TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI TP.HCM
VIỆN CÔNG NGHỆ THÔNG TIN VÀ ĐIỆN, ĐIỆN TỬ

PHƯƠNG PHÁP TOÁN CHO MÁY HỌC

CHƯƠNG III: GIẢI TÍCH MA TRẬN

TS. Trần Thế Vinh

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Giải tích ma trận (Matrix Calculus) đóng vai trò quan trọng trong các phương pháp toán học cho máy học, đặc biệt khi xử lý các thuật toán tối ưu hóa như hồi quy tuyến tính, mạng neuron, hay học sâu. Nó cung cấp công cụ để tính đạo hàm của các hàm số phức tạp liên quan đến ma trận và vector, thường xuất hiện trong việc tối ưu hóa hàm mất mát (loss function).

Ứng dụng:

- Tối ưu hóa hàm mất mát (Loss Function Optimization)
- Mạng nơ-ron và học sâu (Neural Networks and Deep Learning)
- Phân tích thành phần chính (PCA)
- Học không giám sát (Unsupervised Learning)
- Xử lý dữ liệu và biểu diễn (Data Processing and Representation)
- Máy vector hỗ trợ (Support Vector Machines - SVM)

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Xấp xỉ hàm số một biến số

1. Đạo hàm cấp cao: Xét hàm số $y = f(x)$

- Đạo hàm (đạo hàm cấp 1) : $y' = f'(x) = \lim\limits_{t \to x} \frac{f(t)-f(x)}{t-x}$,
- Đạo hàm cấp n : $y^{(n)} = f^{(n)}(x) = \{f^{(n-1)}(x)\}'$ (quy ước: $f^{(0)}(x) = f(x)$)

Ví dụ 1. Với $f(x) = \sin x$: $f'(x) = (\sin x)' = \cos x = \sin\left(x + \frac{\pi}{2}\right)$;

$$f^{(2)}(x) = (\sin x)^{(2)} = \left(\sin(x + \frac{\pi}{2})\right)' = \sin\left(x + 2.\frac{\pi}{2}\right)$$

$$f^{(n)}(x) = (\sin x)^{(n)} = \sin\left(x + n.\frac{\pi}{2}\right)$$

2. Đạo hàm cấp cao của một số hàm cơ bản:

- $(\sin x)^{(n)} = \sin\left(x + n.\frac{\pi}{2}\right)$, $(\cos x)^{(n)} = \cos\left(x + n.\frac{\pi}{2}\right)$, $\forall n = 0,1,2,\ldots$
- $(x^k)^{(n)} = \begin{cases} k(k-1)\ldots(k-n-1).x^{k-n}, \text{nếu } n \le k \\ 0, \text{nếu } n > k \end{cases}$, $\forall n = 0,1,2,\ldots$
- $(a^{b.x})^{(n)} = (b.\ln a)^n.a^{b.x}$, $\left(\frac{1}{x+a}\right)^{(n)} = (-1)^n.\frac{n!}{(x+a)^{n+1}}$, $\forall n = 0,1,2,\ldots$
- $(\ln(x+a))^{(n)} = (-1)^{n-1}.\frac{(n-1)!}{(x+a)^n}, \forall n = 1,2,\ldots$

```python
# Install sympy package
%pip install sp

import sympy as sp

# Khai báo biến
x = sp.Symbol('x')

# Nhập hàm từ bàn phím
expr_str = input("Nhập hàm cần tính đạo hàm: ")

# Chuyển chuỗi thành biểu thức SymPy
f = sp.sympify(expr_str)

# Nhập cấp đạo hàm
n = int(input("Nhập cấp đạo hàm n: "))

# Tính đạo hàm cấp n
df_n = sp.diff(f, x, n)

# Hiển thị kết quả
print(f"Đạo hàm cấp {n} của f(x) là: {df_n}")
```

Ứng dụng:

Trong bài toán tối ưu, ta thường cần ước lượng hàm số xung quanh một điểm nào đó. Việc sử dụng đạo hàm giúp ta tìm xấp xỉ tuyến tính hoặc bậc cao của hàm số.
Ví dụ1:
Khi tìm nghiệm của một phương trình trong Gradient Descent, ta có thể sử dụng **đạo hàm bậc hai (Hessian matrix)** để xác định tốc độ hội tụ hoặc cải thiện tốc độ cập nhật.

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Xấp xỉ hàm số một biến số

3. Công thức Taylor: Tồn tại điểm c bao hàm giữa x và t sao cho:

$$f(x) = f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}.(x-t)^2 + \cdots + \frac{f^{(n)}(t)}{n!}.(x-t)^n + \frac{f^{(n+1)}(c)}{(n+1)!}.(x-t)^{n+1} \quad (3.1)$$

Chú ý:

(3.1) gọi là *công thức khai triển Taylor của hàm $f(x)$ tại t đến cấp n*

Biểu thức $R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}.(x-t)^{n+1}$ gọi là *phần dư thứ n của khai triển Taylor*

Với t = 0, (3.1) trở thành:

$$f(x) = f(0) + \frac{f'(0)}{1!}.x + \frac{f''(0)}{2!}.x^2 + \cdots + \frac{f^{(n)}(0)}{n!}.x^n + \frac{f^{(n+1)}(c)}{(n+1)!}.x^{n+1} \quad (3.2)$$

(3.2) gọi là công thức khai triển Mac Laurin.

```python
import sympy as sp

# Khai báo biến x
x = sp.Symbol('x')

# Nhập hàm từ bàn phím
expr_str = input("Nhập hàm cần xấp xỉ chuỗi Taylor: ")
f = sp.sympify(expr_str)

# Nhập điểm mở rộng Taylor
t = float(input("Nhập điểm t (mở rộng Taylor): "))

# Nhập cấp xấp xỉ Taylor
n = int(input("Nhập bậc n của chuỗi Taylor: "))

# Khai báo chuỗi Taylor
taylor_series = sum(sp.diff(f, x, i).subs(x, t) / sp.factorial(i) * (x - t) ** i for i in range(n + 1))

# Hiển thị kết quả
print(f"Chuỗi Taylor cấp {n} tại x = {t} là: {taylor_series}")
```

4. Xấp xỉ hàm số bởi đa thức

Khi hàm f(x) có các đạo hàm bị chặn thì $R_n(x) = o\left((x-t)^{n+1}\right)$ $(n \to +\infty)$, tức là phần dư sẽ dần về 0 (vô cùng bé) khi $n \to +\infty$.

Khi đó từ (3.1) có công thức xấp xỉ:

$$f(x) \approx f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}.(x-t)^2 + \cdots + \frac{f^{(n)}(t)}{n!}.(x-t)^n \quad (3.3)$$

Với n = 1, khi x khá gần t, ta có công thức xấp xỷ:

$$f(x) \approx f(t) + \frac{f'(t)}{1!}(x-t), \text{ hay: } f'(t) \approx \frac{f(x)-f(t)}{x-t} \quad (3.4)$$

Khi đó (3.2) trở thành:

$$f(x) \approx f(0) + \frac{f'(0)}{1!}.x + \frac{f''(0)}{2!}.x^2 + \cdots + \frac{f^{(n)}(0)}{n!}.x^n \quad (3.5)$$

Các công thức (3.3), (3.5) là các công thức tính xấp xỷ giá trị hàm $f(x)$ bởi đa thức.

Ví dụ 2. Khi n đủ lớn, từ ((3.5) ta có: $e^x \approx 1 + x + \frac{1}{2!}.x^2 + \cdots + \frac{1}{n!}.x^n$

Đặc biệt: $e \approx 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!}$

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Ý nghĩa của khai triển Taylor:

Khai triển Taylor giúp mô phỏng, dự đoán, tối ưu hoá và giải quyết nhiều vấn đề phức tạp theo một cách đơn giản hơn.

Ví dụ: Trong các hệ thống nhận diện giọng nói như Siri, Google Assistant sử dụng mô hình xử lý tín hiệu âm thanh, trong đó khai triển Taylor giúp xấp xỉ biến đổi tần số và nhận dạng giọng nói chính xác hơn như cải thiện khả năng nhận diện giọng nói và chuyển đổi văn bản thành lời nói, giúp tối ưu hoá thuật toán dịch ngôn ngữ tự động.

Ứng dụng:

1. Trong tối ưu hoá (Gradient Descent):

Gradient Descent là một phương pháp phổ biến để tìm cực tiểu của hàm mất mát (loss function). Để hiểu các hàm mất mát thay đổi khi cập nhập tham số, ta có thể dùng khai triển Taylor bậc hai:

$$J(\theta + \Delta\theta) \approx J(\theta) + \nabla J(\theta).\Delta\theta + \frac{1}{2}\Delta\theta^T HJ(\theta)\Delta\theta$$

Trong đó:

- $J(\theta)$ là hàm mất mát.
- $\nabla J(\theta)$ là gradient (đạo hàm bậc nhất)
- $HJ(\theta)$ là ma trận Hessian (đạo hàm bậc 2).

Dựa vào khai triển Taylor:

- Gradient Descent chỉ dùng thành phần bậc 1 ($\nabla J(\theta)$) để cập nhập tham số.
- Newton's Method dùng thêm ma trận Hessian ($HJ(\theta)$) để tìm cực tiểu nhanh hơn.

Khai triển Taylor giúp hiểu sự thay đổi của hàm mất mát quanh một điểm và cho phép thiết kế thuật toán tối ưu hiệu quả hơn.

2. Hồi quy tuyến tính:

$y = f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_n x^n$ Hàm này chính là một đa thức Taylor bậc n, cho phép xấp xỉ các hàm phi tuyến bằng mô hình

3. Phát triển mạng Neuron:

Hàm kích hoạt (activation functions) phổ biến trong mạng neuron như ReLU, Sigmod, Tanh có thể được khai triển Taylor để phân tích sự hội tụ và độ mượt của mạng.

4. Bayesian Machine Learning: Trong thống kê Bayes, hàm hậu nghiệm (posterior) thường rất phức tạp. Khai triển Taylor giúp xấp xỉ nó bằng một mô hình đơn giản hơn, như trong Laplace Approximation:

$$logP(\theta|X) \approx logP(\theta^*) + \frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*)$$

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Hàm nhiều biến số và đạo hàm riêng của hàm nhiều biến số

1. Khái niệm về hàm nhiều biến số

Trong thực tế, thường gặp một đại lượng Y phụ thuộc vào nhiều đại lượng khác nhau $X_1, X_2, \dots, X_n$ mà với một giá trị đã cho của $X_1, X_2, \dots, X_n$ có một và chỉ một giá trị tương ứng của Y. Ta nói Y là một hàm của các biến $X_1, X_2, \dots, X_n$ (hàm n biến).

Theo ngôn ngữ toán học: một hàm n biến số là một ánh xạ

$$f: D \ni x = (x_1, x_2, \dots, x_n) \to y = f(x_1, x_2, \dots, x_n) \in \mathbb{R} \ (D \subset \mathbb{R}^n, \text{D gọi là tập xác định})$$

Ví dụ 3: Với hình chữ nhật có số đo chiều rộng là x, số đo chiều dài là y thì khi chiều rộng và chiều dài thay đổi, diện tích S = x . y là hàm của 2 biến x, y. Với hình hộp chữ nhật có các kích thước x, y, z thì khi x, y, z thay đổi thể tích V = x . y . z là hàm của 3 biến x, y, z.

Ví dụ 4: Với ma trận vuông $A = (x_{ij})_{n\times n}$ thì

- $\det(A)$ là hàm của n^2 biến $x_{ij}, i = \overline{1,n}, j = \overline{1,n}$
- $\text{trace}(A) = x_{11} + x_{22} + \dots + x_{nn}$ là hàm của n biến $x_{11}, x_{22}, \dots, x_{nn}$
- Các chuẩn, các khoảng cách trong không gian R^n là các hàm n biến.

2. Đạo hàm riêng của hàm nhiều biến. Cho hàm n biến: $u = f(x_1, x_2, \dots, x_n)$

- Nếu chỉ cho x_i thay đổi còn các biến khác cố định (hằng số) thì u chỉ là hàm của một biến x_i và đạo hàm của hàm u với biến là x_i được ký hiệu là $\frac{\partial u}{\partial x_i}$ (hoặc u'_{x_i}) và gọi là đạo hàm riêng (cấp 1) của hàm $u = f(x_1, x_2, \dots, x_n)$ theo biến x_i. Khi đó $\frac{\partial u}{\partial x_i}$ cũng là hàm n biến.
- Đạo hàm riêng của $\frac{\partial u}{\partial x_i}$ theo biến x_j là $\frac{\partial}{\partial x_i}\left(\frac{\partial u}{\partial x_i}\right)$ được gọi là đạo hàm riêng cấp 2 của hàm u lần lượt theo các biến x_i

```python
import numpy as np
import sympy as sp
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

# Nhập số biến của hàm số
n = int(input("Nhập số biến của hàm số: "))

# Tạo danh sách biến
variables = sp.symbols(f'x1:{n+1}')

# Nhập hàm dưới dạng chuỗi
expr_str = input(f"Nhập hàm f({', '.join(map(str, variables))}): ")
expr = sp.sympify(expr_str)  # Chuyển chuỗi thành biểu thức SymPy

# Tính đạo hàm riêng theo từng biến
gradient = [sp.diff(expr, var) for var in variables]

# Giải hệ phương trình ∇f = 0 để tìm điểm dừng
critical_points = sp.solve(gradient, variables, dict=True)

print("\nĐạo hàm riêng theo từng biến:")
for var, grad in zip(variables, gradient):
    print(f"∂f/∂{var} = {grad}")

print("\nĐiểm dừng:")
if not critical_points:
    print("Không có điểm dừng.")
else:
    for i, point in enumerate(critical_points, 1):
        print(f"Điểm {i}: {point}")

    # Xác định loại điểm cực trị
    H = sp.hessian(expr, variables)  # Ma trận Hessian
    print("\nPhân loại điểm dừng:")

    for i, point in enumerate(critical_points, 1):
        H_eval = H.subs(point)  # Thay giá trị điểm dừng vào Hessian
        H_np = np.array(H_eval).astype(np.float64)  # Chuyển về ma trận số thực

        det_H = np.linalg.det(H_np)  # Định thức của Hessian

        if det_H > 0:
            if H_np[0, 0] > 0:
                point_type = "Cực tiểu"
            else:
                point_type = "Cực đại"
        elif det_H < 0:
            point_type = "Điểm yên ngựa"
        else:
            point_type = "Không xác định"

        print(f"Điểm {i}: {point}, Loại: {point_type}")
```

Ví dụ 5: Tính các đạo hàm riêng cấp 2 của hàm $u = 2x.e^{-yz} - x^2.z$, ta có:

+ Các đạo hàm riêng cấp 1: $\frac{\partial u}{\partial x} = 2e^{-yz} - 2xz$; $\frac{\partial u}{\partial y} = -2xze^{-yz}$; $\frac{\partial u}{\partial z} = -2xye^{-yz} - x^2$ + Các đạo hàm riêng cấp 2: $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial y}(2e^{-yz} - 2xz) = -2z.e^{-yz}$; $\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}(2e^{-yz} - 2xz) = -2z, \dots$

3. **Cực trị của hàm nhiều biến**

Cho hàm n biến: $u = f(x_1, x_2, \dots, x_n)$ có tập xác định D. Nói rằng hàm u đạt cực đại (cực tiểu) tại $a \in D$, nếu $\exists$ hình cầu mở $B(r, a) \subset D$ sao cho: $f(x) \leq f(a)$ $\left(f(x) \geq f(a)\right), \forall x \in B(r, a)$

a. Điều kiện cần của cực trị: Giả sử hàm u = f(x) đạt cực trị tại a. khi đó nếu f(x) có tất cả các đạo hàm riêng tại a thì $\frac{\partial u}{\partial x_i} = 0, \forall i = 1,2, \dots, n$

b. Điều kiện đủ của cực trị: Với hàm $u = f(x_1, x_2, \dots, x_n)$ có các đạo hàm riêng cấp 2, ma trận: $H_u(a) = \left(\frac{\partial^2 u(a)}{\partial x_i \partial x_j}\right)_{n \times n}$ gọi là ma trận Hessian của hàm u tại điểm a

Mệnh đề: Giả sử a là nghiệm của hệ ph.trình : $\frac{\partial u}{\partial x_i} = 0, \forall i = \overline{1, n}$. Khi đó gọi a là điểm dừng của u và:

(1) Nếu ma trận Hessian $H_u(a)$ là ma trận xác định dương thì u đạt cực tiểu tại a

(2) Nếu ma trận Hessian $H_u(a)$ bán xác định dương, hoặc bán xác định âm nhưng $\exists b \in D, b \neq 0$ sao cho: $(b).H_u(a).[b] = 0$, thì chưa thể kết luận cực trị tại a

(3) Nếu $\exists b, c \in D$ mà: $(b).H_u(a).[b] > 0$ và: $(c).H_u(a).[c] < 0$ thì u không đạt cực trị tại a.

Hệ quả (cực trị hàm 2 biến): Cho hàm 2 biến $z = f(x, y)$ có các đạo hàm riêng cấp 2.

Đặt: $A = \frac{\partial^2 u}{\partial x^2}$, $B = \frac{\partial^2 u}{\partial x \partial y}$, $C = \frac{\partial^2 u}{\partial y^2}$, $\Delta = A.C - B^2$

Xét hệ phương trình: $\begin{cases} \frac{\partial u}{\partial x} = 0 \\ \frac{\partial u}{\partial \dots} \end{cases}$ (*)

```python
# Vẽ đồ thị nếu có thể
if n == 1:
    x = variables[0]
    f_lambdified = sp.lambdify(x, expr, 'numpy')

    x_vals = np.linspace(-10, 10, 400)
    y_vals = f_lambdified(x_vals)

    plt.plot(x_vals, y_vals, label=f'f({x}) = {expr}')

    # Vẽ điểm cực trị
    for point in critical_points:
        x_crit = float(point[x])
        y_crit = f_lambdified(x_crit)
        plt.scatter(x_crit, y_crit, color='red', marker='o', label="Cực trị")

    plt.xlabel(str(x))
    plt.ylabel('f(x)')
    plt.legend()
    plt.title("Biểu đồ 2D của hàm số")
    plt.grid()
    plt.show()

elif n == 2:
    x, y = variables[:2]
    f_lambdified = sp.lambdify((x, y), expr, 'numpy')

    x_vals = np.linspace(-10, 10, 50)
    y_vals = np.linspace(-10, 10, 50)
    X, Y = np.meshgrid(x_vals, y_vals)
    Z = f_lambdified(X, Y)

    fig = plt.figure(figsize=(8, 6))
    ax = fig.add_subplot(111, projection='3d')
    ax.plot_surface(X, Y, Z, cmap='viridis', alpha=0.7)

    # Vẽ điểm cực trị
    for point in critical_points:
        x_crit, y_crit = float(point[x]), float(point[y])
        z_crit = f_lambdified(x_crit, y_crit)
        ax.scatter(x_crit, y_crit, z_crit, color='red', s=100, marker='o')

    ax.set_xlabel(str(x))
    ax.set_ylabel(str(y))
    ax.set_zlabel('f(x, y)')
    ax.set_title("Biểu đồ 3D của hàm số với điểm cực trị")
    plt.show()

else:
    print("\nKhông thể trực quan hóa hàm số với nhiều hơn 2 biến.")
```

Khi đó:
(a) Nếu hệ (*) vô nghiệm thì hàm u không có cực trị
(b) Khi (*) có nghiệm a:
- Nếu tại a có: $\Delta(a) < 0$, thì u không đạt cực trị tại a
- Nếu tại a có: $\Delta(a) > 0$, thì u đạt cực trị tại a, cụ thể: với A(a) > 0 thì u đạt cực tiểu tại b, với A(a) < 0 thì u đạt cực đại tại a.
- Nếu $\Delta(a) = 0$ thì chưa thể kết luận về cực trị tại a

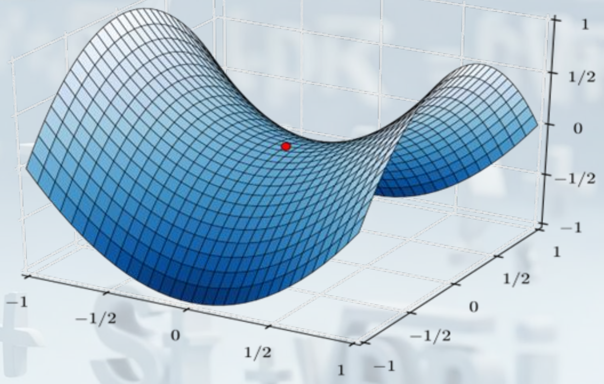
Ví dụ 5: Tìm cực trị của hàm $u = x^2 + y^2 + z^2 - 2z$

Dễ thấy điểm dừng là a = (0, 0, 1) duy nhất và ma trận Hessian $H_u(a) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ là ma trận xác định dương. Vậy hàm u đạt cực trị duy nhất (và là cực tiểu) tại a = (0, 0, 1) và $u_{\text{cực tiểu}} = -2$

Ví dụ 6: Tìm cực trị của hàm $z = x^2 - y^2$(hàm Hyperbolic - Paraboloid)
(Đồ thi của hàm này còn được gọi là mặt yên ngựa (hình bên))
- Dễ thấy điểm dừng duy nhất của hàm này là a = (0, 0)
- Tính $A = \frac{\partial^2 z}{\partial x^2} = 2$, $B = \frac{\partial^2 z}{\partial x \partial y} = 0$,

$$C = \frac{\partial^2 z}{\partial y^2} = -2, \Delta = A.C - B^2 = -4 < 0$$

Vậy hàm này không có cực trị. Điểm dừng (0, 0) không phải là điểm cực trị. Điểm này trên mặt yên ngựa được gọi là điểm yên ngựa hay điểm minimax



GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Đạo hàm của hàm có giá trị là số vô hướng

1. Đạo hàm của hàm $f: \mathbb{R}^n \ni x = (x_1, x_2, \dots, x_n) \to f(x) \in \mathbb{R}$

Đạo hàm bậc nhất (Gradient) của f, kí hiệu $\boldsymbol{\nabla f(x)} \triangleq \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{n\text{x}1}$ (3.6)

Đạo hàm bậc 2 (Hessian), kí hiệu $\boldsymbol{\nabla^2 f(x)} \triangleq \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n\times n}$ (3.7)

Lưu ý:

(1) Ánh xạ $f: \mathbb{R}^n \to \mathbb{R}$ có đạo hàm bậc 1 $\boldsymbol{\nabla f(x)}$ là một ma trận trong $\mathbb{R}^{n\times 1}$, và có đạo hàm bậc hai $\boldsymbol{\nabla^2 f(x)}$ là một ma trận trong $\mathbb{R}^{n\times n}$

(2) Đạo hàm cấp 2: $\nabla_x^2 f(x)$ của hàm số nhiều biến số còn gọi là Hessian là ma trận đối xứng.

Ví dụ 7. Cho hàm $f(x) = f(x_1, x_2) = x_1^2 - 2x_1x_2 + 2x_2^2$

Có: $\nabla_x f(x) = \begin{bmatrix} 2x_1 - 2x_2 \\ 4x_2 - 2x_1 \end{bmatrix}$; $\nabla_x^2 f(x) = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$

```python
import sympy as sp

def compute_derivatives():
    # Nhập số biến
    n = int(input("Nhập số biến n: "))

    # Khởi tạo biến x_1, x_2, ..., x_n
    x = sp.symbols(f'x1:{n+1}')

    # Nhập hàm số dưới dạng chuỗi và chuyển thành biểu thức SymPy
    f_expr = input("Nhập hàm số f(x1, x2, ..., xn): ")
    f = sp.sympify(f_expr)

    # Tính Gradient
    grad = [sp.diff(f, xi) for xi in x]
    grad_vector = sp.Matrix(grad)

    # Tính Hessian
    hessian = sp.Matrix([[sp.diff(grad[i], x[j]) for j in range(n)] for i in range(n)])

    print("\nGradient (∇f):")
    sp.pprint(grad_vector)

    print("\nHessian (∇²f):")
    sp.pprint(hessian)

compute_derivatives()
```

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

2. Đạo hàm của hàm có đầu vào là ma trận (biến ma trận) $f: \mathbb{R}^{m\times n} \ni X = \left(x_{ij}\right)_{m\times n} \to f(X) \in \mathbb{R}$

$$\text{Đạo hàm (bậc 1), kí hiệu: } \nabla_X f(X) \triangleq \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f(X)}{\partial x_{m1}} & \frac{\partial f(X)}{\partial x_{m2}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{m\times n} \tag{3.8}$$

Lưu ý: Đạo hàm $\nabla_X f(X)$ của hàm) $f: \mathbb{R}^{m\times n} \to \mathbb{R}$ là một ma trận trong $\mathbb{R}^{m\times n}$

Ví dụ 8. Xét hàm: $f: \mathbb{R}^{2\times 3} \ni X = \begin{bmatrix} x & y & z \\ u & v & t \end{bmatrix} \to f(X) = \|X\|_F^2 \in \mathbb{R}$

Có: $f(X) = \|X\|_F^2 = x^2 + y^2 + z^2 + u^2 + v^2 + t^2$

Vậy đạo hàm: $\nabla_X f(X) = \begin{bmatrix} 2x & 2y & 2z \\ 2u & 2v & 2t \end{bmatrix} = 2.X$

Ứng dụng:

✓ Gradient Descent (Hạ dốc theo gradient):

- Trong huấn luyện mô hình máy học, ta cần tối ưu một hàm mất mát $J(\theta)$ theo tham số θ.
- Gradient của $J(\theta)$ giúp ta cập nhập tham số theo quy tắc:

$$\theta = \theta - \alpha \nabla J(\theta)$$

Trong đó α là tốc độ học (learning rate).

✓ Các thuật toán tối ưu như Adam, RMSprop đều dựa trên đạo hàm riêng để điều chỉnh tốc độ học.

```python
import sympy as sp

def main():
    # Bước 1: Nhập số dòng (m) và số cột (n)
    m = int(input("Nhập số dòng m: "))
    n = int(input("Nhập số cột n: "))

    # Bước 2: Nhập tên biến cho từng phần tử ma trận X
    # Tạo một cấu trúc lưu tên biến cho mỗi vị trí (i, j)
    variable_names = []
    for i in range(m):
        row_names = []
        for j in range(n):
            var_name = input(f"Nhập tên biến cho phần tử X[{i},{j}]: ")
            row_names.append(var_name)
        variable_names.append(row_names)

    # Tạo dict để ánh xạ "tên biến" -> "biến Symbol trong Sympy"
    var_dict = {}
    for i in range(m):
        for j in range(n):
            name = variable_names[i][j]
            var_dict[name] = sp.Symbol(name, real=True)

    # Xây dựng ma trận Sympy X từ các biến Symbol
    x_symbols = []
    for i in range(m):
        row = []
        for j in range(n):
            row.append(var_dict[variable_names[i][j]])
        x_symbols.append(row)
    X = sp.Matrix(x_symbols)

    # Bước 3: Nhập biểu thức hàm f(X)
    print("\nLưu ý: Hãy dùng đúng tên biến bạn vừa đặt cho các phần tử.\n")
    f_str = input("Nhập biểu thức hàm f(X): ")

    # Dùng sympify với var_dict để chuyển f_str thành biểu thức Sympy
    f_expr = sp.sympify(f_str, locals=var_dict)

    # Bước 4: Tính ma trận đạo hàm (gradient) ∂f/∂X
    gradient_matrix = sp.zeros(m, n)
    for i in range(m):
        for j in range(n):
            gradient_matrix[i, j] = sp.diff(f_expr, X[i, j])

    # Bước 5: In kết quả
    print("\n=== Kết quả ===")
    print("Ma trận X:")
    sp.pprint(X)

    print("\nHàm f(X):")
    sp.pprint(f_expr)

    print("\nĐạo hàm (gradient) ∂f/∂X:")
    sp.pprint(gradient_matrix)

if __name__ == "__main__":
    main()
```

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Đạo hàm của hàm có giá trị vector (đầu ra là vector):

f: $\mathbb{R}^m \to \mathbb{R}^n$:

$$f(x) = f(x_1, x_2, \dots, x_m) = (f_1(x), f_2(x), \dots, f_n(x))^T$$

Đạo hàm của f, kí hiệu:

$$\boldsymbol{\nabla f(x)} \triangleq \left[\frac{\partial f_j(x)}{\partial x_i}\right]_{m\times n} \quad \text{(i: chỉ số hàng, j: chỉ số cột) (Ma trận Jacobian)} \qquad \textbf{(3.9)}$$

Lưu ý:

(1) Đạo hàm $\nabla f(x)$ của hàm f: $\mathbb{R}^m \to \mathbb{R}^n$ là một ma trận trong $\mathbb{R}^{m\times n}$

(2) Đối với f: $\mathbb{R} \to \mathbb{R}^n$: $f(x) = (f_1(x), f_2(x), \dots, f_n(x))^T$, người ta còn quan tâm đến đạo hàm bậc 2:

$$\boldsymbol{\nabla^2 f(x)} \triangleq \left(\frac{d^2 f_1(x)}{dx^2}, \frac{d^2 f_2(x)}{dx^2}, \dots \frac{d^2 f_n(x)}{dx^2}\right) \qquad \textbf{(3.10)}$$

(1) f: $\mathbb{R}^m \to \mathbb{R}^n$, $\nabla f(x) \triangleq \left[\frac{\partial f_j(x)}{\partial x_i}\right]_{m\times n} = ([\nabla f_1(x)]\ \ [\nabla f_2(x)] \dots [\nabla f_n(x)])$

Ví dụ 9. Xét $f(x, y) = (x + y, x.y, x - y) \Rightarrow \nabla f(x) = \begin{bmatrix} 1 & y & 1 \\ 1 & x & -1 \end{bmatrix}$

```python
import sympy as sp

def main():
    # Bước 1: Nhập số chiều đầu vào m
    m = int(input("Nhập số chiều đầu vào m: "))

    # Tạo các biến x1, x2, ..., x_m
    x_symbols = sp.symbols('x1:' + str(m + 1), real=True)

    # Bước 2: Nhập số chiều đầu ra n
    n = int(input("Nhập số chiều đầu ra n: "))

    print(f"\nCó {m} biến: {x_symbols}")
    print(f"Cần nhập {n} thành phần f_1, f_2, ..., f_{n} (mỗi thành phần là 1 hàm R^m -> R).")
    print("Ví dụ: x1 + x2, x1*x2, x1 - x2,... \n")

    # Bước 3: Nhập các thành phần f_j
    f_list = []
    for j in range(n):
        f_str = input(f"Nhập biểu thức cho f_{j + 1}(x1, x2, ..., x{m}): ")
        # Ensure the expression uses the correct variable names
        f_expr = sp.sympify(f_str, locals={f'x{i+1}': x_symbols[i] for i in range(m)})
        f_list.append(f_expr)

    # Xây dựng vector-hàm f(x)
    f_vector = sp.Matrix(f_list)

    # Bước 4: Tính ma trận Jacobian (kích thước n x m)
    jacobian_matrix = sp.zeros(n, m)
    for i in range(n):
        for j in range(m):
            jacobian_matrix[i, j] = sp.diff(f_list[i], x_symbols[j])

    # Bước 5: In kết quả
    print("\n--- Vector-hàm f(x) ---")
    sp.pprint(f_vector)

    print("\n--- Ma trận Jacobian (n x m) ---")
    sp.pprint(jacobian_matrix.T)

if __name__ == "__main__":
    main()
```

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Các tính chất quan trọng của đạo hàm

1. Đạo hàm của tích (Qui tắc tích: Product Rule)**:** Giả sử f(X), g(X) có cùng đầu vào là biến ma trận và có cùng chiều. Khi đó:

$$\nabla\left(\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\right) = \left(\nabla \boldsymbol{f}(\boldsymbol{X})\right).\mathrm{g}(\mathrm{X}) + \left(\nabla \mathrm{g}(\boldsymbol{X})\right).\mathrm{f}(\mathrm{X}) \qquad \textbf{(3.11)}$$

Lưu ý:

(1) Trong (3.11) vector được viết theo cột.

(2) Đối với các hàm một biến số thông thường: u(x), v(x), có :

$$\{u(x).v(x)\}' = u'(x).v(x) + v'(x).u(x) \qquad \textbf{(3.12)}$$

Trong (3.12) các tích có thể giao hoán, nhưng trong (3.11) thì không thể.

Ví dụ 10. Tìm $\nabla_{\mathrm{X}}\{\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\}$: $\mathrm{f(x,y) = (x, x+y, y)}$; $\mathrm{g(x,y) = (x, x-y, y)}, \forall \mathrm{X} = (x,y)^{\mathrm{T}} \in \mathbb{R}^2$

Dùng quy tắc tích, có: $\nabla \boldsymbol{f}(\boldsymbol{X}) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \nabla \mathrm{g}(\boldsymbol{X}) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$

$$\nabla\left(\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\right) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.\begin{bmatrix} \mathrm{x} \\ \mathrm{x-y} \\ \mathrm{y} \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.\begin{bmatrix} \mathrm{x} \\ \mathrm{x+y} \\ \mathrm{y} \end{bmatrix} = \begin{bmatrix} 4\mathrm{x} \\ 0 \end{bmatrix}$$

Tính trực tiếp: $\mathrm{f^T.g}\colon \mathbb{R}^2 \to \mathbb{R}$, $\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X}) = [\mathrm{x} \;\; \mathrm{x+y} \;\; \mathrm{y}].\begin{bmatrix} \mathrm{x} \\ \mathrm{x-y} \\ \mathrm{y} \end{bmatrix} = 2\mathrm{x}^2 \Rightarrow \nabla\left(\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\right) = \begin{bmatrix} 4\mathrm{x} \\ 0 \end{bmatrix}$

2. Quy tắc chuỗi (Chain Rule): Nếu $\mathrm{u(X) = g(f(X))}$ thì:

$$\nabla_X \mathrm{u}(\boldsymbol{X}) = \nabla_X \boldsymbol{g}\left(\boldsymbol{f}(\boldsymbol{X})\right) = \nabla_X \boldsymbol{f}(\boldsymbol{X}).\nabla_f \boldsymbol{g}(\boldsymbol{f}) \qquad \textbf{(3.13)}$$

Chú ý: (3.13) tương tự như quy tắc đạo hàm hàm số hợp một biến u(x) = g(f(x)):

u'(x) =(g(f(x))' = g'(f).f'(x) = f'(x).g'(f)

tuy nhiên trong vế phải (3.13) không thể thay đổi thứ tự nhân 2 ma trận

Ví dụ 11.

Cho f(x, y, z) = (y, z, x), g(x, y, z) = (x + z, y – z). Tìm $\nabla_X \boldsymbol{g}\left(\boldsymbol{f}(\boldsymbol{X})\right)$

Giải: Có $\mathrm{f}\colon \mathbb{R}^3 \to \mathbb{R}^3$, $\mathrm{g}\colon \mathbb{R}^3 \to \mathbb{R}^2$; $\mathrm{u = g \circ f}\colon \mathbb{R}^3 \to \mathbb{R}^2$

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Dùng qui tắc chuỗi: có $\boldsymbol{\nabla_X f(X)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \boldsymbol{g(f)} = \mathrm{g}(\mathrm{f}_1, \mathrm{f}_2, \mathrm{f}_3) = (\mathrm{f}_1 + \mathrm{f}_3, \mathrm{f}_2 - \mathrm{f}_3),$

$$\boldsymbol{\nabla_f g(f)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow \boldsymbol{\nabla_X g\big(f(X)\big)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} . \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Tính trực tiếp: Có $\boldsymbol{g\big(f(X)\big)} = \boldsymbol{g}(\mathrm{y}, \mathrm{z}, \mathrm{x}) = (\mathrm{y} + \mathrm{x}, \mathrm{z} - \mathrm{x}) \Rightarrow \boldsymbol{\nabla_X g\big(f(X)\big)} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

Đạo hàm của một số hàm thường gặp.

1. **Hàm:** f: $\mathbb{R}^n \to \mathbb{R}$: $\boldsymbol{f(x) = a^T . x}$, $\boldsymbol{x} \in \mathbb{R}^n$, a là vector cho trước $\in \mathbb{R}^n$

   Có: $\boldsymbol{f(x) = a^T . x = \sum_{i=1}^n a_i x_i \Rightarrow \nabla f(x) = (a_1\ a_2 \ldots a_n)^T = a}$

   Nhận xét: Vì $\mathrm{x}^T.\mathrm{a} = \boldsymbol{a^T . x}$, nên: $\boldsymbol{\nabla}(\mathrm{x}^T.\mathrm{a}) = \boldsymbol{\nabla}(\boldsymbol{a^T.x}) = \mathrm{a}$

2. **Hàm:** $\boldsymbol{f(x)} = \mathrm{A}.\boldsymbol{x}$, $\mathrm{x} \in \mathbb{R}^n (\mathrm{A} = (\mathrm{a_{ij}})_{m \times n}$ cho trước)

Ta có: $\mathrm{f}: \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{f(x)} = \left[\sum_{j=1}^n \mathrm{a_{1j}.x_j} \ \ \sum_{j=1}^n \mathrm{a_{2j}.x_j} \ \ldots . \sum_{j=1}^n \mathrm{a_{mj}.x_j}\right]^T$

$$\boldsymbol{\nabla f(x)} = [\nabla \mathrm{f_1(x)}\ \ \nabla \mathrm{f_2(x)}\ \ldots\ \nabla \mathrm{f_m(x)}] = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} = \mathrm{A^T}$$

Đặc biệt: $\boldsymbol{\nabla}(\boldsymbol{x}) = \boldsymbol{\nabla}(\mathrm{I_n}.\boldsymbol{x}) = \mathrm{I_n}$

3. **Hàm:** $\boldsymbol{f(x)} = \mathrm{x^T.A.}\boldsymbol{x}$, $\mathrm{x} \in \mathbb{R}^n (\mathrm{A} = (\mathrm{a_{ij}})_{n \times n}$ cho trước)

- Tìm theo quy tắc tích: $\boldsymbol{\nabla f(x)} = \boldsymbol{\nabla}\big(\mathrm{x^T}.(\mathrm{A}.\boldsymbol{x})\big) = \boldsymbol{\nabla}(\mathrm{x}).(\mathrm{Ax}) + \boldsymbol{\nabla}(\mathrm{Ax}).\mathrm{x} = \mathrm{Ax} + \mathrm{A^T.x} = \boldsymbol{(A + A^T).x}$
- Tìm trực tiếp: Có $\mathrm{f}: \mathbb{R}^n \to \mathbb{R}$: $\boldsymbol{f(x)} = \mathrm{x^T.A.}\boldsymbol{x} = \boldsymbol{\sum_{i=1}^n \sum_{j=1}^n a_{ij}.x_i.x_j}$

$$\boldsymbol{\nabla f(x)} = \left(\sum_{i=1}^n (\mathrm{a_{i1} + a_{1i}})\mathrm{x_i} \quad \sum_{i=1}^n (\mathrm{a_{i2} + a_{2i}})\mathrm{x_i} \quad \cdots \quad \sum_{i=1}^n (\mathrm{a_{in} + a_{ni}})\mathrm{x_i}\right)^T = \boldsymbol{(A + A^T).x}$$

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

4. **Hàm** $\boldsymbol{f}(\boldsymbol{x}) = \|\boldsymbol{A}.\boldsymbol{x} - \boldsymbol{b}\|_2^2 \ \left(\mathrm{A} = (\mathrm{a_{ij}})_{\mathrm{m\times n}}, \mathrm{b} \in \mathbb{R}^{\mathrm{m}}, \mathrm{x} \in \mathbb{R}^{\mathrm{n}}\right)$

Có: $\boldsymbol{f}(\boldsymbol{x}) = \mathrm{g}(\mathrm{h}(\mathrm{x}))(\mathrm{h}(\mathrm{x}) = \mathrm{Ax} - \mathrm{b}, \mathrm{g}(\mathrm{x}) = \|\mathrm{x}\|_2^2)$. Từ các ví dụ trên và quy tắc chuỗi

$$\nabla f(x) = \nabla g(h(x)) = (\nabla h(x))^T . \nabla_h g(h) = A^T.2.(A.x - b) = 2.A^T.(A.x - b)$$

5. **Hàm:** $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}.\boldsymbol{x}.\boldsymbol{x}^{\mathrm{T}}.\boldsymbol{b}, \ \mathrm{x} \in \mathbb{R}^{\boldsymbol{n}}$ ($\boldsymbol{a}.\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$cho trước)

Có: $f(x) = (a^T.x).(x^T.b) = (x^T a)^T.(x^T.b)$

$$\nabla f(x) = \nabla\{(x^T a)^T.(x^T.b)\} = \nabla(x^T a).(x^T.b) + \nabla(x^T.b).\ (x^T a)$$
$$= a.(x^T.b) + b.(x^T.a) = a.x^T b + b.a^T.x = a.b^T x + b.a^T.x = (a.b^T + b.a^T).x$$

6. **Hàm** $\boldsymbol{f}(\mathrm{X}) = \boldsymbol{trace}(\boldsymbol{A}.\boldsymbol{X}) \left(\mathrm{A} = (\boldsymbol{a_{ij}})_{\boldsymbol{n\times m}} \in \mathbb{R}^{\boldsymbol{n\times m}}, \boldsymbol{X} = (\boldsymbol{x_{ij}})_{\boldsymbol{m\times n}} \in \mathbb{R}^{\boldsymbol{m\times n}}\right)$

Có $f: \mathbb{R}^{m\times n} \to \mathbb{R};\ \ A.X = (c_{ij})_{n\times n}$ với: $c_{ii} = \sum_{k=1}^{m} a_{ik}.x_{ki}\ ; f(X) = \sum_{i=1}^{n} c_{ii}$

$$\nabla_X f(X) \triangleq \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f(X)}{\partial x_{m1}} & \frac{\partial f(X)}{\partial x_{m2}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} = A^T$$

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

7. Hàm $\boldsymbol{f}(\mathrm{X}) = \boldsymbol{a}^{\mathrm{T}}.\mathrm{X}.\boldsymbol{b},\ \mathrm{X} = (\boldsymbol{x_{ij}})_{\boldsymbol{m\times n}} \in \mathbb{R}^{\mathrm{m\times n}}$ $(\mathrm{a} \in \mathbb{R}^{\boldsymbol{m}},\ \mathrm{b} \in \mathbb{R}^{\boldsymbol{n}})$

Có: : f: $\mathbb{R}^{\mathrm{m\times n}} \to \mathbb{R}$; $\boldsymbol{f}(\boldsymbol{X}) = \sum_{j=1}^{n}\sum_{i=1}^{m} \boldsymbol{a_i b_j x_{ij}}$

$$\nabla_X f(X) \triangleq \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f(X)}{\partial x_{m1}} & \frac{\partial f(X)}{\partial x_{m2}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & \cdots & \vdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix} = a.b^T$$

8. Hàm $\boldsymbol{f}(\mathrm{X}) = \|\mathrm{X}\|_{\mathrm{F}}^{2}\ \left(\mathrm{X} = (\boldsymbol{x_{ij}})_{\mathrm{n\times n}} \in \mathbb{R}^{\mathrm{n\times n}}\right)$

Có f: $\mathbb{R}^{\mathrm{n\times n}} \to \mathbb{R}$: $\boldsymbol{f}(\mathrm{X}) = \sum_{\mathrm{j}=1}^{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x}_{\mathrm{ij}}^{2}$;

$$\nabla_X f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f(X)}{\partial x_{n1}} & \frac{\partial f(X)}{\partial x_{n2}} & \cdots & \frac{\partial f(X)}{\partial x_{nn}} \end{bmatrix} = \begin{bmatrix} 2x_{11} & 2x_{12} & \cdots & 2x_{1n} \\ 2x_{21} & 2x_{22} & \cdots & 2x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ 2x_{n1} & 2x_{n2} & \cdots & 2x_{nn} \end{bmatrix} = 2.\mathrm{X}$$

GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

9. Hàm $\boldsymbol{f(x) = trace\left(X^T.A.X\right)}, \left(\mathrm{X} = \left(\boldsymbol{x_{ij}}\right)_{\boldsymbol{m\times n}} \in \mathbb{R}^{\mathrm{m\times n}}, \mathrm{A} \in \mathbb{R}^{\mathrm{m\times m}}\right)$

Có: $\mathrm{f}: \mathbb{R}^{\mathrm{m\times n}} \to \mathbb{R}$. Kí hiệu các cột của X là: $\mathrm{x_1\ x_2} \dots \mathrm{x}_n$; $\mathrm{X} = [\mathrm{x_1\ \ x_2\ \ \dots\ \ x}_n]$

Khi đó: $\boldsymbol{X^T.A.X} = [\mathrm{x_1\ \ x_2\ \ \dots\ \ x}_n]^{\boldsymbol{T}}.\boldsymbol{A}.[\mathrm{x_1\ \ x_2\ \ \dots\ \ x}_n] = \left(\mathrm{x_i^T.A.x_j}\right)_{\mathrm{n\times n}}$

$$\boldsymbol{f(x) = trace\left(X^T.A.X\right)} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i^T.A.x_i}$$

Vì thế theo kết quả 3.6.3, ta có:

$$\nabla_X f(X) = \sum_{\mathrm{i=1}}^{\mathrm{n}} \boldsymbol{\nabla}(\mathrm{x_i^T.A.x_i}) = \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{A + A^T}).\mathrm{x_i} = (\mathrm{A + A^T}).\mathrm{X}$$

Nhận xét: $\nabla \boldsymbol{trace\left(X^T.X\right)} = \boldsymbol{\nabla trace\left(X^T.\mathrm{I}.X\right)} = 2\mathrm{I.X} = 2\mathrm{X}$ (là k.quả 3.4.8)

10. Hàm $\boldsymbol{f(X) = detX}\ \left(\mathrm{X} = \left(\boldsymbol{x_{ij}}\right)_{\boldsymbol{n\times n}} \in \mathbb{R}^{\mathrm{n\times n}}, \mathrm{X}\ \text{khả nghịch}\right)$

Khai triển theo hàng i có $\mathrm{detX} = \sum_{\mathrm{j}} \boldsymbol{x_{ij}}.\mathrm{X}_{ij}$ (X_{is} là phần phụ đại số của x_{ij} trong X nên không chứa x_{ij}, $\forall \mathrm{s} = 1,2,\dots,\mathrm{n}$), từ đó: $\frac{\boldsymbol{\partial}(\mathrm{detX})}{\boldsymbol{\partial}\mathrm{x}_{ij}} = \mathrm{X}_{ij}$. Vì thế:

$$\nabla_{\mathrm{X}} f(X) = \nabla_{\mathrm{X}}\mathrm{detX} = \left(\frac{\boldsymbol{\partial}(\mathrm{detX})}{\boldsymbol{\partial}\mathrm{x}_{ij}}\right)_{\mathrm{n\times n}} = \left(\mathrm{X}_{ij}\right)_{\mathrm{n\times n}} = (\mathrm{X}^*)^{\mathrm{T}} = (\mathrm{detX}).\mathrm{X}^{-1}$$

11. (BT) Hàm $\boldsymbol{f}(\mathrm{X}) = \|\boldsymbol{A}.\mathrm{X} - \mathrm{B}\|_{\mathrm{F}}^2$ $((\mathrm{X} \in \mathbb{R}^{\mathrm{m\times n}}, \mathrm{A} \in \mathbb{R}^{\mathrm{k\times m}}, \mathrm{B} \in \mathbb{R}^{\mathrm{k\times n}})$

Bằng cách tương tự như trong 3.6.8, ta có:

$$\boldsymbol{\nabla f(X) = \nabla \|A.X - B\|_F^2 = 2A^T(A.X - B)}$$

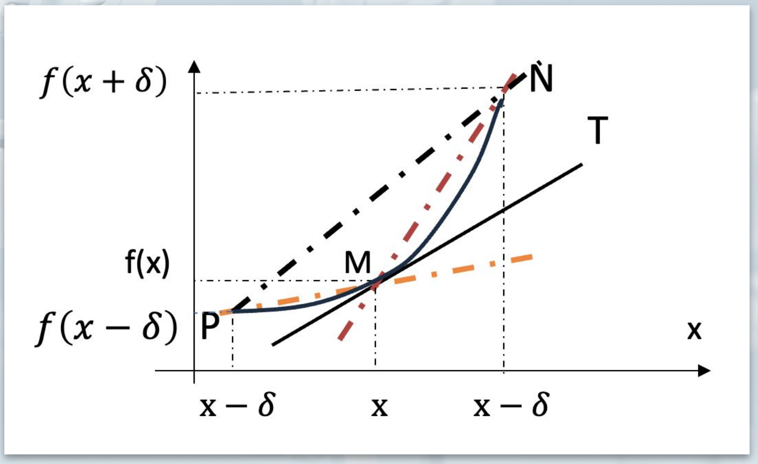
GIẢI TÍCH MA TRẬN TRONG MÁY HỌC

Bảng các đạo hàm cơ bản: Tổng hợp các kết quả trên, ta có bảng các đạo hàm cơ bản:

Dạng hàm f		Dạng hàm f	

Kiểm tra đạo hàm

Việc tính toán tìm đạo hàm của hàm nhiều biến là khá phức tạp và dễ mắc lỗi. Trong thực nghiệm để kiểm tra sự chính xác của việc tính toán đạo hàm, thường dựa trên định nghĩa đạo hàm của hàm một biến.



1. Xấp xỉ đạo hàm của hàm một biến số.
Với $\delta > 0$ đủ nhỏ ta có khai triển Taylor:

$f(x+\delta) = f(x) + f'(x).\delta + \alpha(\delta^2)$ ($\alpha(\delta^2)$ là VCB cấp cao hơn δ^2) (3.14)
$f(x-\delta) = f(x) - f'(x).\delta + \beta(\delta^2)$ ($\beta(\delta^2)$ là VCB cấp cao hơn δ^2) (3.15)

Từ (3.10) có: $f'(x) = \frac{f(x+\delta)-f(x)}{\delta} + o(\delta)$ ($o(\delta)$ là VCB cấp cao hơn δ)
hay có xấp xỉ: $f'(x) \approx \frac{f(x+\delta)-f(x)}{\delta}$, với sai số là $o(\delta)$ (3.16)
Từ (3.11) có: $f'(x) = \frac{f(x)-f(x-\delta)}{\delta} + o(\delta)$ ($o(\delta)$ là VCB cấp cao hơn δ)
hay có xấp xỉ: $f'(x) \approx \frac{f(x)-f(x-\delta)}{\delta}$, với sai số là $o(\delta)$ (3.17)
Từ (3.10), (3.11) có: $f'(x) = \frac{f(x+\delta)-f(x-\delta)}{2\delta} + o(\delta^2)$ ($o(\delta^2)$ là VCB cấp cao hơn δ^2)
nên có xấp xỉ : $f'(x) \approx \frac{f(x+\delta)-f(x-\delta)}{2\delta}$, với sai số là $o(\delta^2)$ (3.18)

Cách tính xấp xỉ này được gọi là Numerical Gradient. Công thức xấp xỉ (3.18) được sử dụng phổ biến để tính Numerical Gradient cho hàm số một biến số do có độ chính xác cao hơn: Khi δ đủ bé thì sai số $o(\delta^2) \ll o(\delta)$.

- Dựa vào ý nghĩa hình học của đạo hàm: Với hàm số y = f(x), đạo hàm f'(x) là hệ số góc tiếp tuyến MT đồ thị tại điểm M(x, f(x)), $\frac{f(x+\delta)-f(x)}{\delta}$ là hệ số góc của cát tuyến bên phải MN với $N(x+\delta, f(x+\delta))$; $\frac{f(x)-f(x-\delta)}{\delta}$ là hệ số góc của cát tuyến bên trái MP với $P(x-\delta, f(x-\delta))$. Rõ ràng là so với các cát tuyến MN, MP cát tuyến PN (có hệ số góc $\frac{f(x+\delta)-f(x-\delta)}{2\delta}$) có phương gần với tiếp tuyến MT hơn, tức là: $\frac{f(x+\delta)-f(x-\delta)}{2\delta}$ xấp xỉ cho f'(x) chính xác hơn là $\frac{f(x+\delta)-f(x)}{\delta}$ hay $\frac{f(x)-f(x-\delta)}{\delta}$.

2. Xấp xỉ đạo hàm của hàm nhiều biến số.

a. Khai triển Taylor cho hàm nhiều biến:

Giả sử $f: \mathbb{R}^n \to \mathbb{R}, f(x) = f(x_1, x_2, \dots, x_n)$ là hàm số n biến số. Với số dương δ đủ nhỏ, ta có công thức khai triển:

$$f(x + \delta.y) = f(x) + \delta. < \nabla f(x), y > + o(\delta) \qquad (3.19)$$

$$f(x + \delta.y) = f(x) + \delta. < \nabla f(x), y > + \frac{\delta^2}{2}.y^T.\nabla^2 f(x).y + o(\delta^2) \qquad (3.20)$$

Khai triển Taylor là cơ sở lý thuyết cho rất nhiều thuật toán tối ưu bằng cách xấp xỉ, đặc biệt là *Gradient descent* và *Newton step*

b. Xấp xỉ đạo hàm của hàm nhiều biến. Đ/v hàm nhiều biến $f(x) = f(x_1, \dots, x_n)$ xấp xỉ (3.18) được áp dụng cho từng biến (coi các biến khác là cố định) trong đạo hàm: $\boldsymbol{\nabla f(x)} \triangleq \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right)^T$.