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.

Gradyan

$m$ 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.

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

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 rahatliktir, bir kisayoldur, bir ziplama noktasidir, yani matematiksel olarak turetilerek ulasilan ana kurallardan biri denemez. Fakat cok ise yaradigina suphe yok.

Tek Sayi Parametreye Gore Matris Turevi

Eger bir $A$ matrisinin tum ogeleri bir tek sayi parametresi $\theta$'e bagli ise, o matrisin $\theta$'ya gore tum elemanlarinin teker teker $\theta$'ya gore turevidir,

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

Cok Parametreli Matris Turevi

Simdi ilginc bir varyasyon; 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.

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

Vektor Turevleri

$a, x \in \mathbb{R}^n$ 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. Bu durumu ustte gorduk, sonuc bir gradyan olacaktir,

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

Peki $a^T x$'in $x^T$'ye gore turevi nedir? $x^T$'nin yatay bir vektor olduguna dikkat, yani satir vektorudur, o zaman sonuc ta yatay bir vektor olur (kiyasla gradyan dikeydi).

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

$$= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} = a^T$$

Matris Turevleri

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

$$\frac{\partial}{\partial x^T}[Ax] = A$$

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, o zaman

$$\frac{\partial}{\partial x^T}[Ax] =$$

Kaynaklar

Duda, Hart, *Pattern Classification*

Kaynaklar

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

Duda, Hart, *Pattern Classification*