

MIT OCW Çok Degiskenli Calculus - Ders 9

Bu dersin konusu birden fazla degisken iceren fonksiyonlari minimizasyonu ile ugrasirken yardimci olacak kismi turev (partial derivative) kavrami. Çok degiskenli bir fonksiyon $f(x, y)$ 'nin birden fazla turevi vardir. Mesela bunlardan bir tanesi

$$\frac{\partial f}{\partial x} = f_x$$

Bu turev x 'in degistirildigi ama y 'nin sabit tutuldugu bir durumu gosterir.

$$\frac{\partial f}{\partial y} = f_y$$

ise y 'in degistirildigi ama x 'nin sabit tutuldugu bir durumu gosterir.

Simdi her ikisinin birden degistirildigi durumda ne olacagini gosteren yaklasiksal (approximate) formulu gorelim. Degisim matematiksel olarak soyle

$$x \sim x + \Delta x$$

$$y \sim y + \Delta y$$

O zaman z icin

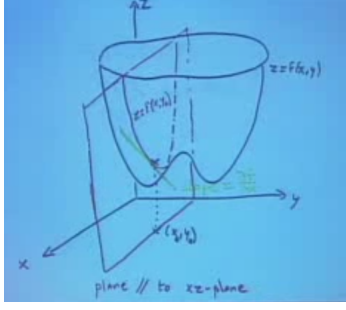
$$z = f(x, y)$$

yaklasiksal degisim soyle olur

$$\Delta z \approx f_x \Delta x + f_y \Delta y \tag{1}$$

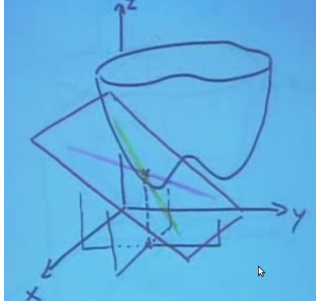
Tekrar vurgulamak gerekirse bu yaklasiksal bir formül, daha “dogru” bir temsil için 2., 3. turevleri iceren daha yuksek dereden (higher order terms) terimlerin de olması gerekir, fakat bu terimler 1. derece lineer bir yaklasiksallik için kullanılmaz.

Bu formulu nasıl dogrularız? Bunu yapmanın yollarından biri teget düzlem yaklasiksallaması (tangent plane approximation). Mesela $z = f(x, y)$ fonksiyonuna olan teget bir düzlemi düşünelim.



Hatırlarsak $\frac{\partial f}{\partial x}$ kısmi turevi x 'in degistigi ama y 'nin sabit tutuldugu bir durumu tarif ediyordu. Yukaridaki grafige gore bu bir anlamda iki cukurlu kap gibi duran z fonksiyonun bir kesitine bakmak gibi (unutmayalım, fonksiyon sadece kabin disinda tanimli, ici bos). Bu kesit uzerine f 'in bir yansimasi oluyor, o yansima ustteki grafikte bir parabol seklinde. Bu parabolda x degistikce o noktanin parabol uzerindeki cizgisel tegeti de degisiyor (grafikteki yesil cizgi) ki bu cizgisel egim $\frac{\partial f}{\partial x}$ 'e esit.

Eger ayni seyi x 'in sabit y 'nin degistigi durum icin yapsaydim, benzer bir kesit elde edecektim.



Bu iki kesit uzerinden elde edilen ikinci teget cizgi birinci ile beraber kullanil-inca bir duzlemi tanımlamak icin kullanilabilir (iki cizgi paralel bir duzlem tanımlamak icin yeterlidir), ki teget duzlem yaklasiksallamasi icin kullanilacak duzlem budur. Formüsel olarak bunu nasıl yapacagimizi gosterelim.

f_x ve f_y iki teget cizgiyi tanımlamak icin kullaniliyorsa, bu formülleri bir araya koyarak duzlemi temsil edebilirim. Eger

$$\frac{\partial f}{\partial x}(x_0, y_0) = a$$

ise bu demektir ki birinci teget çizgi (yesil çizgi) L_1 şöyledir:

$$L_1 = \begin{cases} z = z_0 + a(x - x_0) \\ y = y_0 \end{cases}$$

Bu çizgi için y 'yi sabit tutuyorum, z 'deki değişimi z_0 üstüne eğim a 'nin katları kadar (x 'in değişimi oranında çarparak) ekleyerek hesaplıyorum.

Benzer şekilde

$$\frac{\partial f}{\partial y}(x_0, y_0) = b$$

$$L_2 = \begin{cases} z = z_0 + b(y - y_0) \\ y = y_0 \end{cases}$$

Hem L_1 hem de L_2 $z = f(x, y)$ 'ye tegettir. Bu iki çizgi beraber bir düzlem oluşturur. Bu formül

$$z = z_0 + a(x - x_0) + b(y - y_0) \quad (2)$$

formuludur.

Formül 1, üstteki formülün yaklaşık halidir. Eğer teget düzlem üzerinde olsaydık, \approx işareti = işaretiye dönüşecekti. Bu yaklaşıksallık ufak Δx ve ufak Δy için geçerli. Yani yaklaşık formül, f 'nin grafiği teget düzleme yakın diyor.

Maksimum Minimum Problemleri

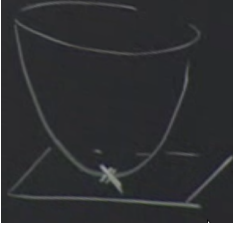
Kismi türevlerin kullanım alanlarından biri optimizasyon problemleridir. Mesela çok değişkenli bir fonksiyonun maksimumunu bulmak gibi. Eğer fonksiyon tek değişkenli olsaydı, hemen türevini alıp sonucu sifıra eşitleyebilirdik, ve buna göre bir çözüm arardık. Çok değişkenli fonksiyonlarda kısmi türevler kullanmak lazım.

Bu derste iki değişkenli duruma bakacağız fakat aynı prensipler, 10, 15, milyon tane değişken için aynı.

Lokal bir minimum için hem $f_x = 0$ hem $f_y = 0$ olmalıdır. Bu niye doğudur? Yine formül 1'e bakarsak, hem $f_x = 0$ hem $f_y = 0$ olduğu zaman Δz sıfır olacaktır, yani birinci derecede düşünürsek $f(x, y)$ 'de değişim yok demektir.

Teget düzlemlerin dilinden konuşursak, minimum anında teget düzlem tama-

men yatay olacaktır.



Formul 2 baglaminda dusunursek, bu durum $a = 0$ ve $b = 0$ oldugu ana tekabul ediyor ve o anda duzlemi tanımlayan $z = z_0$ formuludur.

Tanim

Eger $f_x(x_0, y_0) = 0$ ve $f_y(x_0, y_0) = 0$ ise o zaman x_0, y_0 f 'in kritik noktasidir. Not: Birden fazla degisken icin tabii ki tum kısmi turevlerin o noktada sifir olması gerekir.

Ornek

$$f(x, y) = x^2 - 2xy + 3y^2 + 2x - 2y$$

Bakalım bunu minimize ya da maksimize edebilecek miyiz?

$$f_x = 2x - 2y + 2 = 0$$

$$f_y = -2x + 6y - 2 = 0$$

Ustteki iki denklemi ayni anda cozmeliyiz.

Bu tur durumlarda iki denklemi birbiriyle toplayip basitlestirmeye calismak iyi bir yontemdir. Fakat unutmayin, elimizde her zaman iki tane denklem olmalı, iki denklemi ortadan kaldırıp birdenbire tek denklem ile yola devam edemeyiz.

Toplami yaparsak

$$4y = 0$$

elde ederiz. Bunu alıp birinci denkleme sokalım, sonuc

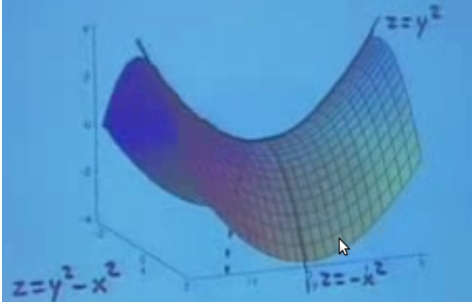
$$2x + 2 = 0$$

$$x = -1$$

Demek ki kritik nokta $(x, y) = (-1, 0)$.

Peki bu kritik noktanin minimum mu maksimum mu oldugunu nereden bilecegiz? Eger tek degiskenli bir fonksiyona bakiyor olsaydik, ikinci tureve bakabilirdik. Benzer bir seyi burada da yapabilirdik, ama sadece birinci turevden bile elimizde iki tane var, ikinci turevlerden cok daha fazlasi olacak. O duruma bakacagiz, simdilik daha az otomatik olarak isi nasil anlayacagimizla ilgilenelim.

Elimizde birden fazla minimum olabilir. Turev(lerin) sifir oldugu noktada bir duzluk vardir, bu bir lokal minimumdur. Yani o noktaya yakin oldugumuz surece (ki lokalligin tanimi bu) bu minimum gecerlidir. Baska bir noktada, turev(lerin) yine sifir oldugu ama daha asagi noktada bir minimum daha olabilirdi. Maksimumlar icin ayni durum gecerli.



Yanliz bir diger secenek daha var. Bu secenek kritik noktanin ne maksimum, ne minimum oldugu durumdur. Bu durumda kritik noktadan hangi “yone dogru” bakiyorsak, degisik bir cevap elde ederiz. Bu at egeri gibi gozuken grafigin orta noktasina, 0,0,0 noktasina bakalim, burada teget duzlem tam yatay. Bu noktaya eger noktasi (saddle point) deniyor. Eger $z = y^2$ yonune dogru bakarsak min durumdayiz, eger $z = -x^2$ yonune dogru bakarsak maks durumdayiz.

2. turevlerden bahsetmisik, ve bu derste kritik noktanin ne oldugunu daha az otomatik bulacagimizi soyledik (2. turevler bir dahaki derste).

Bu yontemde kareler kullanacagiz. Niye kareler? Cunku karesel ifadeler en az sifir olabilirler – bir deger ne olursa olsun, eksi bile olsa karesi alinirsa arti olur, ve bu tur ifadeler sadece sifirda “en az” olurlar.

O zaman $f(x, y)$ 'i karelerin toplami olarak tekrar temsile ugrasalim. $f(x, y)$ 'de zaten kareler var ama tum formulu bir seylerin karesi olarak gostere-

bilirsek, hedefimize erisebiliriz. Tek problem xy terimi, ama $x^2 - 2xy$.. diye giden bir baska formül biliyoruz, Kareyi Tamamlama ile onu kullanalım.

$$f(x, y) = (x - y)^2 + 2y^2 + 2x - 2y$$

Basitlesti ama biraz daha basitlesebilir. Acaba $(x - y)^2$ icindeki $(x - y)$ ile disaridaki $2x - 2y$ arasindaki bir baglanti kurabilir miyiz? Iceriye bir $+1$ eklersek bu olabilir, o zaman disaridaki $2x - 2y$ iptal olur. Icerideki 1 'i dengelemek icin ise disari bir -1 ekleriz.

$$= ((x - y) + 1)^2 + 2y^2 - 1$$

Iste, tum formül artik karelerden olusuyor. Bu formül eger

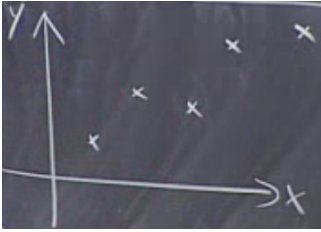
$$= \underbrace{((x - y) + 1)^2}_{\geq 0} + \underbrace{2y^2}_{\geq 0} - 1$$

ise ancak ≥ -1 olabilir. Ve kritik nokta $(-1, 0)$ 'da f 'in degeri hakikaten -1 'dir. Ustteki iki terimin niye ≥ 0 oldugundan bahsettik. Demek ki bu nokta bir minimum. Yani biraz cebirsel takla, ve ufak bir numarayla istedigimiz sonuca erismis olduk.

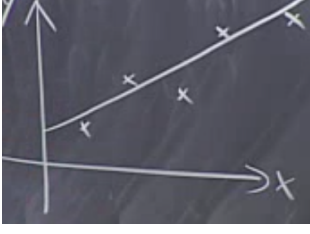
Simdi min/maks probleminin ilginç bir uygulamasini gorelim. Bu uygulamayi min/maks kategorisinde gormeyebilirsiniz, ama aslinda problem min/-maks ile çok güzel bir şekilde cozuluyor.

Deneyisel bilimlerde en az karelesel interpolasyon (least squares interpolation) adli bir teknik kullanilir. Mesela bir deney yapariz, ve deneyden gelen verileri aliriz. Mesela kurbagalari inceliyoruz, ve kurbağa bacak uzunlugunu kurbağa göz buyuklugu arasinda bir baglanti ariyoruz. Ya da baska bir seyi olcuyoruz, genel olarak bir x degiskeni icin onun etki ettigi, alakali oldugu bir y degiskenini olcuyoruz.

Olculen iki degiskeni grafikleyince, su ortaya cikiyor diyelim.



Arada bir korelasyon oldugunu goruyoruz. Bu konu hakkında bilimsel bir makale yaziyor olsaydik, bu grafikte soyle bir cizgi cizerdik,



Fakat veri noktalarinin tam ortasindan gecen bu cizgiyi nasil cizecegiz? Yani cozmek istedigimiz problem verilen $(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$ seklindeki deney verileri icin “en iyi uyan (best fit)” cizgi $y = ax + b$, en iyi yaklasiksalik nedir?

Bu problemde onemli bir puf noktasina isaret edelim: $y = ax + b$ denkleminde bilinmeyenler nedir? Genellikle ogrenciler x ve y degiskenine bakiyorlar. Fakat bu dogru degil. Bir optimal cizgi ortaya cikarmak istiyorsak ilgilendigimiz x ve y degil, esas ilgilendiklerimiz a ve b katsayilari. Cizginin hangi sekilde oldugunu onlar kontrol ediyorlar, yani dogru uyum icin cizginin nereden gectiginin hesaplanmasi problemindeki bilinmeyenler onlar. Yani uyum icin en iyi a ve b 'yi bulmamiz gerekiyor.

Bu noktada “en iyi a ve b ” ifadesinin ne olduguna karar vermemiz lazim. En iyi, a ve b 'nin bir fonksiyonunun minimize edilmesi olabilir, ki bu fonksiyon deneysel veri ile bir teorik cizgi arasindaki uyum hatalarinin toplamini temsil edebilir. Yani hata, o cizginin deney noktalarindan ne kadar uzakta oldugunun toplami ile temsil edilebilir.

“Uzakligi” hesaplamanin da degisik yollari olabilir. Mesela her noktanin bir cizgiye olan duz uzakligi olculebilir. Ya da deneysel noktadan dikey olarak yukari / asagi cikip cizgiye gelinceye kadar olan uzaklik. Ya da uzakligi en fazla olan tek noktanin mesafesi azaltilmaya ugrasilabilir (ama bu son yontem pek iyi bir secim olmayabilir, cunku belki deney sirasinda uykuya dalmissinizdir, ve cok yanlis bir nokta olmuussunuzdur, ve o nokta tum uyum hesabini bozukluga ugratir).

Bu tur seceneklerden bir tanesi en iyisidir, ve evrensel olarak kullanilan yaklasim da o'dur. En Az Kareler demistik, bu yontemde hata noktalarin cizgiye olan uzakliklari karesinin toplamidir. Bu yontem iyi sonuclar veriyor ve hesap icin oldukca temiz bir formül ortaya cikartiyor. Demek ki “en iyi”

tanımı hata noktaların çizgiden olan sapmasının karesinin toplamının minimize edilmesi demek. Sapma nedir? Tahmin edilen ile gerçek veri noktasının farkıdır.

$$y_i - (ax_i + b)$$

O zaman problem

$$\text{Minimize Et } D = \sum_{i=1}^n \left[y_i - (ax_i + b) \right]^2$$

Tekrar vurgulayalım, bu fonksiyonda bilinmeyenler a ve b . x_i ve y_i deneyden gelen veriler.

Minimize etmek için şimdiye kadar öğrendiklerimizi kullanabiliriz. Kritik noktayı bulalım.

Yani istediğimiz

$$\frac{\partial D}{\partial a} = 0$$

$$\frac{\partial D}{\partial b} = 0$$

esitliklerini doğru olduğu an. Kritik nokta burada. Kısmi türevleri alalım.

$$\frac{\partial D}{\partial a} = \sum_{i=1}^n 2(y_i(ax_i + b))(-x_i) = 0$$

$$\frac{\partial D}{\partial b} = \sum_{i=1}^n 2(y_i(ax_i + b))(-1) = 0$$

Çözmemiz gereken denklemler bunlar.

Eğer dikkat edersek bu denklemler a ve b bağlamında lineer. Denklemlerde biraz kalabalıklık var, onları acip tekrar düzenleyerek bu lineerliği görmeye uğrasalım.

İlk önce '2' terimini atalım, ona gerek yok. İki denklem ayrı ayrı şöyle olur:

$$\sum_{i=1}^n (x_i^2 a + x_i b - x_i y_i) = 0$$

$$\sum_{i=1}^n (x_i^2 a + b - y_i) = 0$$

a ve b leri yanyana getirelim. Yine ayrı ayrı

$$\left(\sum_{i=1}^n x_i^2 \right) a + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n x_i y_i$$

$$\left(\sum_{i=1}^n x_i \right) a + nb = \sum_{i=1}^n y_i$$

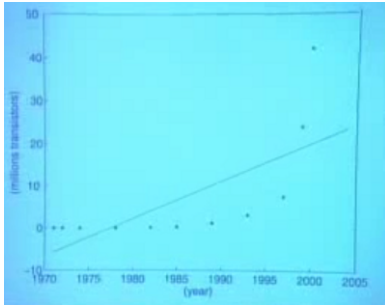
Parantezler icindeki x_i 'li ifadeler korkutucu gorunuyor olabilir, fakat bunlar deney verisinden gelen sayilarin toplamindan ibaret, onlar elimizde sayisal olarak mevcut zaten. Deney verisini alip, hepsini toplayinca bu sayiyi elde edecegiz.

Sonuc olarak elimize gecen 2 x 2 boyutlarında bir lineer sistem. Yani x_i ve y_i iceren ifadeleri hesapladigimiz anda bu sistemi elde ederiz, ve 2 x 2 bir lineer sistemi cozmeyi zaten biliyoruz. Ve kritik noktayi Boylece elde ederiz. Bir sonraki derste gorecegimiz 2. kısmi turevleri kullanacagiz yontemle de bu noktanin min/maks oldugunu anlariz. Bu testi uygulusak ustteki yontemin hakikaten bir min urettigini gorebilirdik.

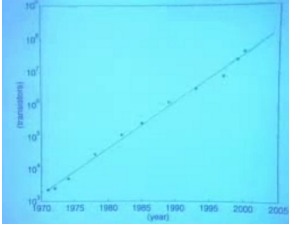
En Az Kareler interpolasyonu cok daha genel kullanimlarda da ise yarar.

Ornek

Bilgisayar dunyasinda Moore Kanunu denen bir kural vardir, bu kural bilgisayar ciplerinin nasil surekli daha hizli, daha iyiye dogru gittigini anlatir. Unlu cip ureticisi Intel baskani Andy Grove tarafından ortaya atilmistir, bir hipotezdir, fakat sasirtici bir sekilde dogru cikmistir. Kuralin olctugu bir mikropcipin icine koyulabilecek transistor sayisidir. Bu olcunun bir ornegi alttadir.



Bu veriye lineer olarak uyum yapamazdik. Fakat logaritmik skalayı kullanırsak, yani transistor sayısı yerine onun logaritmasını alıp grafiklersek,



veriler daha çizgisel olurlar. Bu demektir ki zaman ve transistor sayısı arasındaki ilişki üstel (exponential) bir ilişkidir. Zaten kuralın söylediği de bu, kural her 18 ayda bir çip içindeki transistor sayısının ikiye katlandığını söyler. Bir sonraki büyüklüğün o büyüklüğün o anki halinin belli bir kati olması (aynen nüfus artışında olduğu gibi) üstel bir ilişkiye işaret eder.

Peki en iyi üstel uyumu nasıl buluruz? Böyle bir uyumun formül hali şudur. Düz çizgi formülü yerine bir üstel ifade içerir.

$$y = ce^{ax}$$

Fakat bu formülü direkt hata hesabında kullanırsak, ele geçen formüller çok karmaşık hale geliyor. Ama üstteki numarayı hatırlayalım, log skalasında bakınca her şey lineer çıkıyor. O zaman uyumu şu şekilde yapabiliriz, üstteki formülün log'unu alalım:

$$\ln(y) = \ln(c) + ax$$

ki bu formül lineerdir. Bunun üzerinde En Az Kareler yöntemini kullanabiliriz.

Üstel ilişki yerine karesel bir ilişki de olabilirdi, mesela

$$y = ax^2 + bx + c$$

o zaman en iyi parabolü uydurmaya uğraşabiliriz, bu uyum için a , b ve c 'yi bulmamız gerektirdi. Yani bunu minimize edecektik:

$$D(a, b, c) = \sum_{i=1}^n (y_i - (ax_i^2 + bx_i + c))^2$$

Burada kısmi türevler 3 tane ayrı denklem üretir, ve ilişki yine lineer çıkar, 3 x 3 boyutunda bir sistem elde ederiz.

Yani soyledigimiz gibi, bu problemler ilk basta minimizasyon problemi gibi gozukmeyebiliyordu, ama oyle olduklarini simdi gormus olduk.