Matris Türevleri

Aksi belirtilmedikce altta a, x gibi vektorler kolon vektorleri olacaktir, yani $m \times 1$, ya da $n \times 1$ gibi boyutlara sahip olacaklardir.

m boyutlu vektor x'i alan ve geriye tek sayi sonucu donduren bir $f(x)$ fonksiyonunun x'e gore turevini nasil aliriz? Yani $x \in \mathbb{R}^m$ ve bir vektor,

$$x = \left[\begin{array}{c} x_1 \\ \vdots \\ x_m \end{array}\right]$$

Bu durumda x'in her hucresine / ogesine gore kismi turevler (partial derivatives) alinir, sonucta tek boyutlu / tekil sayili fonksiyon, turev sonrasi m boyutlu bir sonuc vektorunu yaratir, yani

$$\frac{\partial f}{\partial x} = \left[\begin{array}{c} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{array}\right]$$

Bu sonuc tanidik gelmis olabilir, bu ifade gradyan olarak ta bilinir.

$$\frac{\partial f}{\partial x} = \nabla f = grad\ f(x)$$

Turev bir kolon vektoru olarak cikti cunku x de bir kolon vektoruydu. Elde edilen vektor surpriz degil cunku tek, skalar bir deger veren bir fonksiyonun x icindeki *her ogensinin* nasil degistigine gore bunun fonksiyon uzerindeki etkilerini merak ediyorduk, ustteki vektor oge bazinda bize aynen bunu gosteriyor. Yani tek skalar sonuc m tane turev sonucuna ayriliyor, cunku tek sonucun m tane secenege gore degisimini gormek istedik.

Not olarak belirtelim, gradyan vektoru matematiksel bir kisayoldur, yani matematikte kuramsal olarak turetilerek ulasilan ana kurallardan biri denemez. Fakat cok ise yaradigina suphe yok.

Eger bir A matrisinin tum ogeleri tek sayi θ gibi bir degiskene bagli ise, o matrisin θ'ya gore turevi, tum elemanlarinin teker teker θ'ya gore turevidir,

$$\frac{\partial A}{\partial \theta} = \left[\begin{array}{cccc} \frac{\partial a_{11}}{\partial \theta} & \frac{\partial a_{12}}{\partial \theta} & \cdots & \frac{\partial a_{1n}}{\partial \theta} \\ \frac{\partial a_{21}}{\partial \theta} & \frac{\partial a_{22}}{\partial \theta} & \cdots & \frac{\partial a_{2n}}{\partial \theta} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{m1}}{\partial \theta} & \frac{\partial a_{m2}}{\partial \theta} & \cdots & \frac{\partial a_{mn}}{\partial \theta} \end{array}\right]$$

Simdi ilginc bir durum; diyelim ki hem fonksiyon $f(x)$'e verilen x cok boyutlu, hem de fonksiyonun sonucu cok boyutlu! Bu gayet mumkun bir durum. Bu durumda ne olurdu?

Eger f'in turevinin her turlu degisimi temsil etmesini istiyorsak, o zaman hem her girdi hucresi, hem de her cikti hucresi kombinasyonu icin bu degisimi saptamaliyiz. Jacobian matrisleri tam da bunu yapar. Eger m boyutlu girdi ve n boyutlu cikti tanimlayan f'in turevini almak istersek, bu bize $m \times n$ boyutunda bir matris verecektir! Hatirlarsak daha once gradyan sadece m boyutunda bir *vektor* vermisti. Simdi sonuc bir matris.

$$ J(x) = \frac{\partial f(x)}{\partial x} = \left[\begin{array}{ccc} \frac{\partial f_1(x)}{\partial \theta} & \dots & \frac{\partial f_1(x)}{\partial \theta} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial \theta} & \dots & \frac{\partial f_n(x)}{\partial \theta} \end{array}\right] $$

Vektorlere gelelim. $a, x \in \mathbb{R}^n$ ise, $a^T x$'in x'e gore turevi nedir?

$a^T x$ bir noktasal carpim olduguna gore sonucu bir tek sayidir (scalar). Bu noktasal carpimi bir fonksiyon gibi dusunebiliriz, bu durumda demektir ki tek sayili bir fonksiyonun cok boyutlu x'e gore turevini aliyoruz, yani gradyan durumu tekrar vuku buldu,

$$ \frac{\partial (a^T x)}{\partial x} = \left[\begin{array}{c} \frac{\partial (a^T x)}{\partial x_1} \\ \vdots \\ \frac{\partial (a^T x)}{\partial x_n} \end{array}\right] $$

$$ = \left[\begin{array}{c} \frac{\partial (a_1 x_1 + ... + a_n x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial (a_1 x_1 + ... + a_n x_n)}{\partial x_n} \end{array}\right] = \left[\begin{array}{c} a_1 \\ \vdots \\ a_n \end{array}\right] = a $$

Niye her satirda a_1, a_2 gibi degerler elde ettigimizin sebebi bariz herhalde, cunku mesela ilk satirda x_1'e gore turev alindigi durumda $a_1 x_1$ haricindeki tum terimler yokolacaktir, cunku o terimler icinde x_1 yoktur ve Calculus'a gore sabit sayi sayilirlar.

Peki $a^T x$'in x^T'ye gore turevi nedir?

x^T'nin yatay bir vektor olduguna dikkat, yani satir vektoru, o zaman sonuc yatay bir vektor olacaktir (kiyasla gradyan dikeydi).

$$ \frac{\partial (a^T x)}{\partial x^T} = \left[\begin{array}{ccc} \frac{\partial (a^T x)}{\partial x_1} & \dots & \frac{\partial (a^T x)}{\partial x_n} \end{array}\right] \tag{1} $$

$$ = \left[\begin{array}{ccc} a_1 & \dots & a_n \end{array}\right] = a^T $$

Eger bir $x \in \mathbb{R}^m$ vektorunden A matrisi x ile carpiliyor ise, bu carpimin x'e gore turevi nedir?

$$\frac{\partial(Ax)}{\partial x^T} = A$$

Ispat: Eger $a_i \in \mathbb{R}^n$ ise (ki devrigi alininca bu vektor yatay hale gelir, yani altta bu yatay vektorleri ust uste istifledigimizi dusunuyoruz),

$$A = \left[\begin{array}{c} a_1^T \\ \vdots \\ a_m^T \end{array} \right]$$

O zaman Ax ne olur? *Matris Carpimi* yazisindaki "satir bakis acisi" dusunulurse, A'in her satiri, ayri ayri x'in tum satirlarini kombine ederek tekabul eden sonuc satirini olusturur (tabii bu ornekte x'in kendisi bir vektor o yuzden "satirlari" tek bir sayidan ibaret),

$$Ax = \left[\begin{array}{c} a_1^Tx \\ \vdots \\ a_m^Tx \end{array} \right]$$

Ustteki bir vektor, her ogesi tek sayi. Turevi alirsak (dikkat yatay vektore gore turev aliyoruz), ve (1)'i dikkate alirsak,

$$\frac{\partial Ax}{\partial x^T} = \left[\begin{array}{c} \frac{\partial(a_1^Tx)}{\partial x^T} \\ \vdots \\ \frac{\partial(a_m^Tx)}{\partial} \end{array} \right] = \left[\begin{array}{c} a_1^T \\ \vdots \\ a_m^T \end{array} \right] = A$$

Su turev nasil hesaplanir?

$$\frac{\partial(x^TA^T)}{\partial x}$$

Carpimi acalim, x devrigi yatay bir vektor, A'nin satirlari a_i'ler ise devrik sonrasi kolonlar haline gelirler, yani

$$\left[\begin{array}{ccc} x_1 & \dots & x_n \end{array} \right] \left[\begin{array}{ccc} \uparrow & & \uparrow \\ a_1 & \dots & a_m \\ \downarrow & & \downarrow \end{array} \right]$$

Simdi matris carpimi satir bakisini kullanalim, carpan x satiri bir tane, o zaman sonuc tek satir olacak. Bu tek x satirinin ogeleri, A^T satirlarini teker teker carpip toplayacak ve o tek sonuc satirini meydana getirecek. Yani $x_1, x_2, ...$ sirayla a_1'in tum ogelerini carpiyor, ayni sekilde a_2 icin aynisi oluyor, vs. Bu durumu daha temiz bir sekilde alttaki gibi belirtebiliriz,

$$= \begin{pmatrix} x^T a_1 & \dots x^T a_m \end{pmatrix}$$

Bu ifadenin turevini almak cok kolay,

$$= \begin{pmatrix} \frac{\partial(x^T a_1)}{\partial x} & \dots \frac{\partial(x^T a_m)}{\partial x} \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & \dots a_m \end{pmatrix}$$

$$\frac{\partial(x^T A^T)}{\partial x} = A^T \tag{2}$$

Baska bir turev: Diyelim ki $x \in \mathbb{R}^n, z \in \mathbb{R}^m$. Yeni bir vektor $c = A^T z$ tanimlayalim ki vektor n boyutunda. O zaman

$$\frac{\partial(z^T A x)}{\partial x} = \frac{\partial(c^T x)}{\partial x} = c = A^T z$$

Diger bir turev

$$\frac{\partial(z^T A x)}{\partial z} = Ax$$

Ustteki (2)'nin bir sonucu olarak gorulebilir.

Eger $x = x(\alpha), z = z(\alpha)$ olursa, turev almada Zincir Kuralini kullanalim,

$$\frac{\partial(z^T A x)}{\partial \alpha} = \frac{\partial(z^T A x)}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial(z^T A x)}{\partial z}\frac{\partial z}{\partial \alpha}$$

$$\frac{\partial(z^T A x)}{\partial \alpha} = A^T z \frac{\partial x}{\partial \alpha} + Ax \frac{\partial z}{\partial \alpha}$$

Eger $z = x = \alpha$ dersek,

$$\frac{\partial(x^T A x)}{\partial \alpha} = \frac{\partial(x^T A x)}{\partial x} = A^T x \frac{\partial x}{\partial \alpha} + Ax \frac{\partial x}{\partial \alpha}$$

$$= Ax + A^T x = (A^T + A)x$$

Diyelim ki A simetrik bir matris, o zaman $A^T = A$, ve

$$= (A^T + A)x = 2Ax$$

x^Tx ifadesinin x vektorune gore turevi nedir? En basit yol, x^TAx kalibini kullanmak, ve $A = I$ yani birim matrisi koymak. Bu durumda

$$= \frac{\partial (x^T I x)}{\partial \alpha} = 2Ix = 2x$$

Daha zor yoldan, bu bir noktasal carpim olacaktir, $x_1x_1 + x_2x_2 + .. + x_nx_n$ yani $x_1^2 + x_2^2 + .. + x_n^2$. Bu tek bir skalar sonuc, o sonucun her x ogesine gore turevinin alinmasi,

$$\begin{bmatrix} \frac{\partial (x_1^2 + x_2^2 + .. + x_n^2)}{\partial x_1} \\ \dots \\ \frac{\partial (x_1^2 + x_2^2 + .. + x_n^2)}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 \\ \dots \\ 2x_n \end{bmatrix} = 2x$$

Kaynaklar

Economics 627 Econometric Theory II, Vector and Matrix Differentiation, `http://faculty.arts.ubc.ca/vmarmer/econ627/`

Duda, Hart, *Pattern Classification*

`http://math.umn.edu/~mathe233/math2263/resources/sum11_math2263_chainrule.pdf`