Essential Mathematics for Machine Learning

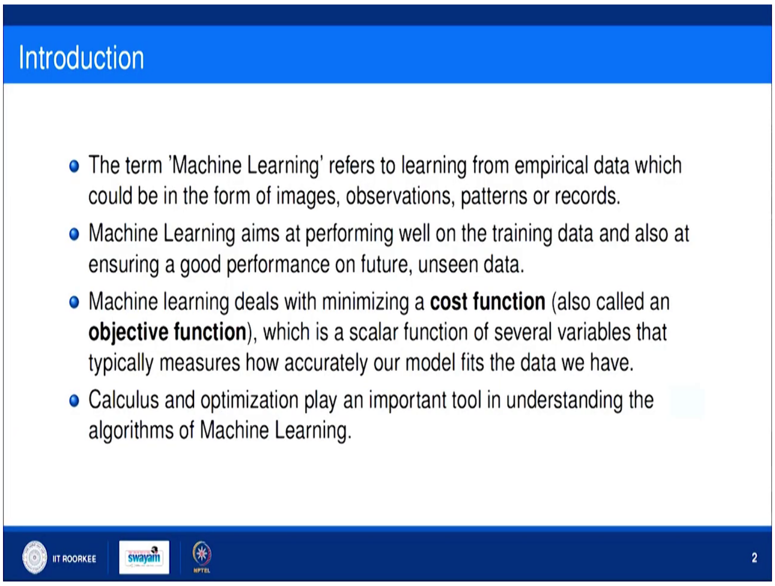
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 21

Basic Concepts of Calculus - I

Hello everyone. So, welcome to a lecture series on Essential Mathematics for Machine Learning. So, this is the 1st lecture on Basic Concepts of Calculus. So, let us start what the essential part of calculus which is used in machine learning.

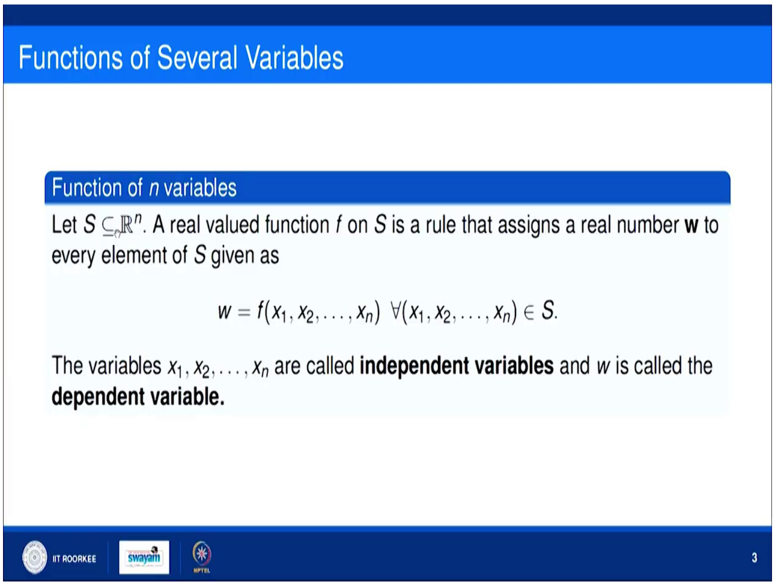
(Refer Slide Time: 00:39)



So, the term machine learning basically refers to learning from empirical data which could be in the form of images, observations, patterns or records. Machine learning aims at performing well on the training data and also at ensuring a good performance on future, unseen data. So, this is a main purpose of any machine learning that whatever paying data we are having, we train that accordingly and we try to predict the future we develop an algorithm to try to predict the future ok.

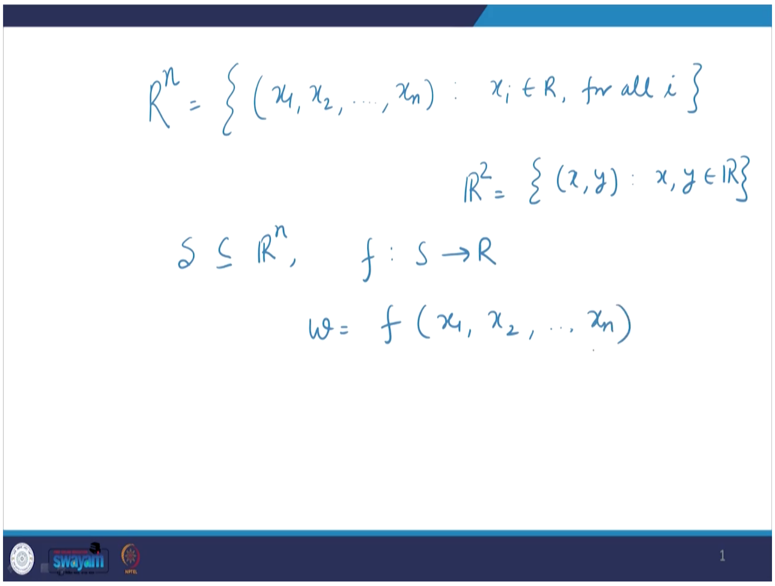
Machine learning deals with minimizing a cost function, also called as objective function which is a scalar function of several variables that typically measures how accurately our model fits the data we have. So, that is the main important theme of optimization here. Calculus and optimization play an important role in understanding the algorithms of machine learning. So, let us first start with calculus that what are the various components of calculus which we are having and which are essential in machine learning.

(Refer Slide Time: 02:06)



So, the first is before understanding, before going to the various other aspects of calculus, let us first understand functions of n variables. So, what do you mean by function of n variables? Let us suppose S a subset of R n.

(Refer Slide Time: 02:14)

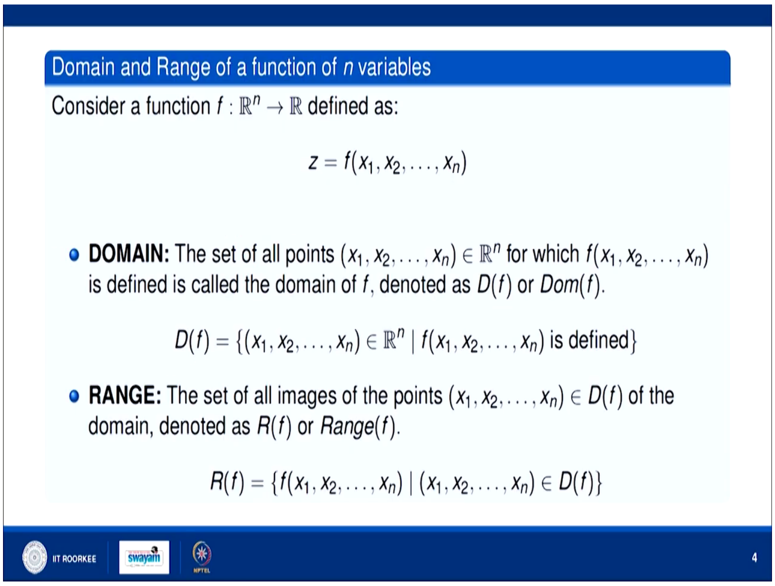


So, what do you mean by R n? So, let us see R n is what you mean by R n? R n is basically if you take R n, R n is basically a set of n tuples x 1, x 2, up to x n such that x i belongs to R, for all i. So, this is basically R n. If you take R 2 suppose, if you take R 2 what is R 2? R 2 is basically all x, y such that x and y belongs to R; that means, two-dimensional geometry if you speak in geometry point of view.

So, R 2 is nothing but two-dimensional geometry, R 3 is basically three-dimensional geometry. So, R n is basically n dimensional Euclid, n-dimensional space where which is x 1, x 2, up to x n; such that all x i belongs to R.

So, if we take a S, a subset of R n and we define the function f from S to R in this way suppose w equal to f of x 1, x 2, up to x n then, this function is called function of several variables or function of n variables. So, here this x i here this x 1, x 2, up to x n are called independent variables and w is called dependent variable ok.

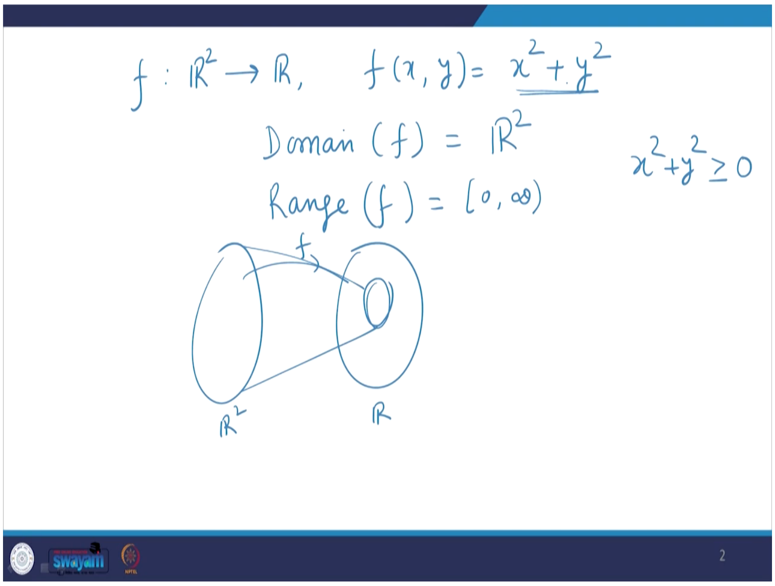
(Refer Slide Time: 03:45)



So, we have already seen a function of single variable. In the same way we are trying to extend the function of single variable to function to n variables ok. Now, how we can define domain and range of function of n variables? So, again consider function f from R n to R which is defined as z is equal to f of x 1, x 2, up to x n.

Now, domain of the function is defined as the set of all the points; set of all the points x 1, x 2, up to x n in R n, for which this function f is defined. So, that is called domain of f and the range of f is simply the set of all images of point x 1, x 2, up to x n of the domain set denoted by R of f range of f.

(Refer Slide Time: 04:42)



So, let us discuss what do you mean by domain or range of a function by giving some examples. Say we take a function f from R 2 to R and the function f is defined like this: f of x, y suppose is equal to x square plus y square a simple example. Suppose function is. So, here we are having two input variables. So, this function is a function of two unknowns and the image is giving that this f maps this to R which is x square plus y square.

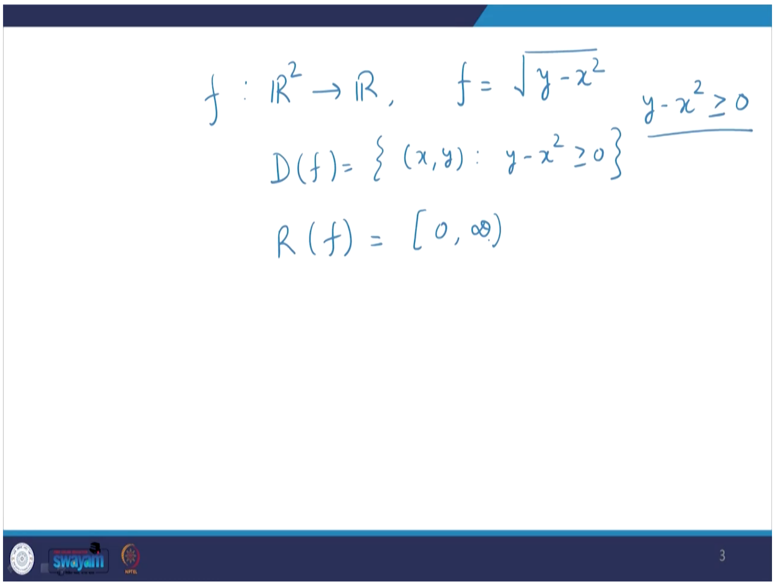
So, what will be the domain of this function? See that domain of this function domain of f will be nothing but entire R 2 because it is always defined for every x and y.

If you talk about range of f; if you talk about range of f, so, range of f; range of f for all those; see you are having this R 2 and here it is R. We are what x, y which we will choose here so that f is defined that is called domain of f and through this map, this f suppose map to this set.

So, this set is basically range of f. So, here what should be the range? See this value x square plus y square can never be negative, x square plus y square for every x and y in R is always greater than equal to 0. It may be 0 when x and y both are 0.

So, we can say that range of this function f will be from 0 to infinity ok. It can never be negative. So, the range, thus domain of this function is entire R 2 and the range of this function is from close interval 0 to open interval infinity.

(Refer Slide Time: 06:46)



So, in a similar way, suppose take another example. Let f is from R 2 to R and say f is defined by this; f is defined as under root of say y minus x square.

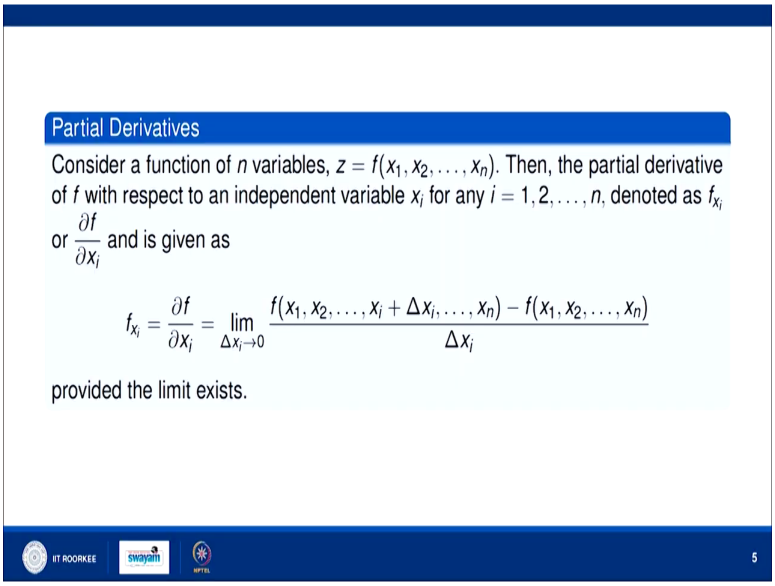
So, again for this f, there are two input variables x and y and it maps that those two points in this way by the mapping under root y minus x square. So, what should the domain of f?

Domain of f is all those point, where this f is defined and this f is defined whenever y minus x square is greater than equal to 0, otherwise this f is not defined in real lines. So, we can say that domain of this f will be all x, y; such that y minus x square is greater than equal to 0. So, the collection of all those x, y, where y is bigger than or equal to x square will constitute the domain of this function f.

Now, if you talk about range of this function f. So, range of this function f will be now whatever x, y you choose here, this will map to this will map x, y to in this way and under root can never be negative.

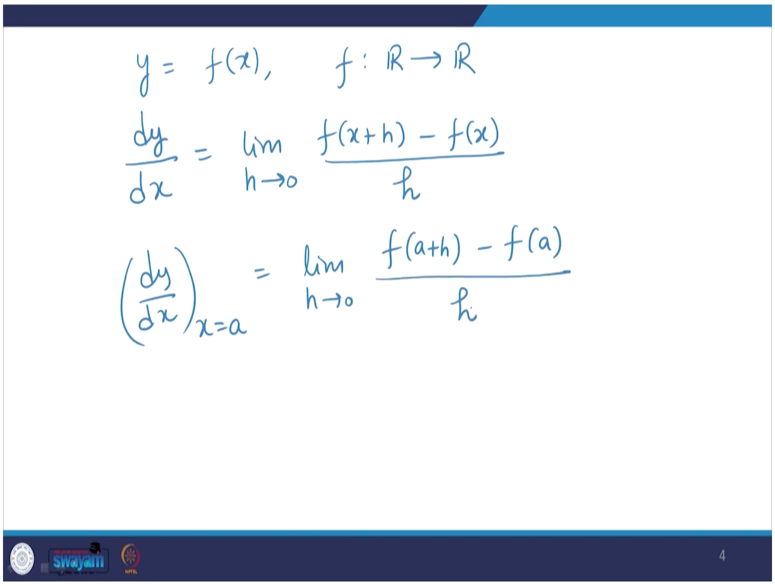
So, we so, this may be 0 when x and y both are 0 or y equal to x square, if y equal to x square it would be 0. So, we can say that range of this function will be again for this example 0 to infinity, close interval 0 up to infinity. So, in this way, we can defined domain and range of several variable functions ok.

(Refer Slide Time: 08:34)



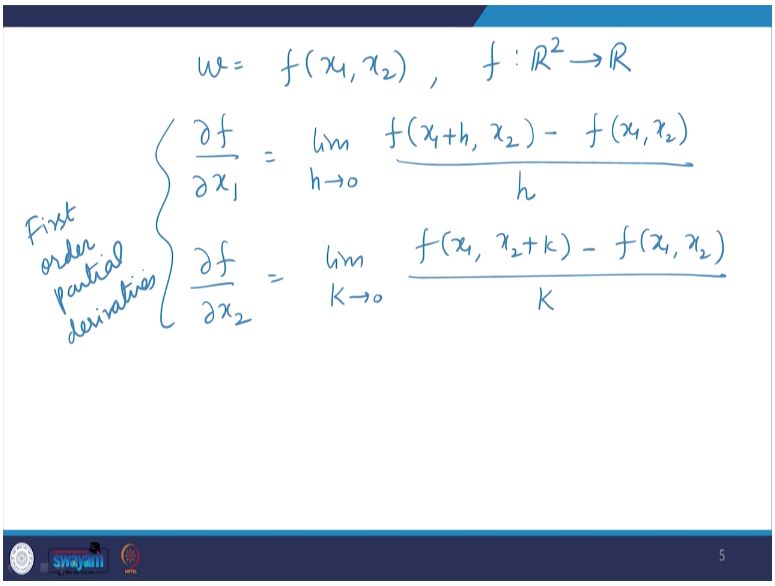
Now, come to partial derivatives. How can you define partial derivatives? So, here is a here consider a function of n variables which is given by z equal to f x 1, x 2, up to x n and then the partial derivative of f with respect to x i is defined like this. So, let us understand what do you mean by this.

(Refer Slide Time: 08:55)



See we all know what is function of single variable, if function is f is from R to R. If f is from R to R, we all know that we all know that we define dy by dx which is nothing but limit h tending to 0 f of x plus h minus f x upon h. So, this is provided this limit exist. So, this is called derivative of the function f respect to x. In this way we defined f dash x at x. If we define dy by dx at x equal to a, then at x equal to a in the same way we can defined like this limit h tending to 0 f of a plus h minus f a upon h.

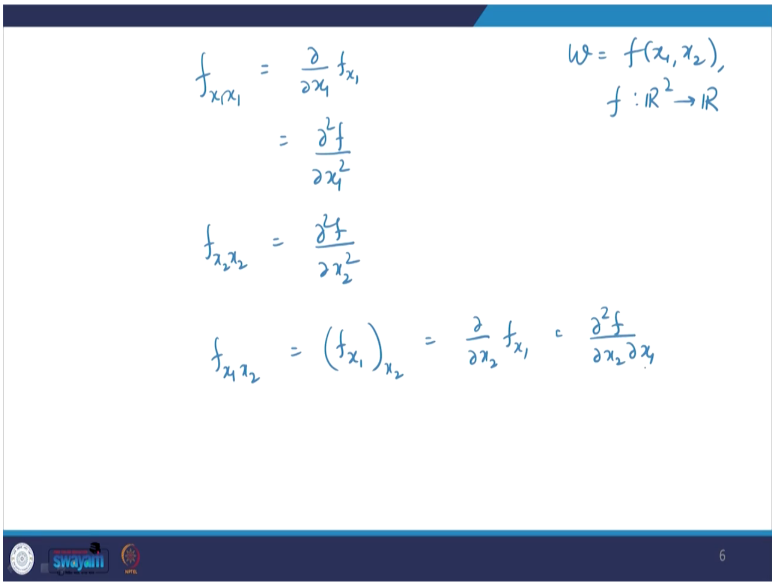
(Refer Slide Time: 09:55)



Now, let us come to function of two variables. Say w is equal to f of x 1, x 2, where f is from R 2 to R a two-variable function. Now here we are having two independent variables x 1 and x 2 and w is a dependent variable. So, here we have a concept of partial derivatives. So, if we define del f upon del x 1. So, del f upon del x 1 is nothing but, limit h tending to 0 it is f of x 1 plus h keeping x 2 fixed minus f of x 1, x 2 upon h.

If we are dealing with partial derivative of f with respect to x 1, then this h will be only with x 1 and the other part will be x 1, x 2. Similarly, if we defined del f upon del x 2, then it may be defined as limit k tending to 0 f of x 1, x 2 plus k minus f of x 1, x 2 upon k. So, in this way we can define first order partial derivatives. So, these are called basically first order partial derivatives; first order partial derivatives with respect to x 1 and x 2 provided this limit exist ok.

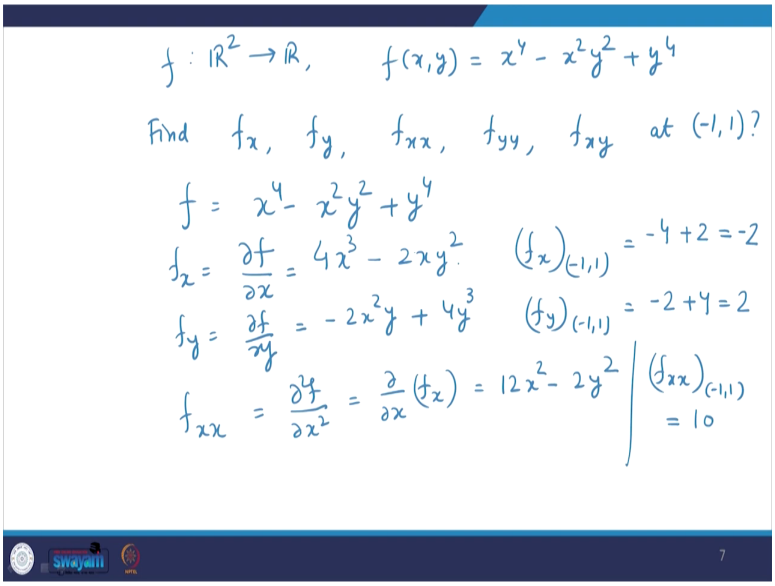
(Refer Slide Time: 11:37)



In the same way, if the higher order derivatives suppose f x 1 x 1; if you have a function w of two-variables x 1 and x 2, where f is a function from R 2 to R, then the higher order derivative, second order derivative may be defined as del of del x 1 of f x 1 which is nothing but second order derivative of f respect to x 1.

If we want to define del x 2 x 2 that means, second order derivative of f with respect to x 2. If we want to define del x 1 x 2, so, this means, f of x 1 with respect to x 2 and that means, del by del x 2 of f x 1 and that simply means second derivative of f respect to x 2 x 1.

(Refer Slide Time: 12:49)



And similarly, other higher order derivatives, partial derivatives can be defined. So, let us understand this by a simple example, simple illustrations. So, let f is function from say R 2 to R and let f of x, y is given as let us suppose it is given by x 4 minus x square y square plus y raised to the power 4 and we have to find f x, f y, suppose f xx, f yy, f xy at minus 1 comma 1.

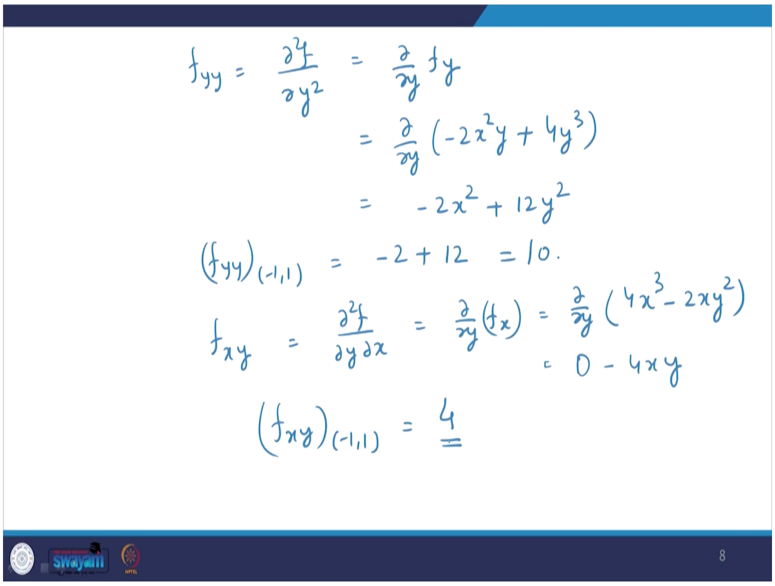
Suppose this we want to find out. So, how we can find? This is very easy problem. So, let us discuss. So, f is what? f is x raise to power 4 minus x square y square plus y raised to power 4. So, f x is what? First order partial derivative of f respect to x. So, if you want to differentiate partially f respect to x that means, we have to take treat x as a variable and all the variables as constant.

So that means, it is 4x cube plus 2 it is minus sorry minus will come here because it is negative; so, this is negative of 2xy square. So, f x at minus 1 comma 1 will be you substitute x as minus 1 and y as 1. So, this will be minus 4, this will be minus plus 2 so, the value is minus 2.

Now, if you want to compute f y it is what? Del f upon del y, which is; now if you want to partially differentiate f respect to y, so, you will take y as a variable and all other variables as constant. So, you take this will be 0, now this will be derivative of y square with respect to y will be 2y. So, that is minus 2x square y plus 4y cube and f y at minus 1 comma 1 will be nothing but it will be we have to put x equal to minus 1 and y as 1, so, it is minus 2 and it is again plus 4. So, it is 2.

Now, suppose you want to compute f xx. So, what will be f xx? It will be second derivative of f with respect to x. So, it is nothing but del by del x of f x. f x we have already computed. So, again differentiate partially respect to x so; that means, 12x square minus 2y square. Now put x equal to minus 1, y is equal to 1. So, the value of f xx at minus 1, 1 will be equal to; will be equal to 10.

(Refer Slide Time: 15:46)



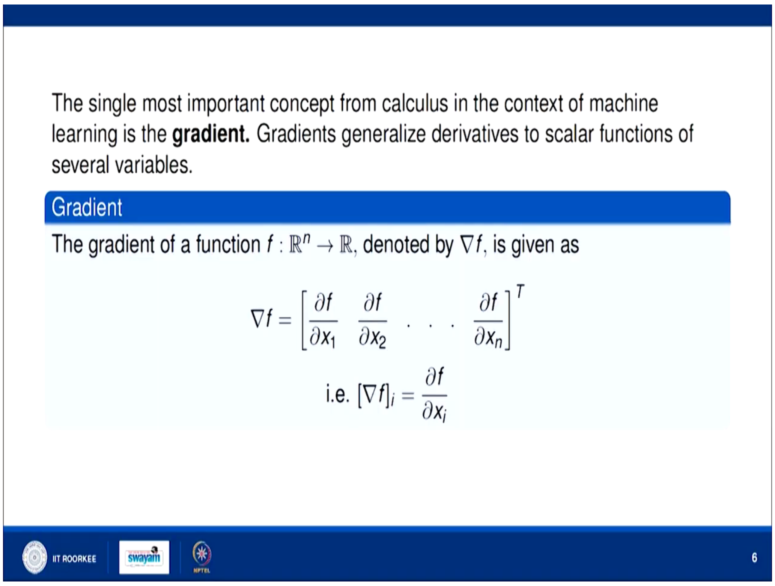
Now, suppose you want to compute f yy now, f we have already computed. So, what is f yy? It will be del square f upon del y square which is del by del y of f y. So, f y we have already computed. So, f y is basically minus 2 x square y plus 4y cube. So, that will be del by del y of minus 2x square y plus 4y cube. So, this will be equal to minus 2x square plus 12y square. Now, at minus 1 comma 1 it will be minus 2 plus 12 which is again 10 ok.

Now the last is f xy, f x y means, del square f upon del y del x; that means, del by del y of f x. So, this is del by del y of now what is f x? f x is 4y cube minus 2xy square, so, it is 4y cube minus 2xy square.

Student: Sir x cube.

So, it is yeah it is 4x cube; it is 4x cube minus 2xy square. So, this is 12, so, this is 0 minus it is 4xy and the value of this at minus 1 comma 1 will be nothing but 4. So, in this way simple mathematical calculations, we can compute the first order partial derivatives, second order partial derivatives or higher order partial derivatives ok.

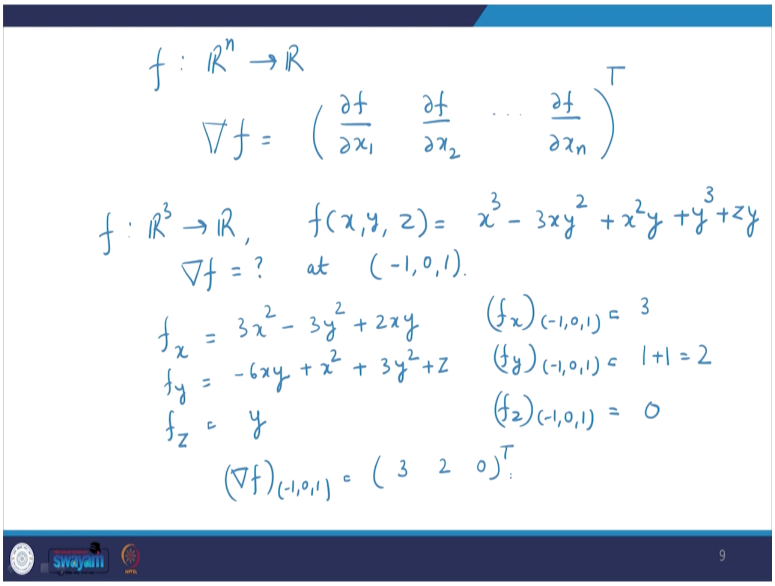
(Refer Slide Time: 17:32)



Now, why we are studying this? So, because these concepts are important machine learning, whenever you deal, whenever develop an algorithm ok. So, in dealing in developing those algorithms, you need the knowledge of basic calculus.

So, gradient concept is very important in machine learning and what do you mean by gradient? How we define gradient? Gradient is basically of a function f from R n to R which is a several variable function is denoted by del of f and is defined is given as by this vector that is del f upon del x 1, del f upon del x 2 of this. So, let us understand what do you mean by this.

(Refer Slide Time: 18:17)



So, here you are having a function f from R n to R ok. Function is of n variable; n variable are the input variables and function maps to R. Now, del of f is nothing but is a vector of this type del f upon del x 1; first you differentiate this respect to x 1, then this with respect to x 2 and so on respect to x n and we take it as a column vector.

Suppose, you are having a function say f of from R2 to R 3 to R and it is defined like this. Suppose, f x, y, z is defined like this, suppose it is x cube minus 3xy square plus x square y plus y cube suppose plus zy. So, suppose this function you are having and you want to compute gradient of f; you want to compute gradient of f; gradient of f is to compute at say minus 1, 0, 1.

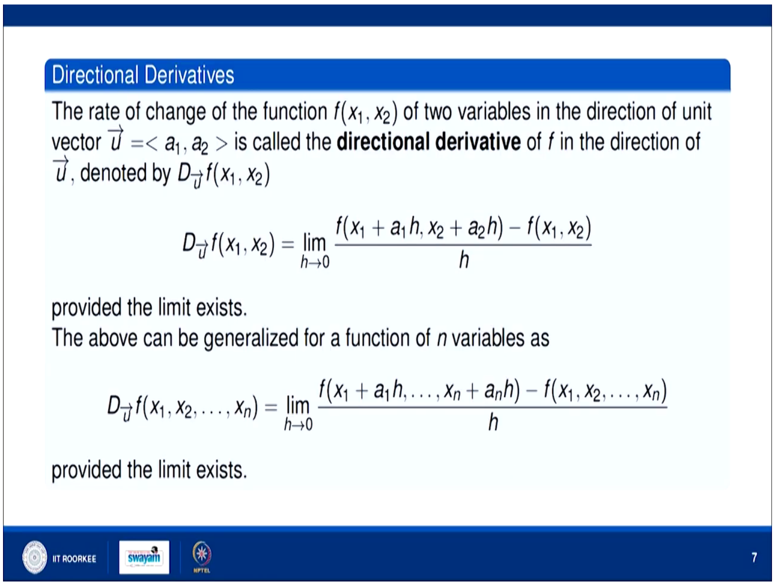
So, you want to compute gradient of this f at a point say minus 1, 0, 1. So, how we can proceed? So, simply we have to find out the first order partial derivatives of f respect to x, y and z and we want, we will compute these first order partial derivatives at this point that will give gradient of f at this point. So, let us complete this problem.

So, what will be f x? f x is 3x square minus 3y square plus 2xy. So, f x at minus 1, 0, 1 will be? So, if you take f x as minus 1, 0, 1, so, this will be this is 3, this is 0, this is 0. So, it will be 3 only.

Now, you compute f y. f y will be what? f y will be minus 6xy plus x square plus 3y square and f y at this point minus 1, 0, 1 will be so, y is 0 only x is here so, it should be 1 yeah ok. So, it will be plus z also and z is 1. So, it is 1 plus 1 that is 2 and f z will be what? f z will be only y. So, f z at minus 1, 0, 1 will be this will be 0.

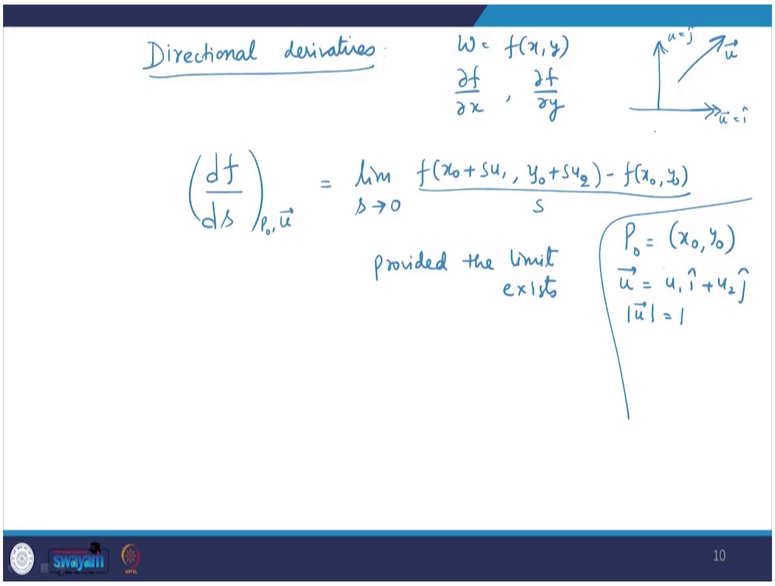
So, now what will be gradient of f? So, gradient of f at this point minus 1, 0, 1 will be this vector this 3, 2, 0 this transpose. So, this will be the gradient of f. So, in this way, we can find out gradient of f. But this small term gradient of f plays an important role in various aspects. So, let us understand this one by one.

(Refer Slide Time: 21:34)



Now, first come to directional derivative, then we will see more geometrical interpretations of gradient of f.

(Refer Slide Time: 21:48)



So, what do you mean by directional derivative of f? So, let us discuss that again. So, the thing is directional derivative. See if we take a function of two-variables x and y. So, we normally compute del f upon del x or we compute del f upon del y.

So, what is del f upon del x means? Del f upon del x geometrically means that rate of change of f along x axis. You are having x axis and y axis. So, what is the rate of change of f along this axis that will be computed by del f upon del x. If you are interested to find out what is the rate of change of f along y axis. So, that can be computed by del f upon del y that is the rate of change of f along y axis.

We normally deals only two directions either x axis or y axis, but if we ask that what will be the rate of change of f along any direction u. So, how can we compute the rate of change of f along any direction u? So, that comes from directional derivatives. So, how it comes from directional directive let us see.

So, the rate of change of f along any direction; along any direction this is a parameter s along any direction u at a point P naught is given by suppose P naught is this point suppose P naught is x naught, y naught. Here I am taking a function of two variables, in the same way I can extend the concept to function of n variables also.

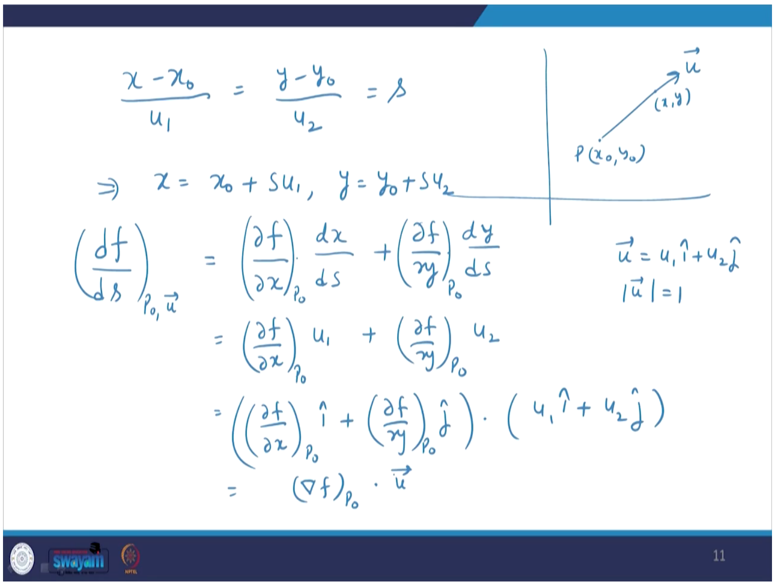
So, if I am taking a fixed point P naught as x naught, y naught and u is a direction, u is given as u 1 i cap plus u 2 j cap, where u is a unit vector that means, mod of u is 1 ok. So, the rate of change of f along u at a point P naught is mathematically defined as limit h tending to 0; a limit s tending to 0 sorry limit s tending to 0 f of x naught plus s u 1 y naught plus s u 2 minus f of x naught, y naught upon s. So, mathematically it can be defined like this. So, provided this limit exist; provided the limit exist.

Student: Y naught (Refer Time: 24:37).

So, this is y naught. So, this is the definition of finding directional derivative of f along any direction u at a point P naught. See if you take considering x axis so that means, in the x axis u will be what? i cap only ok. i cap means, u 1 is 1 and u 2 is 0. So, if you put u 1 equal to 1 and u 2 equal to 0 here so, this definition is nothing but del f upon del x and if you are taking u along y axis, so, in y axis u is nothing but j cap.

So, if it is j cap that means, u 1 is 0 and u 2 is 1. So, if u 1 is 0 and u 2 is 1 you substitute here so, that is nothing but del f upon del y. So, in that sense we can say that this definition generalizes the definition of del f upon del x and del f upon del y also. So, the first way of finding a direction derivative along u in the at a point P naught is this limit definition.

(Refer Slide Time: 25:59)



Now, if the function is if the first order partial derivative function exist at a point P naught, then this definition can be further extended. So, how we can extend this let us see. So, you are having a fixed point x naught, y naught which I am calling as P. This is a direction u which is given by; which is given by u 1 i cap plus u 2 j cap, it is a unit vector, mod of u is 1, I am assuming, along u which is a unit vector.

Suppose now x and y are any point on this vector in this direction; x, y is any point ok. So, how can we how can you find the how can you find the equation of this line in the direction u? So, the equation of line can be find by x minus x naught upon u 1 is equal to y minus y naught upon u 2 because x minus x naught and y minus y naught are dr's of this vector and u 1, u 2 are also dr's direction ratios of this vector and both are parallel. So, the ratios will be same.

So, suppose it is equal to s; so, what this implies? It implies x is equal to x naught plus s u 1 and y will be equal to y naught plus s u 2 ok. Now, you want to compute derivative of direction derivative of f at P naught in the direction u, so, that will be equal to.

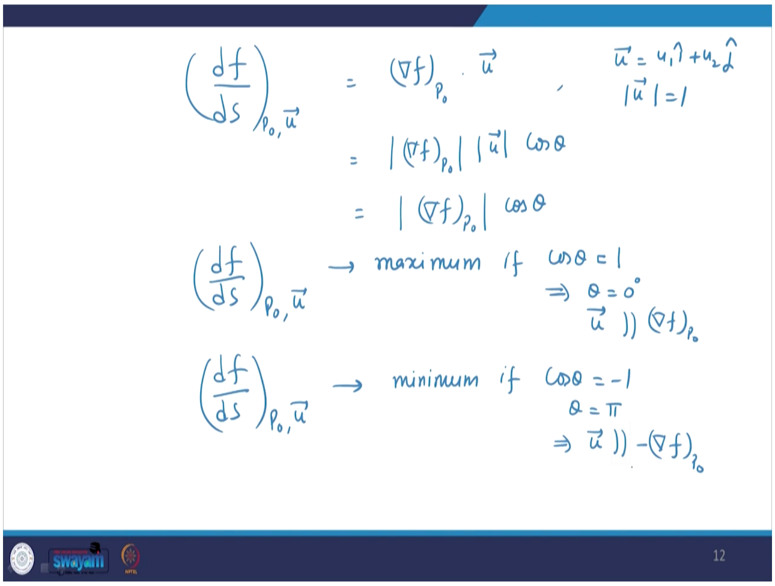
Now this f you have you want to differentiate f with respect to s. So, this f is a function of x and y and where x and y are the functions of s. So, here this function is function of only one-variable because see this is see this x naught is fixed, u 1 is fixed direction is fixed. So, this x is a function of only one-parameter s. So, instead of partial derivative, here will be a total derivative plus del f upon del y into del y again it is dy because again it will be a single variable function. So, this will be nothing but dy upon ds.

Now, what is del f upon del x? You leave it as it is. Now, this at a point P naught of course, and this at a point P naught ok. So, this is del f upon del x at P naught and what is dx upon ds from this equation is u 1 plus del f upon del y at P naught and what is del y upon del s from here is u 2.

So, this is nothing but del f upon del x at P naught i cap plus del y upon del upon del y P naught at it is j cap dot with u 1 i cap plus u 2 j cap it is nothing but the dot product of these two vectors and what it is? It is nothing but gradient of f at P naught and what is this? This is nothing but u vector. So, this is u vector.

So, we can say that if first order partial derivatives of function exist, then the directional derivative of f at a point P naught in the direction u is nothing but the dot product of gradient of f at that point with the unit vector along that direction. The direction in which you want to find the directional derivative of f.

(Refer Slide Time: 29:52)



Now, let us try to understand the geometrical interpretation of this. See we have seen that the directional derivative of f at P naught in the direction u is nothing but a dot product of gradient of f at P naught with u vector, where u is nothing but again it is u 1 i cap plus u 2 j cap which is a unit vector mod of u is 1.

Now, a dot b is what? a dot b is mod of a mod of b into cosine of angle between them. So, this is nothing but mod of this vector mod of u and cosine of angle between them this theta is angle between gradient of f at P naught and vector u. Now, mod of u is 1 from here. So, this is nothing but del of f at P naught into cosine of angle between them ok.

Now, if you want to see that which is a direction of f where the rate of change of f increases most rapidly or decreases most rapidly so, that can be seen from this expression. How? See this expression will be maximum when? This expression will maximum if cos theta this is a fixed quantity at P naught is a fixed quantity, it will not change, but this will change the theta will change.

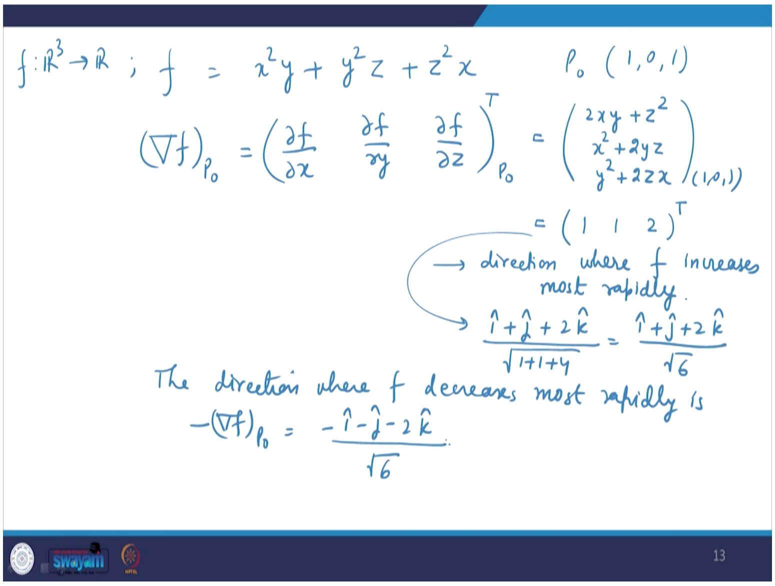
So, in which direction because we do not know which u is, we want to know that which u which is the direction u at which the rate of change of f increases most rapidly; so, that in order to find that direction, this cos theta must be maximum.

So, del df upon ds at P naught in the direction u is maximum if cos theta is maximum and cos theta maximum means 1 because cos theta lies between minus 1 to plus 1 the maximum range of cos theta is 1. If cos theta is 1, this means theta equal to 0. Theta 0 means what? Theta 0 means, the direction of u is same as the direction of gradient of f of P naught. So that means, u is parallel to direction of f at P naught that means, the rate at which the f increases most rapidly is along the direction of gradient of f. So, if we compute gradient of f, it gives what? It gives the rate at which the rate of change of function increases most rapidly and we sometimes called is steepest direction.

So, similarly, if you want to see df upon ds at P naught in the direction u is minimum. When it is minimum? It is minimum if cos theta is minus 1 that means, theta equal to pi. So, this pi means the direction of u is nothing but is negative of gradient of f at that point. So, if direction of u is see if it is opposite to a direction of gradient then that means, that is the direction where it decreases most rapidly. It is a decent direction; this is the ascent direction it increases most rapidly, and this is a decent direction where it decreases most rapidly.

So, if a function is given to us may be of two-variable or three-variable function and a point is given to you and if you want to find out the direction it where it increases most rapidly or decreases most rapidly, that will be simply given by gradient of f at that point or negative of gradient of f at that point.

(Refer Slide Time: 34:05)



So, let us discuss it by another example. Suppose, you are having this function, say you are having this function f x square y plus y square x plus z square x square z suppose x square z and z square x. Suppose P naught is you can take any point say P naught is 1, 0, 1.

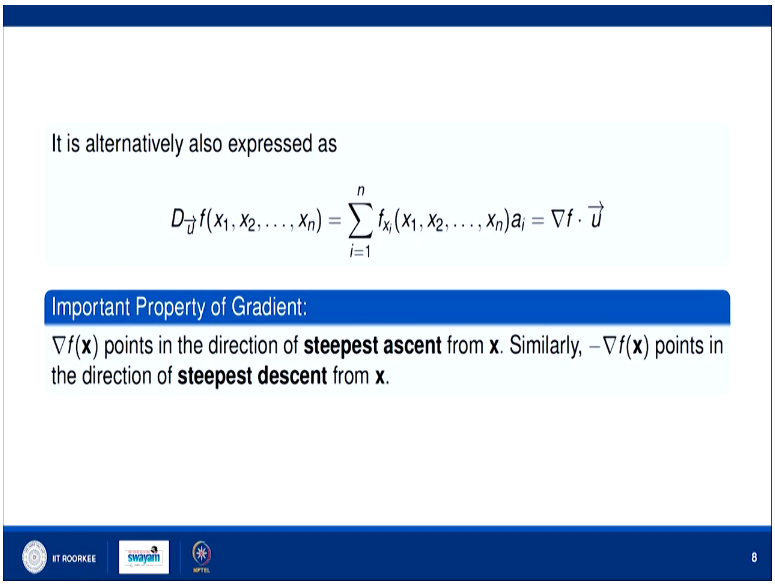
Now the question is for this function which is this function is again this function is from R 3 to R of course there are three input variables x, y and z. So, this function; this function is from R 3 to R ok. Now, if we want to find out the rate at the direction where the rate of f increases most rapidly so, that will be gradient of f at that point at P naught.

So, what is the gradient of f? It is del f upon del x del f upon del y del f upon del z because there are three-variables here at P naught. So, this will be further equal to what is del f upon del x? It is 2xy plus z square del f upon del y is x square plus 2 x 2yz del f upon del z is y square plus 2zx at a point 1, 0, 1. So, this is nothing but x is 1, y is 0 so, it is 0 so, this is 1, x is 1, y is 0, z is 1 this is 1 this 2. So, this transpose.

So, this is this direction; this direction is the direction, where f increases most rapidly. So, this direction is basically; this direction is basically if you see this direction, this is i cap plus j cap plus 2 k cap upon under root of 1 plus 1 plus 4 that is i cap plus j cap plus 2 k cap upon root 6. So, this is a direction where f increases most rapidly.

If somebody asks you that how we can find out a direction where f decreases most rapidly, so, that is nothing but negative of this, negative of gradient. So, the direction where f decreases most rapidly is simply negative of gradient of f at P naught which is equals to minus i cap minus j cap minus 2 k cap upon under root 6. So, in this way, we can find out the rate of change of f along any direction and ok. So, this is direction derivative. So, here I have defined for two-variables and then for n-variables the rate of change of f.

(Refer Slide Time: 37:27)



So, this I have already explained you that gradient of f attacks points in the direction of steepest ascent of x where increases most rapidly that is steepest ascent and where decreases most rapidly that is called steepest descent and that is nothing but negative of gradient of f ok. So, thank you very much for hearing me.

Thank you.