به نام خدا

دانشگاه صنعتی شریف - دانشکده مهندسی کامپیوتر

مقدمه‌ای بر یادگیری ماشین

فصل اول: محاسبات و بهینه سازی

نویسندگان: پرناز بابلیان، کوروش مسلمی

2.مشتق، گرادیان و قاعده زنجیره ای

در این بخش به یادآوری مباحثی از ریاضی 1 و 2 می پردازیم.

1.2.مشتق

در ساده ترین تعریف مشتق یک تابع نرخ تغییرات آن نسبت به آرگومان هایش را نشان می دهد. مشتق تابع f تابعی است به نام f' که به صورت زیر تعریف می شود:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

این حد به ما می گوید نسبت تغییر مقدار تابع، $f(x+h)-f(x)$ ، به تغییر اعمال شده، h ،

این حد به ما می گوید نسبت تغییر مقدار تابع، $f(x+h)-f(x)$ ، به تغییر اعمال شده، h ، وقتی h به صفر میل می کند به چه مقداری همگرا می شود. اگر حاصل متناهی باشد گوییم f در xمشتق پذیر است. در ادامه این همگرایی را برای تابع $f(x)=3x^2-4x$ در $x=1$ به صورت عددی بررسی می کنیم.

```
import numpy as np
def f(x):
    return 3 * x ** 2 - 4 * x

for h in 10.0**np.arange(-1, -6, -1):
    print(f'h={h:.5f}, numerical limit={(f(1+h)-f(1))/h:.5f}')
```

```
h=0.10000, numerical limit=2.30000
h=0.01000, numerical limit=2.03000
h=0.00100, numerical limit=2.00300
h=0.00010, numerical limit=2.00030
h=0.00001, numerical limit=2.00003
```

در کتب مختلف نماد های متفاوتی برای مشتق گیری استفاده می شود که همگی معادل هستند.اگر قرار دهیم $y = f(x)$ نماد های $\frac{d}{dx}f(x)$, $\frac{dy}{dx}$ را نماد های لایب نیتس و علامت پریم (y') نیز نماد نیوتن برای مشتق می نامند. به عبارت دیگر:

$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x) = Df(x) = D_x f(x)$$

نماد های $\frac{d}{dx}$ و D عملگرهای دیفرانسیل نامیده می شوند. در ادامه مشتق چند تابع متداول آمده است:

$$\begin{aligned}
\frac{d}{dx}C &= 0 && \text{for any constant } C \\
\frac{d}{dx}x^n &= nx^{n-1} && \text{for } n \neq 0 \\
\frac{d}{dx}e^x &= e^x \\
\frac{d}{dx}\ln x &= x^{-1}
\end{aligned}$$

همچنین قوانین زیر برای مشتق گیری سریع تر به ما کمک می کنند:

$$\begin{aligned}
\frac{d}{dx}[Cf(x)] &= C\frac{d}{dx}f(x) && \text{Constant multiple rule} \\
\frac{d}{dx}[f(x)+g(x)] &= \frac{d}{dx}f(x) + \frac{d}{dx}g(x) && \text{Sum rule} \\
\frac{d}{dx}[f(x)g(x)] &= f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x) && \text{Product rule} \\
\frac{d}{dx}\frac{f(x)}{g(x)} &= \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g^2(x)} && \text{Quotient rule}
\end{aligned}$$

اکنون آماده ایم مشتقی که به صورت عددی حساب کردیم را به صورت تحلیلی محاسبه کنیم:

$$\frac{d}{dx}[3x^2 - 4x] = 3\frac{d}{dx}x^2 - 4\frac{d}{dx}x = 6x - 4.$$

که در $x = 1$ همان 2 است.همچنین در پایتون نیز می توانیم به صورت خودکار و نمادین این مشتق گیری را انجام دهیم.

```
from sympy import *
x = Symbol('x', real=True)
f = 3*x**2 - 4*x

dfdx = f.diff(x) # <- yes, taking derivatives is this easy!

print("f'(x) =", latex(simplify(dfdx)))
```

```
f'(x) = 6 x - 4
```

2.2.مشتق های جزئی و گرادیان

در این قسمت ابتدا تعمیم تعریف قبلی برای توابع چند متغیره را بیان می کنیم. فرض کنیم تابع $y = f(x_1, x_2, \ldots, x_n)$ یک تابع n متغیره باشد. مشتق جزئی y نسبت به i امین پارامتر (x_i) به صورت زیر تعریف می شود:

$$\frac{\partial y}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}.$$

برای محاسبه $\frac{\partial y}{\partial x_i}$ پارامتر های $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ را ثابت گرفته و مشتق y نسبت به x_i را محاسبه می کنیم. مشابه قبل نماد های زیر معادل هستند:

$$\frac{\partial y}{\partial x_i} = \frac{\partial f}{\partial x_i} = \partial_{x_i} f = \partial_i f = f_{x_i} = f_i = D_i f = D_{x_i} f.$$

مثال) اگر $w = f(x, y, z, t)$ در این صورت:

$$\frac{\partial^7 w}{\partial x\ \partial y\ \partial z\ \partial^3 x\ \partial t} = f_{txxxzyx}(x, y, z, t)$$

مثال) اگر $f(x, y, z) = (sinx)y - z^2$ در این صورت:

$$f_x(x, y, z) = y(cosx) \qquad f_y(x, y, z) = sinx \qquad f_z(x, y, z) = -2z$$

مثال) می خواهیم $\frac{\partial^7}{\partial x \partial y^2 \partial z^4} e^{xyz}$ را با پایتون محاسبه کنیم.

```
x, y, z = symbols('x y z')
expr = exp(x*y*z)
diff(expr, x, y, 2, z, 4)
```

$$x^3 y^2 \left(x^3 y^3 z^3 + 14x^2 y^2 z^2 + 52xyz + 48\right) e^{xyz}$$

اکنون آماده ایم تا گرادیان را تعریف کنیم. فرض کنید f تابعی n متغیره به صورت

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$
$$f := f(x_1, x_2, \cdots, x_n)$$

باشد آنگاه گرادیان f را به صورت زیر تعریف می‌کنیم

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n})^T$$

بنابراین $\nabla_{\mathbf{x}} f(\mathbf{x})$ برداری در $\mathbb{R}^n$ خواهد بود. در مواقعی که ابهامی نباشد به جای $\nabla_{\mathbf{x}} f(\mathbf{x})$ از نماد ∇f استفاده می کنیم.

مثال) فرض کنید

$$f(x_1, x_2) = \frac{1}{x_1} - 4x_2$$

آنگاه بردار گرادیان تابع f به صورت زیر خواهد بود

$$\nabla f = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})^T$$

که در آن

$$\frac{\partial f}{\partial x_1} = \frac{\partial(\dfrac{1}{x_1} - 4x_2)}{\partial x_1} = -\frac{1}{x_1^2} \quad , \quad \frac{\partial f}{\partial x_1} = \frac{\partial(\dfrac{1}{x_1} - 4x_2)}{\partial x_2} = -4$$

بنابراین داریم

$$\nabla f = (-\frac{1}{x_1^2}, -4)^T$$

همچنین می توانیم این بردار را با پایتون محاسبه کنیم.

```
x, y = symbols('x y')
expr = (1/x)-4*y
[expr.diff(xi) for xi in [x,y]]
```

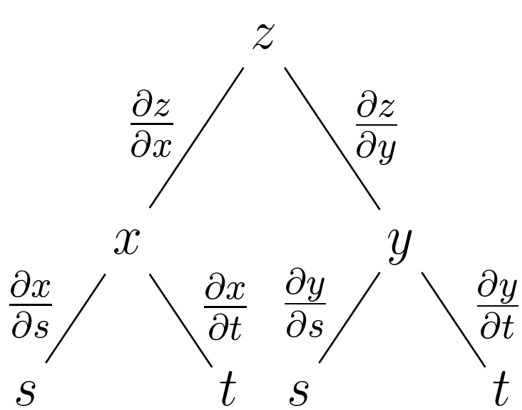
```
[-1/x**2, -4]
```

3.2.قاعده زنجیره ای

فرض کنید $z = f(x, y)$ با مشتقات جزئی مرتبه اول پیوسته و x و y توابعی مشتق پذیر برحسب t و s باشند آنگاه طبق قاعده زنجیره ای داریم:

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}.\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}.\frac{\partial y}{\partial t}$$

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}.\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}.\frac{\partial y}{\partial s}$$



در بخش بعد برهان صوری تری از این حالت ساده ولی نماینده قائده زنجیره ای داده خواهد شد . دو معادله ی فوق را میتوان در معادله ماتریسی زیر گنجاند .

$$\begin{pmatrix} \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{pmatrix}$$

به طور کلی ،هرگاه z تابعی از چند متغیر اولیه باشد . هر یک از این متغیر ها تابع چند متغیر ثانیه باشد ، آنگاه مشتق جزئی z نسبت به یکی از متغیر های ثانویه چند جمله دارد که به ازای هر متغیر ثانویه که z تابع آن است یک جمله برای شرکت در مشتق ظاهر می

شود .

مثال) اگر $w = z(u, v, r)$ و u و v برحسب x,y,r باشند و r تابعی برحسب x,y باشد در این صورت $\frac{\partial z}{\partial x}$ را به صورت زیر محاسبه می کنیم:

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}.\frac{\partial u}{\partial x} + \frac{\partial z}{\partial u}.\frac{\partial u}{\partial r}.\frac{\partial r}{\partial x} + \frac{\partial z}{\partial v}.\frac{\partial v}{\partial x} + \frac{\partial z}{\partial v}.\frac{\partial v}{\partial r}.\frac{\partial r}{\partial x} + \frac{\partial z}{\partial r}.\frac{\partial r}{\partial x}$$

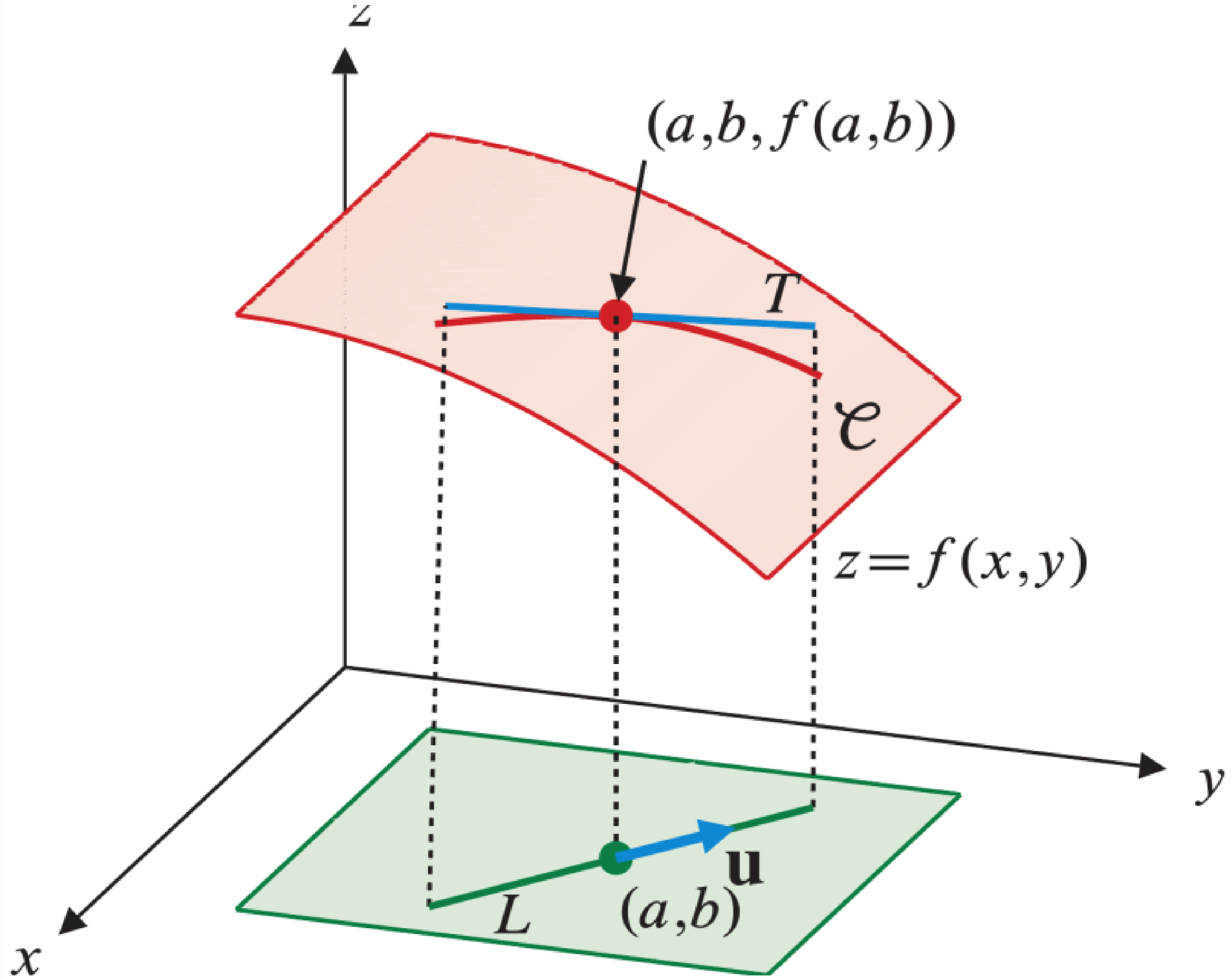
4.2.مشتقهای جهتی (Directional Derivative)

مشتقهای جزئی اول $f_1(a, b)$ و $f_2(a, b)$ میزانهای تغییر $f(x, y)$ در (a,b) را به دست می دهند که به ترتیب در جهتهای محورهای x و y مثبت سنجیده می شوند .جهت را می توان با یک بردار ناصفر (بردار یکه در بهترین راه) مشخص کرد . حال فرض کنیم $u = ui + uj$ یک بردار یکه باشد به طورکه $u^2 + v^2 = 1$ ، مشتق جهتی $f(x, y)$ در (a,b) در جهت u در صفحه xy سنجیده می شود . مشتق جهت‌دار با نمادهای زیر نشان داده می‌شود:

$$D_u f(a, b) = \lim_{h \to 0^+} \frac{f(a + hu, b + hv) - f(a, b)}{h}$$

مشتق های جهتی در جهتهای موازی با محور های مختصات مستقیما با مشتق های جزئی اول داده می شوند :

$$D_i f(a, b) = f_1(a, b)$$
$$D_j(a, b) = f_2(a, b)$$
$$D_{-i} f(a, b) = -f_1(a, b)$$
$$D_{-j} f(a, b) = -f_2(a, b)$$



بردار یکه u خط L را بر(a,b) در قلمرو f را معین میکند . صفحه قائم شامل L نمودار f را در منحنی C قطع می کند که مماسش T در $(a, b, f(a, b))$ دارای شیب $D_u f(a, b)$ است.

5.2.کاربرد گرادیان در یافتن مشتقهای جهتی

هرگاه f در (a,b) مشتق پذیر باشد و $u = ui + uj$ برداری یکه باشد ، مشتق جهتی f در جهت u برابر است با :

$$D_u f(a, b) = u \bullet \nabla f(a, b)$$

اگرv یک بردار ناصفر باشد ، همیشه می توان با تقسیم v بر طول آن ، یک بردار یکه در همان جهت به دست آورد .
لذا ، مشتق جهتی f در (a, b) در جهت v مساوی است با :

$$D_{v/|v|} f(a, b) = \frac{v}{|v|} \bullet \nabla f(a, b)$$

مثال) میزان تغییر $f(x, y) = y^4 + 2xy^3 + x^2y^2$ در (0,1) در هر یک از جهت های زیر را بیابید :

الف) $i+2j$ ب) $j+2i$ ج) $3i$

حل) داریم :

$$\nabla f(x,y) = (2y^3+2xy^2)i + (4y^3+6xy^2+2x^2y)j$$

,

$$\nabla f(0,1) = 2i+4j$$

الف)مشتق جهتی f در (0,1) در جهت $i+2j$ مساوی است با :

$$\frac{i+2j}{|i+2j|} \bullet (2i+4j) = \frac{2+8}{\sqrt{5}} = 2\sqrt{5}$$

ملاحظه می کنیم که $i+2j$ در همان جهت $\nabla f(0,1)$ است در نتیجه مشتق جهتی مثبت و مساوی طول $\nabla f(0,1)$ است.

ب) مشتق جهتی f در (0,1) در جهت $j-2i$ مساوی است با :

$$\frac{-2i+j}{|-2i+j|} \bullet (2i+4j) = \frac{-4+4}{\sqrt{5}} = 0$$

چون $j-2i$ بر $\nabla f(0,1)$ عمود است ، برمنحنی ترازِ f روی (0,1) مماس است ؛ در نتیجه مشتق جهتی در آن جهت صفر است .

پ)مشتق جهتی f در (0,1) در جهت $3i$ مساوی است با :

$$i \bullet (2i+4j) = 2.$$

همانطور که قبلا گفتیم ، مشتق جهتی f در جهت محور x مثبت مساوی $f_1(0,1)$ است.

6.2.تابع های ضمنی و دستگاه معادلات

ابتدا با مثالی از دستگاه معادلات شروع می کنیم و و سپس به تعریف قضیه تابع ضمنی می پردازیم :

مثال) فرض کنید x,y,u,v با معادلات زیر به هم مربوط شده باشند :

$$\begin{cases} u = x^2+xy-y^2 \\ v = 2xy+y^2 \end{cases}$$

الف) $(\partial x/\partial u)_v$ و ب) $(\partial x/\partial u)_y$ را در نقطه ای که $x=2$ و $y=-1$ بیابید .

حل) الف)برای محاسبه ی $(\partial x/\partial u)_v$، x و yرا اتوابعی از u و v گرفته و از معادلات داده شده نسبت به u مشتق گرفته v، را ثابت می گیریم :

$$1 = \frac{\partial u}{\partial u} = (2x+y)\frac{\partial x}{\partial u} + (x-2)\frac{\partial y}{\partial u},$$
$$0 = \frac{\partial v}{\partial u} = 2y\frac{\partial x}{\partial u} + (2x+2y)\frac{\partial y}{\partial u}$$

در$x=2$، $y=-1$ داریم

$$1 = 3\frac{\partial x}{\partial u} + 4\frac{\partial y}{\partial u}$$
$$0 = -2\frac{\partial x}{\partial u} + 2\frac{\partial y}{\partial u}$$

با حذف $\partial y/\partial u$ به نتیجه $(\partial x/\partial u)_v = 1/7$ می رسیم .

ب) برای محاسبه ی $(\partial x/\partial u)_y$ ، xو v را به عنوان توابعی از y و uگرفته و از معادلات داده شده با ثابت گرفتن y نسبت به u مشتق می گیریم :

$$1=\frac{\partial u}{\partial u}=(2x+y)\frac{\partial x}{\partial u},$$
$$\frac{\partial v}{\partial u}=2y\frac{\partial x}{\partial u}$$

در $x=2, y=-1$ معادله اول فورا نتیجه می دهد که $(\partial x/\partial u)_y=1/3$.

7.2.قضیه تابع ضمنی

مثال) دستگاه معادلات خطی زیر را در نظر میگیریم :

$$F(x,y,s,t)=a_1x+b_1y+c_1s+d_1t+e_1=0,$$
$$G(x,y,s,t)=a_2x+b_2y+c_2s+d_2t+e_2=0.$$

این دستگاه را می توان به شکل ماتریسی نوشت :

$$A\begin{pmatrix}x\\y\end{pmatrix}+C\begin{pmatrix}s\\t\end{pmatrix}+\varepsilon=\begin{pmatrix}0\\0\end{pmatrix}$$

که در آن

$$\varepsilon=\begin{pmatrix}e_1\\e_2\end{pmatrix},C=\begin{pmatrix}c_1&d_1\\c_2&d_2\end{pmatrix},A=\begin{pmatrix}a_1&b_1\\a_2&b_2\end{pmatrix}$$

معادلات را می توان نسبت به x و y به عنوان توابعی از s و t حل کرد مشروط براینکه $\det(A)\neq 0$ ، زیرا این وجود ماتریس معکوس A^{-1} را اینجاب می کند ؛ پس :

$$\begin{pmatrix}x\\y\end{pmatrix}=-A^{-1}\left(C\begin{pmatrix}s\\t\end{pmatrix}+\varepsilon\right)$$

ملاحظه می کنیم که $\det(A)=\partial(F,G)/\partial(x,y)$ ؛ در نتیجه صفر نشدن این ژاکوبین حل معادلات نسبت به x و y را تضمین خواهد کرد .

8.2.ماتریس ژاکوبی (Jacobian matrix)

فرض کنید تابع $f:R^n\rightarrow R^m$ با مقادیر $[f_1(x)(x_1...x_n),\dots,f_m(x)(x_1...x_n)]$ وجود دارد. به ماتریسی که از مشتق این تابع در هر نقطه از x_1 تا x_n تشکیل شده ماتریس ژاکوبی گفته می شود که برابر است با :

$$\mathbb{J}=\begin{bmatrix}\dfrac{\partial\mathbf{f}(\mathbf{x})}{\partial x_1}&\cdots&\dfrac{\partial\mathbf{f}(\mathbf{x})}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\nabla^T f_1(\mathbf{x})\\\vdots\\\nabla^T f_m(\mathbf{x})\end{bmatrix}=\begin{bmatrix}\dfrac{\partial f_1(\mathbf{x})}{\partial x_1}&\cdots&\dfrac{\partial f_1(\mathbf{x})}{\partial x_n}\\\vdots&\ddots&\vdots\\\dfrac{\partial f_m(\mathbf{x})}{\partial x_1}&\cdots&\dfrac{\partial f_m(\mathbf{x})}{\partial x_n}\end{bmatrix}$$

بنابراین :

$$\mathbb{J}_{i,j}=\frac{\partial f_i(\mathbf{x})}{\partial x_j}$$

($m=1$, Where $\nabla f=\mathbb{J}_f^T$)

دترمینان ژاکوبین دو تابع $u=u(x,y)$ و $v=v(x,y)$ نسبت به دو متغیر x و y دترمینان زیر است :

$$\frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix}\frac{\partial u}{\partial x}&\frac{\partial u}{\partial y}\\\frac{\partial v}{\partial x}&\frac{\partial v}{\partial y}\end{vmatrix}$$

به همین نحو ، ژاکوبین دو تابع $F(x,y,\dots)$ و $G(x,y,\dots)$ نسبت به متغیر های x و y دترمینان زیر می باشد :

$$\frac{\partial(F,G)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{vmatrix} = \begin{vmatrix} F_1 & F_2 \\ G_1 & G_2 \end{vmatrix}$$

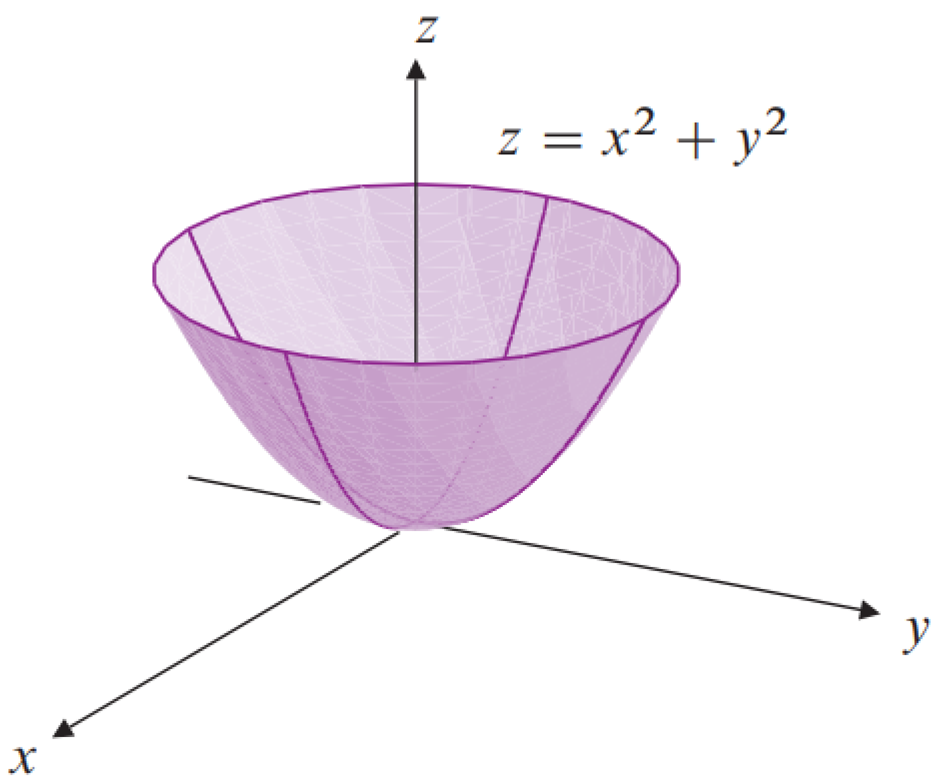
9.2.ماتریس هسین (Hessian matrix)

فرض کنید تابع $f : R^n \to R$ داریم ، اگر تمام مشتقات جزئی دوم f وجود داشته باشد آنگاه ماتریس هسین H از f برابر است با :

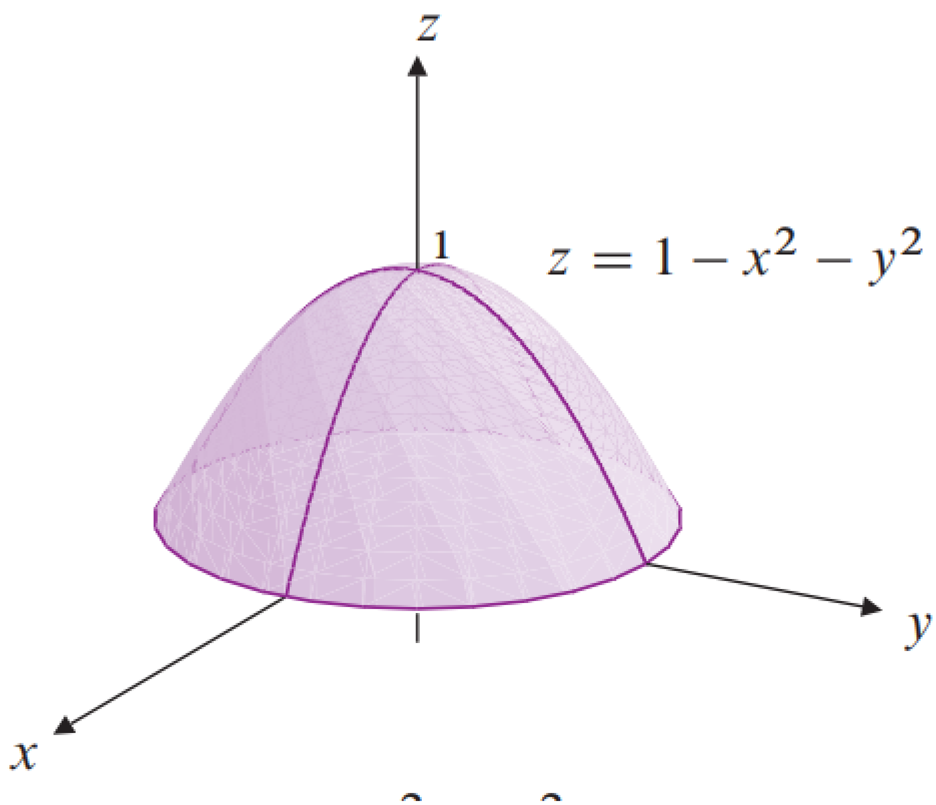
$$(HF_{i,j}) = \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial x_i \partial x_j}$$

ماتریس هسین یک ماتریس متقارن است، از آنجا که طبق فرض مشتقات مرتبه دوم تابع مورد نظر پیوسته اند، نتیجه می شود که ترتیب دیفرانسیل گیری اهمیتی نخواهد داشت (قضیه شوارتز).

10.2.مقادیر اکستریم (Extreme Values)



تابع $f(x,y) = x^2 + y^2$ ، دارای مقدار مینیمم 0 است ؛ این مقدار در (0,0) رخ میدهد که در آن نمودار دارای صفحه مماس افقی است .



به همین نحو تابع بالا دارای مقدار ماکزیمم 1 در (0,0) است . یافتن مقدار ماکزیمم و مینیمم توابع چند متغیره ، مانند توابع یک متغیره ، اسا بسیاری از کاربرد های حساب دیفرانسیل و انتگرال پیشرفته ر مسائلی است که در سایر نظامها ظاهر می شود . بحث را با مرور آنچه از حالت یک متغیره می دانیم آغاز می کنیم ، تابع $f(x)$ در نقطه a در قلمروش دارای مقدار ماکزیمم موضعی(یا مقدار مینیمم موضعی) است اگر به ازای هر x در قلمرو f که به قدر کافی نزدیک x باشد $f(x) \leq f(a)$ (یا $f(x) \geq f(a)$). هرگاه نامساوی مربوطه به ازای هر x در قلمرو f برقرار باشد ، آنگاه می گوییم f دارای مقدار ماکزیمم مطلق (یا مینیمم مطلق) در a است .به علاوه این مقدار اکستریم موضعی یا مطلق می توانند فقط در نقاطی از یکی از سه نوع زیر رخ دهند :

الف) نقاط بحرانی، که $f'(x) = 0$

ب)نقاط منفرد ، که $f'(x)$ موجود نیست ؛ یا

پ)نقاط انتهایی قلمرو f

برای توابع چند متغیره نیز وضعیت مشابهی برقرار است .

شرایط لازم برای مقادیر اکستریم

تابع $f(x, y)$ فقط وقتی می تواند در نقطه (a,b) از قلمرو خود مقدار اکستریم مطلق داشته باشد که (a, b) به یکی از صورت های زیر باشد :

الف) یک نقطه ی بحرانی f یعنی یک نقطه صادق در $\nabla f(a,b)=0$ ؛

ب)یک نقطه منفرد f یعنی نقطه ای که در آن $\nabla f(a,b)$ وجود ندارد ؛یا

پ)یک نقطه مرزی قلمرو f

شرایط کافی برای مقادیر اکستریم

هرگاه f یک تابع n متغیره پیوسته باشد که قلمروش یک مجموعه بسته و کراندار در$\mathbb{R}^n$است ،آنگاه برد f یک مجموعه کراندار از اعداد حقیقی است و نقاطی در قلمرو آن وجود دارند که در آنها f مقادیر ماکزیمم ومینیمم مطلق را می گیرد .

مثال) تابع $f(x,y)=x^2+y^2$ در (0,0) نقطه بحرانی دارد زیرا $\nabla f=2xi+2yj$ و هر دو مولفه ∇f در (0,0) صفر می شوند چون

$$\text{به‌ازای}(0,0)\neq(x,y)\text{،}f(x,y)>0=f(0,0)$$

f باید در این نقطه مقدار مینیمم (مطلق) 0 داشته باشد . اگر قلمرو f محدود نشود ، f مقدار ماکزیمم ندارد . به همین نحو ، $g(x,y)=1-x^2-y^2$ دارای مقدار ماکزیمم (مطلق) 1 در نقطه بحرانی (0,0) می باشد .

11.2.ضرایب لاگرانژ

یک روش یافتن مقادیر اکستریم $f(x,y)$ تحت قید تساوی $g(x,y)=0$بر قضیه زیر استوار است :

فرض کنیم f و g در مجاورت نقطه $P_0=(x_0,y_0)$ روی منحنی C به معادله $g(x,y)=0$ مشتقهای جزئی اول پیوسته داشته باشند . همچنین ، وقتی به نقاط روی C محدود شویم ،تابع $f(x,y)$ در P_0 مقدار ماکزیمم یا مینیمم موضعی داشته باشد و 1) P_0 یک نقطه ی انتهایی C نباشد و 2) $\nabla g(P_0)\neq 0$

در این صورت عددی مانند λ_0 هست به طوریکه (x_0,y_0,λ_0) یک نقطه بحرانی تابع لاگرانژین

$$L(x,y,\lambda)=f(x,y)+\lambda g(x,y)$$

می باشد .

این قضیه پیشنهاد می کند که برای یافتن نقاط روی منحنی $g(x,y)=0$ که $f(x,y)$ در آنها ماکزیمم یا مینیمم است باید نقاط بحرانی تابع لاگرانژین $L(x,y,\lambda)=f(x,y)+\lambda g(x,y)$ را جستجو نماییم . در هر نقطه بحرانی L باید داشته باشیم :

$$\begin{cases} 0=\frac{\partial L}{\partial x}=f_1(x,y)+\lambda g_1(x,y)\\ 0=\frac{\partial L}{\partial y}=f_2(x,y)+\lambda g_2(x,y)\end{cases}$$

یعنی ∇f موازی ∇g است .

$$\text{(معادله‌قید)}=\frac{\partial L}{\partial \lambda}=g(x,y)$$

اما فرض است که مسئله مقید دارای جواب است ، این قضیه وجود جواب را تضمین نمی کند ؛ تنها یک روش برای یافتن جواب که از قبل وجودش معلوم است به دست می دهد . معمولا باید پیش از به کار گیری این روش برای یافتن جواب از وجود آن مطمعن شوید .

منابع

1. حساب دیفرانسیل و انتگرال و هندسه تحلیلی نوشته ی آر.ای.آدامز
2. [فصل دوم کتاب Dive Into Deep Learning]
3. [جزوه ریاضیات برای یادگیری ماشین دانشگاه Berkely]