

Matris Türevleri

Aksi belirtilmedikçe altta a, x gibi vektorler kolon vektorleri olacaktır, yani $m \times 1$, ya da $n \times 1$ gibi boyutlara sahip olacaklardır.

m boyutlu vektor x 'i alan ve geriye tek sayı sonucu döndüren bir $f(x)$ fonksiyonunun x 'e göre türevini nasıl alırız? Yani $x \in \mathbb{R}^m$ ve bir vektor,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

Bu durumda x 'in her hücreğine / ögesine göre kısmi türevler (partial derivatives) alınır, sonuçta tek boyutlu / tekil sayılı fonksiyon, türev sonrası m boyutlu bir sonuç vektorunu yaratır, yani

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{bmatrix}$$

Bu sonuç tanıdık gelmiş olabilir, bu ifade gradyan olarak da bilinir.

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

Türev bir kolon vektoru olarak çıktı çünkü x de bir kolon vektoruydu. Elde edilen vektor surpriz değil çünkü tek, skalar bir değer veren bir fonksiyonun x içindeki *her ögensinin* nasıl değiştiğine göre bunun fonksiyon üzerindeki etkilerini merak ediyorduk, üstteki vektor öge bazında bize aynen bunu gösteriyor. Yani tek skalar sonuç m tane türev sonucuna ayrılıyor, çünkü tek sonucun m tane seçeneğe göre değişimini görmek istedik.

Not olarak belirtelim, gradyan vektoru matematiksel bir kısayoldur, yani matematikte kuramsal olarak türetilerek ulaşılan ana kurallardan biri denemez. Fakat çok ise yaradığına şüphe yok.

Eğer bir A matrisinin tüm ögeleri tek sayı θ gibi bir değişkene bağlı ise, o matrisin θ 'ya göre türevi, tüm elemanlarının teker teker θ 'ya göre türevidir,

$$\frac{\partial A}{\partial \theta} = \begin{bmatrix} \frac{\partial a_{11}}{\partial \theta} & \frac{\partial a_{12}}{\partial \theta} & \cdots & \frac{\partial a_{1n}}{\partial \theta} \\ \frac{\partial a_{21}}{\partial \theta} & \frac{\partial a_{22}}{\partial \theta} & \cdots & \frac{\partial a_{2n}}{\partial \theta} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{m1}}{\partial \theta} & \frac{\partial a_{m2}}{\partial \theta} & \cdots & \frac{\partial a_{mn}}{\partial \theta} \end{bmatrix}$$

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

Eğer f 'in turevinin her türlü değişimi temsil etmesini istiyorsak, o zaman hem her girdi hücresi, hem de her çıktı hücresi kombinasyonu için bu değişimi saptamalıyız. Jacobian matrisleri tam da bunu yapar. Eğer m boyutlu girdi ve n boyutlu çıktı tanımlayan f 'in turevini almak istersek, bu bize $m \times n$ boyutunda bir matris verecektir! Hatırlarsak daha önce gradyan sadece m boyutunda bir *vektor* vermisti. Simdi sonuc bir matris.

$$J(x) = \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial \theta} & \cdots & \frac{\partial f_1(x)}{\partial \theta} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial \theta} & \cdots & \frac{\partial f_n(x)}{\partial \theta} \end{bmatrix}$$

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

$a^T x$ bir noktasal çarpım olduğuna göre sonucu bir tek sayıdır (scalar). Bu noktasal çarpımı bir fonksiyon gibi düşünebiliriz, bu durumda demektir ki tek sayılı bir fonksiyonun çok boyutlu x 'e göre turevini alıyoruz, yani gradyan durumu tekrar vuku buldu,

$$\begin{aligned} \frac{\partial(a^T x)}{\partial x} &= \begin{bmatrix} \frac{\partial(a^T x)}{\partial x_1} \\ \vdots \\ \frac{\partial(a^T x)}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial(a_1 x_1 + \dots + a_n x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial(a_1 x_1 + \dots + a_n x_n)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a \end{aligned}$$

Niye her satırda a_1, a_2 gibi değerler elde ettiğimizizin sebebi bariz herhalde, çünkü mesela ilk satırda x_1 'e göre turev alındığı durumda $a_1 x_1$ haricindeki tüm terimler yokolacaktır, çünkü o terimler içinde x_1 yoktur ve Calculus'a göre sabit sayı sayılırlar.

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

x^T 'nin yatay bir vektor olduğuna dikkat, yani satır vektörü, o zaman sonuc yatay bir vektor olacaktır (kiyasla gradyan dikeydi).

$$\begin{aligned} \frac{\partial(a^T x)}{\partial x^T} &= \begin{bmatrix} \frac{\partial(a^T x)}{\partial x_1} & \cdots & \frac{\partial(a^T x)}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = a^T \end{aligned} \tag{1}$$

Eger bir $x \in \mathbb{R}^n$ 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),

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$$

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),

$$Ax = \begin{bmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

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

$$\frac{\partial Ax}{\partial x^T} = \begin{bmatrix} \frac{\partial(a_1^T x)}{\partial x^T} \\ \vdots \\ \frac{\partial(a_m^T x)}{\partial x^T} \end{bmatrix} = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} = A$$

Su turev nasil hesaplanir?

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

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

$$\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \uparrow & & \uparrow \\ a_1 & \dots & a_m \\ \downarrow & & \downarrow \end{bmatrix}$$

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, \dots 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,

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

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

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

$x^T x$ ifadesinin x vektorune gore turevi nedir? Bu bir noktasal carpim olacaktır, $x_1 x_1 + x_2 x_2 + \dots + x_n x_n$ yani $x_1^2 + x_2^2 + \dots + x_n^2$. Bu tek bir skalar sonuc, o sonucun her x ogesine gore turevinin alınmasi,

$$\begin{aligned} & \begin{bmatrix} \frac{\partial(x_1^2 + x_2^2 + \dots + x_n^2)}{\partial x_1} \\ \dots \\ \frac{\partial(x_1^2 + x_2^2 + \dots + x_n^2)}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 \\ \dots \\ 2x_n \end{bmatrix} = 2x \end{aligned}$$

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

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

Ustteki (2)'nin bir sonucu olarak gorulebilir.

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

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

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

Eğer $z = x = \alpha$ dersek,

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

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

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

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