Artificial Intelligence for Economics

Prof. Adway Mitra

Artificial Intelligence

Indian Institute of Technology Kharagpur

Week – 02

Lecture - 06

Lecture 06 : Unconstrained Optimization

Hello everyone. Welcome to this course on Artificial Intelligence for Economics. I am Adway Mitra, an Assistant Professor at Indian Institute of Technology, Kharagpur. And today we are with starting with our lecture 6, the topic of which is Unconstrained Optimization. So, so far we have been seeing some of the basic applications of artificial intelligence in economics. We have seen some example or examples of many tasks in economics in which artificial intelligence can be applied.

From now on we will be discussing more about the methods about the artificial intelligence themselves. and at the same time we will try to link every each of those methods to the domain of economics. So, today's the methods which we are going to discuss today is the method of optimization. Now, so today we will first learn to formulate an optimization problem, we will learn the concepts of convex and non convex optimization.

We will understand that what is unconstrained optimization and how it can be solved using the algorithm called gradient descent and we will also understand what is multi criteria optimization or multi objective optimization and Pareto optimality. So,like optimization problem let us start with an optimization problem in economics a very simple problem. Let us say I have a budget of M there are two sector let let us say I am the I am the government let us say that I like I have some budget which I want to allocate to to different sectors. As of now let us say that there are only two sectors available in front of me let us say education and health care. So, I the amount of budget I have I have to invest part of it to education part of it to health care let us call them as $S1$ and $S2$.

Now, if I let us also assume that if I invest something in any particular sector, I will get some benefit out of the sector. Now, all benefits in economic supports cannot be quantified like that like for example, that is how do I quantify what is the outcome of the healthcare sector or the education sector or let us say defense sector and things like that. So, like there are of course, some indicators economy which are which can actually act as

proxies to measure the performance of every sector. Let us say what is the average or what is the number of people who are getting college degrees per year. So, that is that may be considered as one performance measure of the of the education sector in case of health care.

let us say the overall life expectancy is one indicator of the performance of the health sector. Like similarly in case of like defense or the law and order the number of violent crimes or terrorist attacks that are taking place that is another performance measure for the defense or the law and order sectors and so on and so forth. So, let us say that each of the sectors which we are considering they have some kind of functions using which like which tell us something about the performance of those sectors. And it is like let us say that the we make an assumption that the performance of that sector is a function of the amount of investment I make on it. So, like let those two functions be called as $f1(x)$ and $f2(x)$.

Now, we make a further assumption here that $f1(x)$ and $f2(x)$ these are two function they both map to the same domain. So, that we are and that domain is the domain of real numbers. Now, because both of them are mapping outcomes are measured in terms of real numbers, I can actually consider adding them up. That is if I want to understand what is the net outcome of my investment of these two sectors, then from each of the sectors I can have their individual performance measures $f1(x)$ and $f2(x)$. And because I have made an assumption that they like both of them lie in the say real space, so I can add them up also.

So, that like I define what is known as my utility function. So, let us say $f1(x)f2(M-x)$. So, the total budget M I have divided as x in sector 1 and $M-x$ in sector 2. So, the net outcome which I am getting I call that as utility function which is $f1(x)+f2(M-x)$. So, what is my task? My task is of course, to maximize the utility with respect to x.

So, we are considering that $f1$ and $f2$ both of them high values are good if like if $f1$ that is not necessarily the case. For example, like if the indicator is function for defense sector is let us say number of terrorist attacks. So, clearly I want that to be down. So, a low value of that is good, but in case of say in case of the health care sector if we are talking about life expectancy then higher the better. number of graduates per year the higher the better.

So, in this case we are considering that $f1$ and $f2$ are both higher the better. If that is not the case then also no worries we can simply take the negative or reciprocal of that. So, what is my task? My task is to choose the value of x such that $F(x)$ is as y as high as possible. Now, so this is the basic optimization problem. So, we have to discuss how to

algorithms to solve problems like this.

So, lot of it depends on what the nature of these functions $f1$ and $f2$. So, let us consider with the simplest case that both of them are smooth and differentiable function and this x is continuous. Of course, in the way we have formulated this x has got to be continuous. But let us say let us say that actually need not be because if it is just money then we know that a money is only discrete, but we can let us just make an assumption that the money is continuous that is I can spend any real or fractional amount of money in one of the companies and the rest in the other. So, in that case our assumption is that x is a continuous variable.

So, in if that is the case then we can simply calculate the derivative of this utility function $F(x)$ which will of course, all week also be continuous and differentiable and we equate it that to 0 and then solve for the optimal value of x. So, this is basically just school high school level calculus. Now we can progressively make the problem more and more complicated. Let us now consider multivariable case also. So, let us say that like we can like have some additional constraints like I am not it is not necessary for me to spend the entire amount x I mean the entire budget M I should need not spend on both of these sectors.

I can actually spend a smaller amount also. So, in that case my utility function is $F(x1,x2)=f1(x1)+f2(x2)$ as earlier, but with the constraint that $x1+x2 \leq M$. So, note that earlier I was saying that $x1+x2=M$. So, that instead of writing $f2(x2)$ I could simply write $f2(M-x)$, but in this case I like $x1+x2$ I am specifying is less than or equal to M. So, like we have higher degrees of freedom in this case.

In the previous case I had only 1 degrees of freedom meaning as soon as I am changing x or setting a value of x the value of the other investment is also automatically fixed. But in this case both of them can vary, however the only constant is that $x1+x2$ their sum must be within my budget. So, we can try to find solutions of this problem as I said we will find some algorithms. So, like we will find those values of $x1$ and $x2$ for which this utility function is as high as possible. So, that is that will those values will comprise my optimal solution.

So, note that there might be different combinations of $x1$ and $x2$ for which we can reach the same value of the of this utility function f. So, like some of those values they may be the optimal value, some of them may be the suboptimal value and so on and so forth. So, like here you can see that we can like plot the those solutions which are which give the same value of the utility function we can just plot them together as curves and we call these as indifference curves. So, in this case let us say that the two variables are $x1$ along

the x axis and $x2$ along the y axis I have plotted. Now let us say that for a the functions $f1$ $f2$ are such that if both of them are 2 then the value of the utility function f that is equal to 4.

Similarly, if the first one is $x1$ and $x2$ is 4, then also the value of the utility function is 4. On the other hand, if it is 5 3 or 3 5, then in both cases the utility function is equal to 15. So, like we can see that as long as like we are on this curve the exact value of the $x1$ and $x2$ the variables which we are taking does not seem to make a difference as far as the end goal is concerned. So, we can say that these solutions in some sense they are all equivalent to each other. So, like and like I am indifferent to whether I choose 3 5 or 5 3 because both are giving me the same outcome.

Now, so that is for like in so these kinds of curves these are known as the indifference curves. Each indifference curve is associated with one particular value of the of this objective or utility function. Now, how to solve an optimization problem? So we usually start by casting the optimization problem as minimization. Now earlier I was saying maximization of the utility function. Now there are certain types of functions which we would like to maximize and certain other functions which we would like to minimize.

But in general when we are formulating it mathematically we follow this convention that like we treat it as a minimization problem. If the actual problem is maximization then no problem we simply fits and fix flip the signs that is if I want to maximize $F(x)$ that means, I want to minimize minus of $F(x)$. So, I convert it to a minimization problem like that. So, like so if we remember the concept of linear regression or the least square regression. So, like we have the squared error loss function that is we have values of the independent variable x.

and we have the values of the target variable or dependent variable y. So, I just try to express the dependent variable y as a function of independent variable x the function is a linear function and I try to minimize my least square error. So, so this is like so I basically I am trying to find the parameters w and b for which this objective function or the in this case it is called as the square the least square error this function is as low as possibility I want to minimize it. So, I this is an optimization problem with respect to the parameters w and b. So, I just like I can see do it this problem quite easily by calculating the gradients of this loss of this least square loss or least square error and I can equate those derivatives to 0 that is I calculate the derivatives with respect to both w and b.

So, I get like those are partial derivatives each of them are equated to 0. So, that I get two equations I solve those equations simultaneously and I get closed form expressions of W and B. So, that is what we had learnt to do in high school for solving this problem

of linear regression. But in this case or in many cases this analytical approach based on calculus does not work. Reasons for not working it might be that the utility function is not even differentiable or it might be that the derivative equations which we get I mean the two simultaneous equations which we got using the partial derivatives.

it is possible that we cannot solve them like simultaneously. So, in such cases we need the what is known as the numerical approach. Now, what is the numerical approach? So, this is based on like approximation algorithms like these are all iterative algorithms. So, iterative algorithm it starts with an initial solution which is of course, not the correct solution, but then at every step we keep increasing which we keep improving the function one step at a time and try to gradually move towards the optimal solution. So, this is what the gradient descent bed optimization looks like.

You start with any initial point $x0$ and next step you move to a point $x1$ which is equal to $x0$ - a small term a small quantity a usually a positive quantity multiplied by the derivative of that function at x naught at the say at the its current value $x0$. So, that gives us a new value of $x1$ now you like you check whether $x1$ is the same as $x0$ or it is a different from $x0$. If it is a in general it will be different. So, if it is different you just continue and you keep on doing the same process over and over till you see that $x1$ and $x0$ are the same. So, like and when you are doing the moving to the next iteration then you next you forget the previous current value $x0$ and you just remember the new value $x1$.

You consider you set $x0$ to $x1$ and then again you do the same process. So, in this like you just keep on doing this in iteration after iteration until this condition is satisfied. So, like this like this is the like as you can understand a very simple algorithm where you just need to calculate the derivative of x this will work even if the variable x it is a vector valued variable. So, like in many cases in real life we will see that it is a the input variable the dependent variable x it is a vector valued variable. So, we can calculate the derivative of the loss function also it will it will be basically the vector of the partial derivatives with respect to each of the variables in the vector.

And the optimization like for vector x the optimization for each of the dimensions of the vector it can be decoupled also from the rest. So, let us say that the vector x it has d dimensions $(x1, x2, \ldots, xd)$. So, I first like I can minimize $x1$ while holding the other variables $x2$ etc. I am holding their variables as their those variables at fixed at a current value at our then after that I am once I have like perturbed the value of $x1$ then I move to the next vector dimension $x2$. And now I again apply the gradient descent to descent to update its value while holding the other value like the other variables that is $x1$ $x3$ $x4$ ... xd they are constant.

After that we update $x3$ holding all other variables as constant and so on and so forth. So, just to see an idea of why it will it should work let us consider a very simple one dimensional function is x^2-4x+4. So, x where x is a real number. So, this is the function which we want to optimize. So, we can like we can easily understand that the solution is of this is $x=2$ that is the point at which it will it will be minimized.

Now, to apply the gradient descent we calculate its derivative with respect to x which is $2x-4$. and we that a parameter a which we call as a learning rate we set it to 0.1. Now, we will see the significance of this learning rate. Now, let us say that our initial estimate of the solution is 5.

Now, of course, that is wrong. So, like at 5 the this function $F(x)$ it is quite far away from its smallest value at $x=2$ its value will be equal to 0, but at 5 it is certainly not 0. In fact, it is closer to 10. But now what we will do is so a 5 is my $x0$ now I will calculate what is $x1$. How I will calculate it? So, I can just like calculate the derivative at my current value 5.

So, 2×5 is 10 - 4 is 6. So, the derivative is 6 I multiply that with the learning rate 0.1. So, that is 0.

6. and then my next x 1 that becomes 5 minus 0.6 which is 4.4. So, I move to the point 4.4 and I find that the value of f has decreased that is $F(4.4)$.

we can understand from this diagram that $F(4.4)$ will definitely be lesser than $F(5)$. So, you can say that my task is to minimize f. So, I have moved a little bit towards that a. Now, I keep on doing the same thing after a few iterations maybe from 4.4 I have come closer to 3 and maybe after still further iterations I have come very close to 0 itself and finally, I have just reached x_{opt} which is $x=2$ at the value at which $F(x)=0$.

So, when that happens when I have reached the optimal then I will see that when I try to change it further then $x1$ like that I find that it is not changing that is. So, in that case I would understand that I have reached the local minima. So, why will it why will that be the case because when I have reached the minima then the derivative will of course, be equal to 0 because we know that it is at saddle points like maxima minima etc. or inflection points the derivative vanishes. So, $x1$ is going to be just $x0$. So, then we have our iterative process has come to an end and we have reached a minima.

So, that is that is our solution. So, does it always converge? So, first of all does it there are two questions like first of all will it necessarily converge and secondly will it

converge at the minima. So, it can be like shown that if the function is such that it has only one minima then definitely it will converge there if at all it converges. it cannot converge at like if it is a like if it is function with a single minima then it cannot converge at any non minima point. However, like the question whether it will at all converge or not that is that really depends on the choice of the learning rate A.

So, so what is the role that the learning rate A plays. So, basically it governs by how far I am moving from the current value to the next values. So, by the way so this explains the term gradient descent. So, I am currently here I just feel that which way is the gradient of the function that is when I am at the current value I can either increase or I can decrease should I increase or should I decrease. So, I that the decision that decision depends on what is the direction in which the function decreases or rather it is slopes the function slopes in which direction. So, obviously, I can understand that if I decrease the value of x here then the function will decrease.

So, I must decrease the value of x. If instead of the initial point between 5 it were something like let us say minus 1 then it would be the reverse then we would see that increasing x would decrease the function. So, we would have moved in the direction of increasing the value of x, but the question is I am increasing it in decreasing it in one particular direction, but by how much. So, that is governed by the learning rate a higher the learning rate the bigger steps I make in the direction of the So, like I can either choose a small value of a or a big value of a both have pros and cons. If I choose a big values of a then there is a risk that I may jump over the optimal solution. Let us say my current with my initial value was here now I let us say I understand that I have to decrease the function because the gradient is the in that direction, but the step which I take is so large that I end up here.

And like here that is I have crossed the minima. So, now, again when I see I see that the gradient is in this direction that is I must increase x. So, I increase x, but by such an amount that I come somewhere here that is I have again overshot the minima. And so, like I keep on just moving from this side to that side, but I am never able to reach the minima. So, that is one way in which I may just fail to converge all together that is I like if I am choosing a large learning rate then the algorithm may not converge at all. On the other hand if I take a small learning rate then it will converge definitely, but it may take a lot of time it may take a lot of steps before I can reach the minimum.

Now if the as I already said that if the function is convex then we will necessarily converge there if it is not convex then we will converge at any minima. In the next slide I will explain what is meant by convex and non-convex functions, but before that I must also add that there is an alternative to this gradient descent which is known as the Newton Raphson's method. So, that is similar only the only however, the difference lies

in this update state. Here I like calculate the that is instead of multiplying the gradient with the learning rate, I multiply the current value of the function with inverse of the gradient. So, like so this has faster convergence and the learning rate parameter which we already discussed is a bit problematic that is gone, but its performance on the other hand it depends on the properties of this of this thing.

So, note that if x is scalar then or one dimensional then calculating this is not a problem at all and this inverse simply means the reciprocal. So, in that case this is Newton's reference method is good and it has no problems. On the other hand if it is a vector if x is vector valued as we discussed earlier then this derivative this is basically the Hessian matrix and I have to calculate the inverse of it and calculating the inverse of that is sometimes a dicey issue. So, like in case of vector valued function this Newton Raphson's may have some computational problems.

Now, I mentioned the term convex function earlier. So, what is the convex function? So, convex function is defined like this that the value of the function at any x which lies let us fix two points $x1$ and $x2$ first on which the function is defined. Now, so like at $x1$ we have the function value $F(x)$ at $x2$ we have $F(x2)$. Now, you consider any point which lies between $x1$ and $x2$ so like this. So, in that case the value of the function at that point $F(x)$ will be less than or equal to the straight line which was joining these two points $F(x1)$ and $F(x2)$. That is like if we assume a linear function which is connecting these two points $x1$ and $x2$.

And convex function f is such that its value will always be less than the linear function between those two points. Now, this is the definition of a convex function it can be shown that a convex function has a unique minima. And all the numerical methods like gradient descent or Newton Raphson's will converge there if they at all converge. We have already seen a situation where it may not converge at all, but if it does converge it will definitely converge at the local minima only it cannot converge anywhere else. But a non-convex function that can have multiple minima that is which we call as the local minima.

So, we are all concept familiar with the concept of local and global minima anyway. Now when we are solving a numerical method, it can converge to any local minima which is closest to the initial point. Where it will, if there are multiple minima or multiple local minima, these numerical methods will converge to any point which is closest to the initial point. So our approach will be to start descent from multiple initial points. Find the local minima corresponding to all of them and simply find the minima of those local minima.

If we do this for like a wide enough variety of initial points, then hopefully the minima of those local minima will be equal to the global minima itself. Now, we come to a different problem this is called as multi objective optimization. So, so far we have been considering that there is one objective function or the utility function which we are trying to minimize, but now let us say that there are multiple objective functions. So, let us consider a simple problem that we problem which many of us face in daily life. So, let us say that I want to fly from city A to city B for which there are different flights of different airlines.

Now, those flights they have they take different times and they have each of them have different costs in thousands. So, this is the table of the time taken and the cost charged by each of these airlines. So, now, obviously, my aim is to minimize the time of flight as well as the cost of the flight. So, which do I choose? So, here from this table we can make certain observations.

Let us look at these two airlines Indigo and Go Air. So, Indigo definitely takes less the time than Go Air 2 versus 3, the cost of the both the flights is the same. So, we can say that like between Indigo and Go Air I will definitely prefer Indigo because the costs are the same and I am gaining in time. but if you consider these two pair Go Air and Vistara between them Go Air costs Go Air is faster than the Vistara it takes less time, but its charge is also more. So, in this case there is no obvious solution in front of me time if I consider time then I should then I should consider Go Air, but if I consider cost then I should consider Now, if you consider this set, if I consider this Air Asia, Air Asia is here and if I consider it with this set of Jet Airways, Akash and Kingfisher.

So, let us say what happens. Air Asia versus Jet Airways both take the same duration, but the cost is least in case of Air Asia. So, I will definitely prefer Air Asia with over Jet Airways, the same holds for Air Asia and Akash. If I consider Kingfisher, then also I am reaching the same conclusion. In case of Kingfisher, in fact, it is the conclusion is even clearer.

If I use Air Asia, I am doing better than Kingfisher on both aspects. So, now this brings us to the notion of dominance in case of multi objective function. So, I say that a solution $x1$ dominates another solution $x2$ assuming that it is a minimization problem if like for every objective function like $fi(x1) \leq fi(x2)$. that is I am considering that smaller is better and there exists at least one objective function for which the solution $x1$ is better than the solution $x2$ right. So, like if you if we come back to the air the airlines example.

So, we can if you consider Indigo versus Go Air let us say. So, like we can say or rather or if we or as we just discussed right now Air Asia and Kingfisher. So, Air Asia

dominates Kingfisher because like Air Asia is as good as is in fact better than Kingfisher on both accounts. Air Asia also dominates Jet Airways because Air Asia is at least as good as Jet Airways on both accounts also there is at least one account on which Air Asia is strictly better than Jet Airways. So, I can say that Air Asia strictly dominates Jet Airways. So, now it is also possible that there are two solutions x1 and x2 which were neither dominates the other.

Like if you consider this case the Air India versus Viztara then Air India is sorry Go Air versus Viztara. Then Goyer is better than Vistara with respect to time, but Vistara is better with respect to cost. So, there is like so we cannot say that either of them dominates the other. That is one is better on one account, the other is better on the other account.

So, there is no dominance. So, now I will say that any solution x that is any choice of airlines which I can make in this case I will call that solution as Pareto optimal if it is not dominated by any other solution that is you cannot find any other solution which is strictly better than this. So, note that here I have discussed like assuming that it is a minimization problem, but it also works like if I am considering a maximization problem or even a mixed problem where one criteria F1 needs to be maximized another criteria $f\,2$ needs to be minimized. So, if you look at this diagram we can say that 0.

1 dominates 0.2 why because the x like with respect to $f\,1$. The point 1 is definitely better than point 2 because $f\,1$ needs to be maximized, but with respect to $f\,2$ also 1 is definitely better than 2 because $f\,2$ needs to be minimized. So, 1 dominates 2. Similarly, it can be shown that 5 dominates 1 because with respect to $f\,1$, 5 is definitely better than 1 with respect to $f\,2$ both are equal. So, like we can say that 5 dominates 1. but if you consider 1 and 4, then there is no dominance because 1 is better than 4 with respect to $f\,2$, but 4 is better than 1 with respect to $f\,1$.

So, a solution such a set of solutions which are not dominated by any other solution is called as a optimal solutions. The non-dominated set of all possible solution that is called as the Pareto optimal set. So, we can say see that if you consider Indigo, SpiceJet, Air India and Air Asia in this table, you will see that like none of them dominate each other. So, they like furthermore none of them is dominated by any other of the airlines also.

So, these 4 they form a Pareto optimal set. In fact, since they are not dominated by anyone else then we say that it is a Pareto optimal set of rank 1. Now, similarly among the things which are left among the remaining ones, so they are so these are dominated by the Pareto optimal sets in this case Indigo, Spice Jet, Air India and Air Asia. Now, if you consider these three cases, now they do not dominate each other plus they dominate whatever is remaining such as Akash and Kingfisher. So, these three they do not

dominate each other, but they dominate the remaining ones. So, in that case we can call them as the dominant set of rank 2 that is if you once you take away the Pareto optimal solution or the dominant set of rank 1, the next best set of solutions that is called as the rank 2.

Similarly, we can define rank 3 also and so on and so forth. Now, we can define like we. So, clearly there are multiple Pareto optimal solutions are possible to every multi objective problem. The boundary defined by the set of all such problems in the objective function space when they are mapped to the space of the objective functions that is known as the Pareto optimal front. So, let us say this is the space of solutions. So, here $x1$ and $x2$ there are two like let us say that it is there are two variables $x1$ and $x2$.

So, from that space at every point of the at every solution I can apply the functions $f1$ and $f2$. So, I can find a new point in the space defined by $f1(x)$ and $f2(x)$ right. So, Like we can say that corresponding to this particular point of $x1$ and $x2$, I can calculate $f1$ and I can calculate $f2$. So the $f1$, $f2$ pair of corresponding to this point is then mapped to this point. Similarly, the $f1$ and $f2$ pair of corresponding to this point is mapped to this point and so on and so forth. So, like it can like it is not very difficult to understand that the points which are lying on the boundary of this region.

So, like this is basically the like that is corresponding to all of these solutions, this is the set of values we obtain in the function space. Now, it is difficult not difficult to understand that the edge of this or the boundary of this region will be defined by these Pareto optimal solutions. So, what the when we are trying to solve a multi objective optimization our essential aim is to come up with a set of Pareto optimal solutions. However, like we like we may want those Pareto optimal solutions to be as diverse as possible that is where each solution is good with respect to a different objective. So, like among these Indigo, Spicejet etc. like they are all Pareto optimal, but I may like I may want diverse like diversity among them.

Let us say that Indigo is very good with respect to time even though it is may not be that good with respect to cost. On the other hand Air Asia may be the exact reverse it may be very good with it may be very cheap, but not so good as far as time is concerned. Both of them are parato optimal solutions, but they are they these two they provide some diversity that is one is providing like very good result with respect to one criteria or one objective another with different objectives. So, like this is the aim of multi objective optimization. How do I solve it? And one way to solve it is by attaching weights to all of the objective functions so that we can like add them up and convert it to the kind of functions which we had earlier.

So, like remember consider that if you remember the original budget constraint problem. So, Now, like we I am trying to optimize both of them, but like I just have some like I make prioritize one over the other. So, these kinds of priorities I may represent by these kinds of weights and then I can add them up as I was doing in to define the utility function in the budget case. So, this solution concept of Pareto optimal solutions in economics it is very important in the domain of resource allocation among multiple sectors that is how to divide a budget among multiple sectors. So, that the return is from at least some of the sectors should be good in better than in terms of others. It is also useful in like in trade where we were trying to divide like or we are trying giving the task of producing different goods to different countries.

Any country should grow or should produce that goods which it is able to do more efficiently and more cheaply than the others and then it can trade. Similarly, in case of income distribution I may transfer the resources from a wealthy individual to a poorer individual in such a way that it the overall social welfare is improved, but the we cannot say that the wealthy individual becomes worse off than they were earlier similarly in case of social welfare maximization. So, with this we come to the end of this lecture 6 on unconstrained optimization.

In the next lecture, we will discuss constrained optimization. So, we will meet again. So, till then please take care and stay well. See you soon. Bye.