,11,13,15 The frequency distribution of the sampling means is as follows X: 3 5 7 9 11 13 15 f : 1 2 3 4 3 2 1 $\mu_{\bar{x}} = \sum \frac{f_i x_i}{f_i} = \frac{144}{16} = 9$

		$\sigma^2{}_{\bar{x}} = \sum \frac{f_i x_i}{f_i}$-$(\mu_{\bar{x}})^2$ =1456/16-9^2 =10 Thus $\mu_{\bar{x}} = 9$ $\sigma_{\bar{x}} = \sqrt{10}$ b) Let us consider the sample without replacement they are as follows (3,7),(3,11),(3,15),(7,11),(7,15),(11,15) The sampling means are 5,7,9,9,11,13 $\mu_{\bar{x}}$=1/6(5+7+9+9+11+13)=9 $\mu_{\bar{x}}$=μ $$\sigma^2{}_{\bar{x}} = \frac{1}{6}\{(5-9)^2 + (7-9)^2 + (9-9)^2 + (11-9)^2 + (13-9)^2\}$$ $$= \frac{40}{6} = \frac{20}{3}$$ Consider $\frac{\sigma^2}{n} x \left[\frac{N-n}{N-1}\right]$ $$= \frac{20}{2} x \left[\frac{4-2}{4-1}\right]$$ $$= \frac{\sigma^2}{n} x \left[\frac{N-n}{N-1}\right] = \sigma^2{}_{\bar{x}}$$

	b)	Certain tubes manufactured by a company have mean life time of 800 hours and S.D of 60hours. Find the probability that a random sample of 16 tubes from the group will have a mean life time a) between 790 hours and 810 hours b) less than 785 hours c) more than 820 hours d) between 770 hours and 830 hours. 3 b) By data μ=800, σ=60, n=15 Therefore $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{4} = 15$ $$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 800}{15} \sim N(0,1)$$ $$a)\ P(790 < \bar{x} < 810)$$ $$= p\left(\frac{790 - 800}{15} < z < \frac{810 - 800}{15}\right)$$ $$= P(0 \cdot 67 < z < 0 \cdot 67)$$ $$= p(0 \cdot 67 < z < 0 \cdot 67)$$ $$= 0.7486$$ $$b)\quad P(\bar{z} < 785)$$ $$= p\left(z < \frac{785 - 800}{15}\right)$$ $$= P(z < -1)$$ $$= P(z > 1)$$ $$= 0.1587$$ $c)$ $=P(\bar{x}>820)$ $= P\left(z > \frac{820-800}{15}\right)$ $$= P(z > 1 \cdot 33)$$ $$= 0.0918$$

		d) $$= P(770 < \bar{x} < 880)$$ $$p\left(\frac{770-880}{15} < \frac{830-800}{15}\right)$$ $$= P(-2 < z < 2)$$ $$= 2P(0 < z < 2)$$ $$= 0 \cdot 9772$$
4	a)	A prototype automotive tire has a design life of 38500 miles with S.D. of 2500 miles. Five suchtires are manufactured and tested. On the assumption that the actual population S.D. is 2500 miles, find the probability that the sample mean will be less than 36000 miles. Assume that the distribution of lifetimes of such tires is normal. 4a) For simplicity we use units of thousands of miles Then $\mu_{\bar{x}}$ = Mean of sample mean $\mu = 38 \cdot 5$ Thousands of miles $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2\cdot 5}{\sqrt{5}} = 1 \cdot 11803$ thousands of miles

		These normally distributed $= P(\bar{x} < 86)$ $$\bar{x}\ is\ also\ the\ following\ normal\ distribution$$ $$= P(\bar{x} < 86)$$ $$= P\left(z < \frac{36 - \mu\bar{x}}{\sigma_{\bar{x}}}\right)$$ $$= P(z < -2 \cdot 24)$$ = 0.0125 That is, if the time perfom ad designed, there is only about a 125% chance that the average of a sample of this size would be so low
	b)	An automobile battery manufacturer claims that its midgrade battery has a mean life of 50 months with a S.D. of 6 months. Suppose the distribution of battery lives of this particular brand is approximately normal. a) On the assumption that the manufacturer claims are true, find the probability that a randomly selected battery of this type will last less than 48 months. b) On the same assumption, find the probability that mean of a random sample of 36 such batteries will be less than 48 months. Given μ=Mean life of a battery=50 months σ = Standard deviation= 6 months Let X : life of a battery $$x \sim N(\mu, \sigma^2)$$ $$\frac{x - \mu}{\sigma} = 2 \sim N(0,1)$$ $$= p(x < 48)$$ $$= P\left(z < \frac{48 - \mu}{\sigma}\right)$$ $$= p(z < -0.33)$$ $$= 0 \cdot 3707$$

		n = Sample Size =36 Therefore, Sample Mean = $\mu_{\bar{x}} = \mu = 50$ $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = 1$$ $$\bar{x} \sim \frac{\mu \bar{x}}{\sigma_{\bar{x}}} = z \sim N(0,1)$$ $$= P(\bar{x} < 48)$$ $$= P\left(z < \frac{48-50}{1}\right) \cdot$$ $$= P(z < -2)$$ $$= 0.0228$$
5	a)	The weights of 1500 ball bearings are normally distributed with a mean of 635 gms and S.D. of 1.36 gms. If 300 random samples of size 36 are drawn from this population. In the case of random sampling with replacement, find how many random samples would have their mean a)between 634.76gms and 635.24 gms, b) greater than 635.6 gms, c)less than 634.5 gms or more than 635.24 gms X= wt of ball bearing N-pop Size =1500 μ= 635 gm σ= 1.36 gm No of random sample =300 n= sample size=36 $$X \sim N(\mu, \sigma^2)$$ $$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$$ $$\mu_{\bar{x}} = \mu = 635$$ $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1 \cdot 36}{6} = 0.2267$$ $$z = \frac{\bar{x} - 635}{0 \cdot 2267} \sim N(0,1)$$

i) The no of random samples have their mean between 634.75 g & 635.24 g

$= \text{No.Of samples} \times p\{634.76 < X < 635.24\}$

$= 300 \times P[\ 1.0587 < z < 1.0587]$

$=300 \times 0.8554$

$=256.62$

$=257$

ii) No of Samples that have there mean greater than 635.6 g

$= 300 \times P[X>635.6]$

$= 300 \times P\ [\ Z> 2.6467]$

$= 300 \times [0.5-0.4960]$

$= 300 \times 0.004$

$=1.2$

$=1$

		d)No of Samples with mean less than 634.5 g or more than 635.24g =300 x [P[$\bar{x} < 634.5] + P[\bar{x} > 635.24$]] =300 x [p[z<-2.21]+P[z>1.06]] = 300 x [0.0136+0.1446] = 300 x[0.1582] =47.46
	b)	500 ball bearings have a mean weight of 142.30 gms and S.D. of 8.5 gms. Find the probability that a random sample of 100 ball bearings chosen from this group will have a combined weight a) between 14061 and 14175 gms b) more than 14460 gms Population size =N=500 μ=142.3 gms σ= 8.5 gms sample size =n=100 x: wt of ball bearing n is large by central limit theorem $$\frac{\bar{x} - \mu\bar{x}}{\sigma_{\bar{x}}} \sim N(0,1)$$ $$\mu_{\bar{x}} = \mu = 142 \cdot 3gm$$ $$\sigma_{\bar{x}} = \frac{8 \cdot 5}{\sqrt{n}} = 0 \cdot 85gm$$

		$$z=\frac{\bar{x}-142\cdot 3}{0\cdot 85}\sim N(0,1)$$ i) P[the combined wt of the group lies between 140.61 gms &141.75 gms] $=P[140.61/n<x,141.75\ /n]$ $=P[140.61<x<141.75]$ $=p[-1.988<Z<-0.647]$ =0.2345 ii) P[THE COMBINED WT OF THE GROUP IS MORE THAN 144.60 GMS] $=P[X>144.6]$ $=P[Z>2.71]$ = 0.0034
6	a)	The mean and S.D of the maximum loads supported by 60 cables are 11.09 tonnes and 0.73 tonnes respectively. Find a) 95% b) 99% confidence limits for mean of the maximum loads of all cables by the company.

		6A) By data X max load supported by a cable then $$\bar{x} = 11.09, \sigma = b0.73\ , n = 60$$ a) 95% confidence limits for the mean of maximum loads are given by $$x \pm 196(\sigma/\sqrt{n})$$ $$\pm 11.09 \pm 1 \cdot 96(0 \cdot 73/\sqrt{60})$$ $$\bar{a}, 1.09 \pm 0 \cdot 18$$ Limits are 10.91 tonner & 11.27 tonner b) 99% confidence limits for the mean of maximum loads are given by $$= \bar{x} \pm 2 \cdot 58(\sigma/\sqrt{n})$$ $$= 11.09 \pm 2 \cdot 58(0 \cdot 73/\sqrt{60})$$ $$= 11.09 \pm 0 \cdot 24$$ $$Limits\ are\ 10.85\ tonners\ \&\ 11.33\ tonners$$
	b)	A sample of 900 men is found to have a mean height of 64inch. If this sample has been drawn from a normal population with standard deviation 20 inch, find the 99% confidence limits for the mean height of the men in the population. 6 b) Given n=900 Let x: ht $$\bar{x}\ = 64\ inch, \sigma =\ 20\ inch$$ 99% confidence limits for the mean is given by $$= \bar{x} \pm 2 \cdot 58(\sigma/\sqrt{n})$$ $$= 64 \pm 2 \cdot 58(20/30)$$ $$= 64 \pm 2 \cdot 58(0.6667)$$ $$= 64 \pm 1 \cdot 720086$$ The limits are 62.279914 & 65.720086
7	a)	A sample of 5000 students in a college was taken and their average height was found to be 62.5Kg with a standard deviation of 22kg. Find the 95% confidential limits of the average weight of the students in the entire University.

		7 a) n=50000, $\bar{x} = Average\ wt = 62.5\ kg$ Σ= 22 kg a) 95% Confidence limits for the mean of maximum loads are given by $\bar{x} \pm 1.96(22/\sqrt{n})$ $= 62 \cdot 5 \pm 1 \cdot 96(22/\sqrt{5000})$ $= 62 \cdot 5 \pm 0 \cdot 6098$ $Limits\ are\ 61.8902\ \&\ 63.1098$
	b)	Systolic blood pressure of 566 males was taken. Mean BP was found to be 128.8mm and SD 13.05mm. Find 95% confidence limits of BP within which the populations mean would lie. 7 b) Let x: systolic blood pressure n= 566 , $\bar{x} = 128.8\ mm, \sigma = 13.05\ mm$ $\bar{x} \pm 1.96(\sigma/\sqrt{n})$ $= 129 \cdot 8 \pm 1 \cdot 96(13.05/\sqrt{566})$ $= 128 \cdot 8 \pm 1.07506$ $Limits\ are\ 127.72494\ \&\ 129.87506$
8	a)	Standard deviation of blood sugar level in a population is 6 mg%. If population mean is not known, within what limits is it likely to lie if a random sample of 100 has a mean of 80mg%? 8 a) x : blood sugar level n= 100, $\bar{x} = 80\ mg, \sigma = 6\ mg$ 95% Confidence limits for the mean is given by $\bar{x} \pm 1.96(\sigma/\sqrt{n})$ $= 80 \pm 1.176$ The limits are 78.824 & 81.176
	b)	To know the mean weights of all 10 year old boys in Delhi a sample of 225 was taken. The mean weight of the sample was found to be 67 pounds with s.d. of 12 pounds. What can we infer about the mean weight of the population?

		8 b) x : wt of a 10 years old boy in Delhi n = 22.5 $\bar{x} = 67 pounds, \sigma = 12\ pounds$ 95% Confidence limits of x is $$\bar{x} \pm 1.96(\sigma/\sqrt{n})$$ = 67 $\pm 1.96(0.8)$ = 67 $\pm$ 1.568 The limits are 65.432 pounds & 68.568 pounds We can interfere that pop mean wt lies in (65.432, 68.568) 99% Confidence limits of x is $$\bar{x} \pm 2.58(\sigma/\sqrt{n})$$ = 67 $\pm 2.58(0.8)$ = 67 $\pm$ 2.064 The limits are 64.936 & 69.064 pounds We can interfere that pop mean wt lies in (64.936, 69.064) pound with 99 % confidence
9	a)	The mean and S.D of the diameters of a sample of 250 rivet heads manufactured by a company are 7.2642 mm and 0.0058mm respectively. Find (a) 99% (b) 95% confidence limits for the mean diameter of all the rivet heads manufactured by the company. X: Diameter of rivert read n = 250 , $\bar{x} = 7.2642\ mm, \sigma = 0.0058\ mm$ 95% Confidence limits of x is = 7.2642 $\pm 1.96(0.0058/\sqrt{250})$ = 7.2642 $\pm$ 0.00072 The limits are 7.26348 mm & 7.26492 mm

		99% Confidence limits of x is $$\bar{x} \pm 2.58(\sigma/\sqrt{n})$$ $= 7.2642 \pm 2.58(0.0058/\sqrt{250})$ $= 7.2642 \pm 0.00095$ The limits are 7.26325 mm & 7.26515 mm
	b)	Spring break can be a very expensive holiday. A sample of 80 students is surveyed, and the average amount spent by students on travel and beverages is \$593.84. The sample standard deviation is approximately \$369.34.Construct a 95% confidence interval for the population mean amount of money spent by spring breakers. X: Money spent by a student in a spring break. n = 80 , $\bar{x} = 593.84$, $\sigma = 369.34$ 95% Confidence limits of x is $$\bar{x} \pm 1.96(\sigma/\sqrt{n})$$ $= 593.84 \pm 1.96(369.34/\sqrt{80})$ $= 593.84 \pm 80.9352$ The limits are 512.9048 & 674.7752
10	a)	400 items are sampled from a normally distributed population with a sample mean x of 22.1 and a population standard deviation(σ) of 12.8. Construct a 95% confidence interval for the true population mean. n = 400 , $\bar{x} = 22.1$, $\sigma = 12.8$ 95% Confidence limits of x is $$\bar{x} \pm 1.96(\sigma/\sqrt{n})$$ $= 22.1 \pm 1.96(12.8/\sqrt{400})$ $= 22.1 \pm 1.2544$ The limits are 20.8456 & 23.3544

	b)	The mean and S.D. marks of a sample of 100 students are 67.45 and 2.92 respectively. Find (a) 95% (b) 99% confidence intervals for estimating the marks of the population. n = 100 , $\bar{x} = 67.45, \sigma = 2.92$ 95% Confidence limits of x is $$\bar{x} \pm 1.96(\sigma/\sqrt{n})$$ $= 67.45 \pm 1.96(2.92/\sqrt{100})$ $= 67.45 \quad \pm 0.57232$ The limits are 66.87768 & 68.02232 99% Confidence limits of x is $$\bar{x} \pm 2.58(\sigma/\sqrt{n})$$ $= 67.45 \pm 2.58(2.92/\sqrt{100})$ $= 67.45 \pm 0.75336$ The limits are 66.69664 & 69.20336
11	a)	A machine is expected to produce nails of length 3 inches. A random sample of 25 nails gave an average length of 3.1 inch with standard deviation 0.3. Can it be said that the machine is producing nails as per specification?($t_{0.05}$ for 24 d.f. is 2.064) a) Let x length of nail µ pop mean of x =3 inch n= 25 s = sample std deviation =0.3 $$\bar{x} = 3 \ inch$$ **t** = $\bar{x} - \mu/s \ (\sqrt{n})$ $= 0.1/0.3 \sqrt{25} = 1.67 - 2.064 = t_{0.05}, df$ Thus the hypo that the machine is producing nails are per specifications is accepted At 5% level of significance.

	b)	Ten individuals are chosen at random from a population and their heights in inches are found to be 63,63,66,67,68,69,70,70,71,71. Test the hypothesis that the mean height of the universe is 66 inches. ($t_{0.05}$=2.262 for 9 d.f.) 11 b) x: height of an individual in inches μ =66 inchs n= 10 $\bar{x} = \frac{\sum x_i}{n} = 67 \cdot 8$ $s^2 = \frac{1}{(n-1)} \Sigma(x;-\bar{x})^2$ $= 9 \cdot 0^{67}$ $s = 3.011$ $t = \frac{\bar{x}-\mu}{s} \cdot \sqrt{n}$ $= \frac{678-66}{3 \cdot 0/1}\sqrt{10} = 1.89 < 2.262$ $= t_{0.05} \ / \ 9\, df$
12	a)	A certain stimulus administered to each of 12 patients resulted in the following increases of blood pressure: 5,2,8,-1,3,0,-2,1,5,0,4,6. Can it be concluded that the stimulus will in general be accompanied by an increase in blood pressure. ($t_{0.05}$ for 11 d.f. is 2.2) Given n=12 X: increase in blood pressure $\bar{x} = \frac{\sum x_i}{n}$= [5+2+8-1+0-2+1+5+0+4+6]/12 $= 31/12 = 2.5833$ $s^2 = \frac{1}{(n-1)} \Sigma(x;-\bar{x})^2$ $= \frac{1}{(n-1)}\left(\Sigma \zeta_i^2 - \frac{1}{n}(\Sigma x_i)^2\right)$ $= \frac{1}{11}\left(185 - \frac{1}{12}(31)^2\right)$ $= 9 \cdot 3 - 38$ $s = 3.088$ $\mu = 0$ $t = \frac{\bar{x}-\mu}{s}\sqrt{n}$ $= 2 \cdot 8979 > 2 \cdot 201_{11\,df}$ $= \quad t_{0.05}$ The hype is rejected it with 95% confidence we can say that the stimulus in general is accompanied with Increase in bp.

	b)	A machinist is making engine parts with axle diameter of 0.7 inch. A random sample of 10 parts shows mean diameter 0.742 inch with a standard deviation of 0.04 inch. On the basis of this sample, would you say that the work is inferior? ($t_{0.05}$=2.262 for 9 d.f.) x: axle diameter n=10 $\bar{x} = 0.742$ S = 0.04 μ = 0.7 $t = \frac{\bar{x} - \mu}{s} \cdot \sqrt{n}$ = 0.042/0.04 $\sqrt{10}$ = 3.3204> 2.262 = $t_{0.05}$ $^{9\,df}$ The hypo is rejected 0.7 on the basis of sample we can say that the work is interior
13	a)	Show that 95% confidence limits for the mean μ of the population are $x \pm \frac{\sigma s}{\;} t_{0.05}$. If $t_{0.05}$ is the tabulated value of t for n-1 degrees of freedom at 5% 1.0.3 ⇒ P[1 t 1 > $t_{0.05}$] = 0.05 ⇒ P[1 t 1 ≤ $t_{0.05}$] = 1 - 0.05 = 0.95 i.e with 95% confidence 1 t 1 ≤ $t_{0.05}$ ⇒ $\left\lvert \frac{\bar{x} - \mu}{\sigma_s} . \sqrt{n} \right\rvert \le t_{0.05}$ ⇒ $- t_{0.05} \le \frac{\bar{x} - \mu}{\sigma_s} . \sqrt{n} \le t_{0.05}$ ⇒ $- \frac{\sigma_s\, t_{0.05}}{\sqrt{n}} \le \bar{x} - \mu \le \frac{\sigma_s}{\sqrt{n}} t_{0.05}$ ⇒ $\bar{x} - \frac{\sigma_s\, t_{0.05}}{\sqrt{n}} \le \mu \le \bar{x} + \frac{\sigma_s}{\sqrt{n}} t_{0.05}$ ∴ 95% confidence limits of μ is $\bar{x} \pm \frac{\sigma_s\, t_{0.05}}{\sqrt{n}}$

	b)	A random sample of 10 measurements of the diameter of a sphere gave a mean of 12 cm and standard deviation 0.15 cm. Find 95% confidence limits for the actual diameter. ($t_{0.05}$=2.262 for 9 d.f.) **13 b)** X = diameter of the sphere n = 10; $\bar{x} = 12\ cm$; S = 0.15 cm ∴ 95% confidence limits are $\bar{x} \pm \frac{S}{\sqrt{n}}\ t_{0.05}$ i.e $12 \pm \frac{12}{\sqrt{10}}$ (2.262) i.e 12 ± 0.1073 i.e 11.8927, 12.1073
14	a)	A random sample of 10 boys had the following I.Q.: 70, 120, 110, 101, 88, 83,95, 98, 107,100. Do these data support the assumption of a population mean I.Q. of 100 at 5% level of significance ($t_{0.05}$=2.262 for 9 d.f.) **14 a)** X = IQ of a boy $\bar{X}$ = (70 + 120 + 110 + 101 + 88 + 83 + 95 + 98 + 107 + 100) / 10 = 972 / 10 = 97.2 $S^2 = \frac{1}{9}\ [96312 - \frac{1}{10}\ 97.2^2$ = 203.7333 ∴ S = 14.2735 ∴ 1 t 1 = $\frac{2.8}{14.273}\ \sqrt{10}$ = 0.6203 < 2.262 ∴ Hypothesis that the population mean IQ of 100 at 5% level of significance is accepted
	b)	A random sample of size 25 from a normal population has the mean 47.5 and s.d 8.4. Does this information refute the claim that the mean of the population is 42.1.($t_{0.05}$=2.064 for 9 d.f.) **14 b)** n = 25; $\bar{X}$ = 47.5; S = 8.4 μ = 42.1 $\therefore t = \frac{\overline{X - \mu}}{s} . \sqrt{n}$ $= \frac{47.5-42}{8.4} . \sqrt{25}$ = 45 / 14 = 3.2143 > 2.064 = $t_{0.05}$; d.f = 24 ∴ Hypothesis is rejected at 5% 1.0.3

15	a)	A process for making certain bearings is under control if the diameter of the bearings have themean 0.5 cm. What can we say about this process if a sample of 10 of these bearings has a mean diameter of 0.506 cm. and S.D. of 0.004cm?($t_{0.05}$=2.262 for 9 d.f.) **15 a)** X = diameter of bearing μ = 0.5; n = 10; $\bar{X} = 0.506$; s = 0.004 $\therefore t = \frac{\overline{X-\mu}}{s} . \sqrt{n}$ $= \frac{0.006}{0.004} . \sqrt{10}$ = 4.743 > 2.262 ∴ The hypothesis is rejected at 5% 1.0.3 i.e the process is not under control
	b)	A machine is supposed to produce washers of mean thickness 0.12cm. A sample of 10 washerswas found to have a mean thickness of 0.128cm and standard deviation 0.008. Test whether themachine is working in proper order at 5% level of significance.. ($t_{0.05}$=2.262 for 9 d.f.) **15 b)** X = thickness of washer μ = 0.12; n = 10 ; $\bar{X} = 0.128$; s = 0.008 $\therefore t = \frac{\overline{X-\mu}}{s} . \sqrt{n}$ $= \frac{0.008}{0.008} . \sqrt{10}$ = 3.1623 > 2.262 = $t_{0.05}$, 9 df ∴ The hypothesis is rejected i.e the machine is not working in proper order at 5% 1.03

Optimization Technique:

Introduction:

- The objective of optimization is to deal with real life problems
- It means getting the optimal output for problems
- In machine learning, optimization is slightly different. Generally, while optimizing, we know exactly how our data looks like and what areas we want to improve, but in machine learning we have no clue how our "new data" looks like, let alone try to optimize on it. Therefore, in machine learning, we perform optimization on the training data and check its performance on a new validation data.

There are various kinds of optimization techniques, which is as follows

- Mechanics Deciding the surface of aerospace design
- Economics Cost Optimization
- Physics Time optimization in quantum computing
- Various popular machine algorithm depends upon optimization techniques like linear regression, neural network, K nearest neighbor etc
- Gradient descent is the most common used optimization techniques in machine learning Optimization Techniques

Gradient Descent Algorithm :

Gradient Descent is an optimization algorithm **for finding a local minimum of a differentiable function**. Gradient descent is simply used in machine learning to find the values of a function's parameters (coefficients) that minimize a cost function as far as possible.

Suppose a large bowl like what you would eat cereal out of or store fruit in This bowl is a plot of the cost function (f)

- A random position on the surface of the bowl is the cost of the current values of the coefficients (cost)
- The bottom of the bowl is the cost of the best set of coefficients, the minimum of the function.

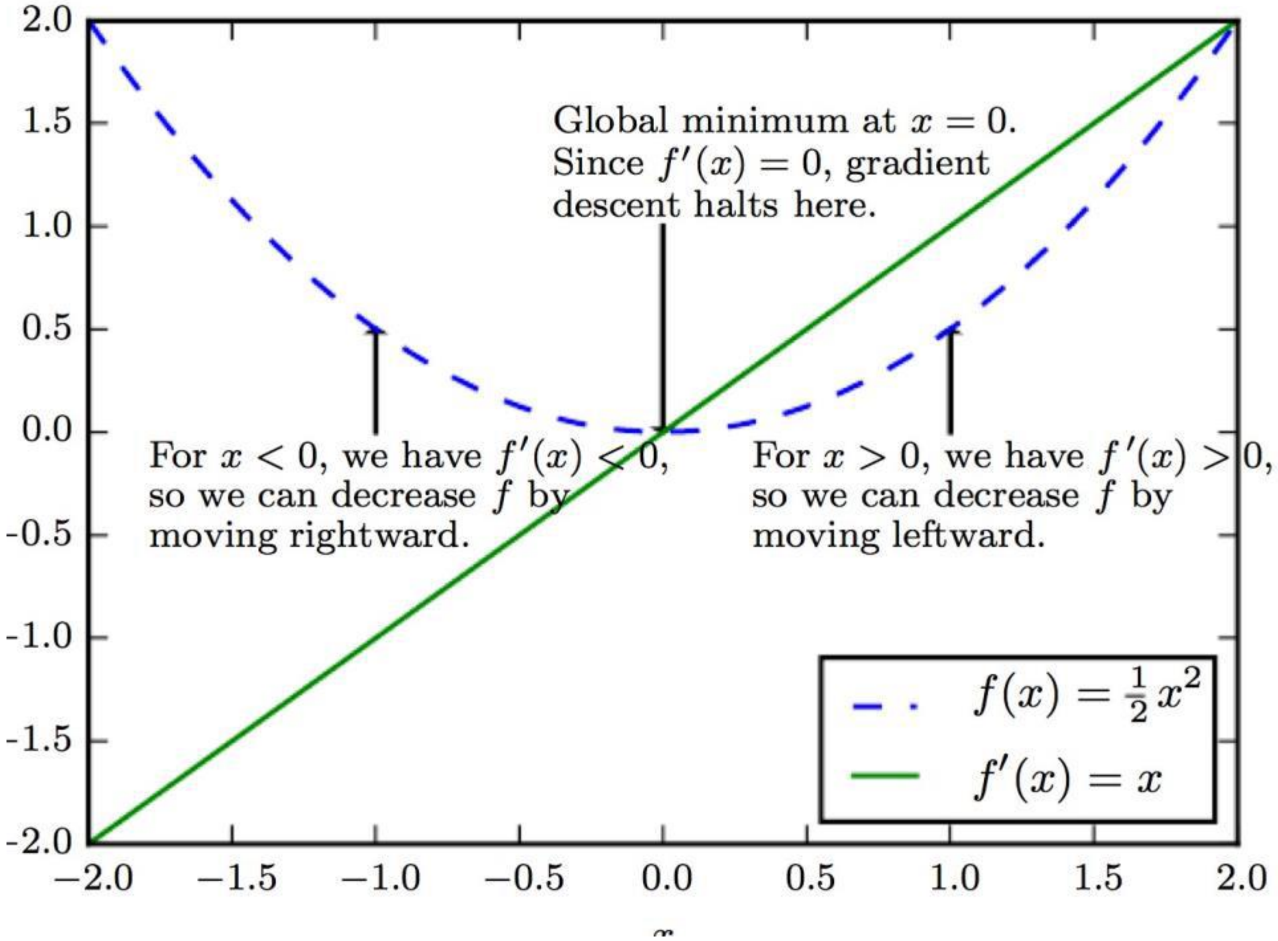
- The goal is to continue to try different values for the coefficients, evaluate their cost and select new coefficients that have a slightly better (lower) cost.

- Repeating this process enough times will lead to the bottom of the bowl and you will know the values of the coefficients that result in the minimum cost.

Given function is f (x) ½ x ^{2} which has a bowl shape with global minimum at x =0, Since f'(x)= x

- For x > f (x) increases with x and f'(x)= 0
- For x < f x) decreases with x and f'(x)= 0
- Use f' (x) to follow function downhill

Reduce f x) by going in direction opposite sign of derivative f' (x)



- We often minimize functions with multiple inputs f :R ^{n} →R
- For minimization to make sense there must still be only one (Scalar) output.

Types of Gradient Descent Algorithms:

It can be classified by two methods:

- Batch Gradient Descent Algorithm
- Stochastic Gradient Descent Algorithm
- Batch gradient descent algorithms, use whole data at once to compute the gradient, whereas in stochastic you take a sample while computing the gradient.

Batch Gradient Descent:

- The objectives of all supervised machine learning algorithms is to best estimate a target function (f) that maps input data (X) onto output variables (Y)
- Some machine learning algorithms have coefficients that characterize the algorithms estimate for the target function (f).

Different algorithms have different representations and different coefficients, but many of them require a process of optimization to find the set of coefficients that result in the best estimate of the target function
Examples of algorithms with coefficients that can be optimized using gradient descent are

- Linear Regression
- Logistic Regression

Advantages of Batch gradient descent:

- It produces less noise in comparison to other gradient descent.
- It produces stable gradient descent convergence.
- It is Computationally efficient as all resources are used for all training samples.

Mini-Batch Gradient Descent:

Mini Batch gradient descent is the combination of both batch gradient descent and stochastic gradient descent. It divides the training datasets into small batch sizes then performs the updates on those batches separately. Splitting training datasets into smaller batches make a balance to maintain the computational efficiency of batch gradient descent and speed of stochastic gradient descent. Hence, we can achieve a special type of gradient descent with higher computational efficiency and less noisy gradient descent.

Advantages of Mini Batch gradient descent:

- It is easier to fit in allocated memory.
- It is computationally efficient.
- It produces stable gradient descent convergence.

Stochastic Gradient Descent:

- Gradient descent can be slow to run on very large datasets
- One iteration of the gradient descent algorithm requires a prediction for each instance in the training dataset, it can take a long time when you have many millions of instances
- When large amounts of data, you can use a variation of gradient descent called stochastic gradient descent
- A few samples are selected randomly instead of the whole data set for each iteration In **Gradient Descent** there is a term called “batch” which denotes the total number of samples from a dataset that is used for calculating the gradient for each iteration.

- Stochastic gradient descent selects an observation uniformly at random, say i and uses fi(w) as an estimator for F(w) While this is a noisy estimator; we are able to update the weights much more frequency and therefore hope to converge more rapidly.

- Updates takes only O(d) computation, though the total number of iterations, T, is larger than in the Gradient Descent algorithm Stochastic Gradient Descent.

Advantages of Stochastic gradient descent:

In Stochastic gradient descent (SGD), learning happens on every example, and it consists of a few advantages over other gradient descent.

It is easier to allocate in desired memory.
It is relatively fast to compute than batch gradient descent.
It is more efficient for large datasets.

Gradient Descent Algorithm :

1: **While** Stopping criterion not met **do**
2: Initialize parameter updates $\Delta\theta_t = 0$
3: for each (x^i, y^i) in D_{tr} do
4: Compute gradient using back propagation $\nabla_{\theta t}\, L(\theta_t; x^i, y^i)$
5: Aggregate gradient $\Delta\theta_t = \Delta\theta_t + \nabla_{\theta t} L$
6: end for
7: Apply update $\nabla_{t+1} = \theta_t \; - \alpha \frac{1}{|\mathrm{D}_{tr}|} \nabla\theta_t$
8: end while