Vector Calculus- 1

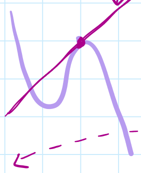
09 March 2024 12:32

Color Scheme:
#C6B0F2
#A07BE9
#AB008B
#EFDDFF

Scalar Derivatives

$f: \mathbb{R} \to \mathbb{R}$ ⟶ some point $x \to f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

input x changes by a small amount, how much f^n changes



$f(x+\varepsilon) = f(x) + \varepsilon \cdot f'(x)$

$y = f(x)$

$\frac{dy}{dx} = f'(x)$

$x \to x + \Delta x \Rightarrow y \approx y + \frac{dy}{dx} \Delta x$ → Also works as an operator!

$f(x) \to \boxed{\frac{d}{dx}} \to f'(x)$

Ex 1: $y = 9(x + x^2)$

$\frac{dy}{dx} = \frac{d}{dx}(9(x+x^2))$

$= 9 + 18x^2$

$\frac{d}{dx}$ acts as a distributive operator → Pull out constants + solve.

Vector Calculus + Partial Derivatives

NNs are NOT single f^n of single parameters → $f(x,y)$ instead of $f(x)$

Each function has multiple parameters ← So when we calculate derivs.

$\cancel{\frac{d}{dy}}$ $\frac{\partial}{\partial y}$ ✓

Derivative wrt one parameter WHILE THE OTHER REMAINS CONSTANT

Ex 1: $f(x,y) = xy$.

$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(xy) = y \frac{\partial}{\partial x}(x) = y$

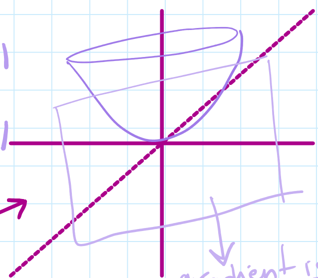
$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(xy) = x \frac{\partial}{\partial y}(y) = x$

$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2}{\partial x^2}(y) = 0$ [as y is a constant]

$\frac{\partial^2 f}{\partial y^2} = \frac{\partial^2}{\partial y^2}(x) = 0$ [x is a constant]

BUT $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}(x) = 1$

$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}(y) = 1$



gradient is a tangent plane that touches $f(\bar{x})$ @ one point.

Ex 2: $f(x,y) = 3x^2 y$.

$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(3x^2 y) = 3y \frac{\partial}{\partial x}(x^2) = (3y)(2x) = 6xy$

$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(3x^2 y) = 3x^2 \frac{\partial}{\partial y}(y) = 3x^2(1) = 3x^2$

$\nabla f(x,y) = \left[\frac{\partial}{\partial x} f(x,y), \frac{\partial}{\partial y} f(x,y)\right]$

$\nabla f(x,y) = [6xy, 3x^2]$

Interpretation

$\bar{x} \to \bar{x} + \Delta\bar{x}$

$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$y \simeq y + \frac{\partial y}{\partial \bar{x}} \Delta \bar{x}$

scalar; vector; vector · vec = scalar

⇒ Vector i/p, Scalar o/p

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

∇f is a vector of partial derivatives

$f(x_1, x_2 \cdots x_n)$

$\nabla f(p) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \cdots \frac{\partial f}{\partial x_n}\right]$

Essentially,

The combination of these partial derivatives gives us the dir^n of STEEPEST ASCENT

"dir^n in which you move to climb up the hill the fastest."

↳ magnitude shows how fast it's changing

$\frac{\partial y}{\partial \bar{x}} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots \frac{\partial y}{\partial x_n}\right]$

$\bar{x}$ is a vector ① x_i is the i^{th} coordinate of $\bar{x}$

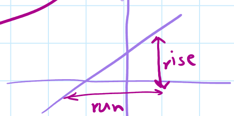
y is a scalar ② x_i is a scalar

$|\bar{x}| = N$

↳ $\frac{\partial y}{\partial x_i}$ = scalar

→ so $\left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2} \cdots\right]$ is a vector of PDs.

Gradient



for scalars: "dir^n of greatest change of a scalar f^n"

$m = \frac{\text{rise}}{\text{run}}$

Vector Calculus

"gradient of a scalar-valued f^n of several variables is a vector field ∇f whose value @ p gives the rate of fastest increase."

Steepest ascent = rel^n w/ rate of change + optimization of f^n

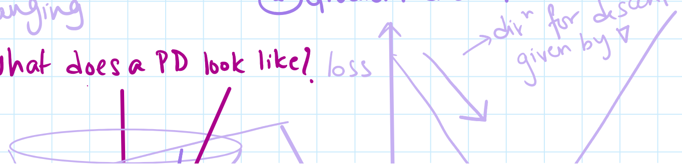
Applications:

① Optimization:

max cost, max profit → optimal params in models.

- Steepest ascent = fastest ↑ in f^n val.
- Allows algos to converge towards the maximum
- Negative of ∇ gives us dir^n of steepest descent to find min.

② Gradient descent



loss

dir^n for descent given by ∇

What does a PD look like?

Case study: Lion Brain

i/p: vectors

o/p: scalar — jump/run

$y = w_1 x_1 + w_2 x_2 + \ldots + w_N x_N$

o/p : scalar — jump/run

$y = w_1x_1 + w_2x_2 + \ldots + w_Nx_N$
$= \sum_{i=1}^{N} w_i x_i$

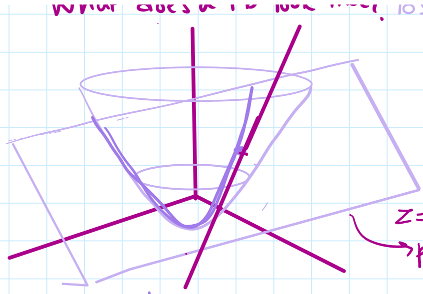
$\rightarrow \nabla y = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots \frac{\partial y}{\partial x_N}\right]$
$= [w_1, w_2, \ldots w_N]$
$= \frac{\partial}{\partial x_i}\left(\sum_{i=1}^{N} w_i x_i\right) = w_i$ → only i^{th} term stays → $\frac{\partial}{\partial x_i}[w_1x_1 + w_2x_2, \ldots w_ix_i \ldots w_Nx_N]$

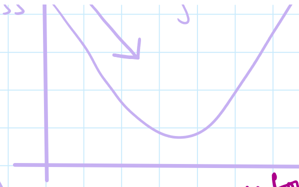
$= \cancel{\frac{\partial}{\partial x_i}(w_1x_1)}^{0} + \cancel{\frac{\partial}{\partial x_i}(w_2x_2)}^{0} + \frac{\partial}{\partial x_i}(w_ix_i) + \ldots \cancel{\frac{\partial}{\partial x_i}(w_Nx_N)}^{0}$

$= w_i \frac{\partial}{\partial x_i}(x_i)$

$= w_i$

What does a 3D look like? loss





$z = x^2 + y^2$ → plane at $y = 2$ → $z = x^2 + 4$, $\frac{\partial z}{\partial x} = 2x$ → we found tangent keeping y const

Matrix I/P, Scalar O/P

→ function of a matrix instead of a vector

Here, we're moving from derivative of one f^n $f(\bar{x})$ to derivative of many functions $f(W)$

vector calculus → matrix calculus.

$f(W), \; W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}$

for a vector: $\nabla f(\bar{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \ldots \frac{\partial f}{\partial x_n}\right)$

$\rightarrow \frac{\partial f}{\partial W} = \begin{bmatrix} \partial f/\partial w_{11} & \partial f/\partial w_{12} & \partial f/\partial w_{13} \\ \vdots & & \\ \partial f/\partial w_{31} & \partial f/\partial w_{32} & \partial f/\partial w_{33} \end{bmatrix}$

Demonstration: $g(x,y) = 2x + y^8$
$f(x,y) = 3x^2y$

$\frac{\partial}{\partial x} g = \frac{\partial}{\partial x}(2x + y^8) = 2$

$\frac{\partial}{\partial y} g = \frac{\partial}{\partial y}(2x + y^8) = 8y^7$

$\nabla g(x,y) = \left[\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right] = [2, 8y^7]$

$\frac{\partial}{\partial x} f = \frac{\partial}{\partial x}(3x^2y) = 6xy$

$\frac{\partial}{\partial y} f = \frac{\partial}{\partial y}(3x^2y) = 3x^2$

$\nabla f(x,y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] = [6xy, 3x^2]$

Before: single f^n, vector i/p
Now: multiple functions, matrix i/p → organize their gradients.

↓

Stack gradient vectors = JACOBIAN
$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$

$J = \begin{bmatrix} \nabla f(x,y) \\ \nabla g(x,y) \end{bmatrix} = \begin{bmatrix} 6xy & 3x^2 \\ 2 & 8y^7 \end{bmatrix}$

$J = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \cdots & \partial f_1/\partial x_n \\ \vdots & & \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 \cdots & \partial f_m/\partial x_n \end{bmatrix}_{m \times n}$

$\rightarrow J = \begin{bmatrix} \partial f/\partial x & \partial f/\partial y \\ \partial g/\partial x & \partial g/\partial y \end{bmatrix}$ → Our input was a set of vectors & our o/p (Jacobian) was also a set of vectors.

So far:

1. gradient of f^n w/ scalar o/p = gradient vector
2. gradient of f^n w/ vector o/p = Jacobian

Take derivative again → HESSIAN → Matrix of 2^{nd} Order Derivatives

Vector of m scalar valued f^n's

$\bar{y} = f(\bar{x}), \quad |\bar{x}| = n$
→ each f_i f^n here can return a scalar

$y_1 = f_1(\bar{x})$
NOT VECTOR ← $y_2 = f_2(\bar{x})$
$\vdots$
$y_m = f_m(\bar{x})$

Special case of an IDENTITY f^n:
$\bar{y} = f(\bar{x})$
$\rightarrow |\bar{x}| = n$
$\rightarrow |\bar{y}| = m$

$\frac{\partial \bar{y}}{\partial \bar{x}} = \begin{bmatrix} \nabla f_1(x) \\ \nabla f_2(x) \\ \vdots \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \bar{x}} f_1(\bar{x}) \\ \vdots \end{bmatrix}$

1. $f_1(\bar{x})$ or y_1 is scalar!!
2. Diff^n each scalar wrt vector

Ex 1: $f(x,y) = xy$

$\frac{\partial f}{\partial x} = y \qquad \frac{\partial f}{\partial y} = x \qquad [\nabla f = (y, x)]$

$\frac{\partial^2 f}{\partial x^2} = 0 \qquad \frac{\partial^2 f}{\partial y^2} = 0$

$\frac{\partial^2 f}{\partial x \partial y} = 1 \qquad \frac{\partial^2 f}{\partial y \partial x} = 1$

$H = \begin{bmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial y \partial x \\ \ldots \end{bmatrix}$

$$\frac{\partial \bar{y}}{\partial \bar{x}} = \begin{bmatrix} \nabla f_1(x) \\ \nabla f_2(x) \\ \vdots \\ \nabla f_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \bar{x}} f_1(\bar{x}) \\ \vdots \\ \frac{\partial}{\partial \bar{x}} f_m(\bar{x}) \end{bmatrix}$$

(1) $f_1(\bar{x})$ or y_1 is scalar!!

(2) Diffⁿ each scalar wrt vector = vector

sounds like a ... GRADIENT!!!

$$\mathcal{H} = \begin{bmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial y \partial x \\ \partial^2 f/\partial x \partial y & \partial^2 f/\partial y^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{n \times n}$$

(1) Grad: vector of n dimensions

(2) $\mathcal{H}$: $n \times n$ matrix

$|\bar{x}| = n$

Back to solving:

J { vector in, vector out

$$\frac{\partial \bar{y}}{\partial \bar{x}} = \begin{bmatrix} \frac{\partial f_1(\bar{x})}{\partial x_1} & \frac{\partial f_1(\bar{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\bar{x})}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_m(\bar{x})}{\partial x_1} & \frac{\partial f_m(\bar{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\bar{x})}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & & \frac{\partial x_n}{\partial x_n} \end{bmatrix} \rightarrow \frac{\partial x_i}{\partial x_j} = 0, \text{ when } i \neq j$$

$$= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & \cdots & & & 1 \end{bmatrix} = I \text{ (identity matrix)} \checkmark$$

Single Variable Chain Rule

There's only a single dataflow path from x to y + none of intermediate functions have more than one param.

How to:

1. Introduce intermediate variables.
2. Compute derivative wrt parameters.
3. Combine all derivatives of interm. variables.
4. Substitute intermediate values.

* forward diffⁿ

$x \rightarrow y$

$$\frac{dy}{dx} = \frac{du}{dx} \cdot \frac{dy}{du}$$

* backward diffⁿ

$y \rightarrow x$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Ex 1: $y = f(g(x)) = \sin(x^2)$

(1) $u = x^2$

$\Rightarrow y = \sin u$

(2) $\frac{du}{dx} = 2x \qquad \frac{\partial y}{\partial u} = \cos u$

(3) $\frac{\partial y}{\partial x} = \frac{\partial u}{\partial x} \cdot \frac{\partial y}{\partial u} = 2x \cos u$

(4) $\frac{\partial y}{\partial x} = 2x \cos x^2$

$y = \sin$

$u = \text{square}$

x

data flow

Ex 2: $f(x) = \ln(\sin(x^3)^2)$

(1) $u = x^3$

$v = \sin u$

$w = v^2$

$y = \ln w$

(2) $\frac{dy}{dx} = \frac{dy}{dw} \cdot \frac{dw}{dv} \cdot \frac{dv}{du} \cdot \frac{du}{dx}$

$= \frac{1}{w} \cdot 2v \cdot \cos u \cdot 3x^2$

(3) $\frac{dy}{dx} = \frac{1}{v^2} \cdot 2v \cdot \cos x^3 \cdot 3x^2$

$= \frac{2}{\sin x^3} \cdot \cos x^3 \cdot 3x^2$

(4) $= 6x^2 \cdot \cot x^3$

direction of data flow

x → cube → sin → square → ln

Single variable total Derivative Chain Rule

Case study: $y = f(x) = x + x^2$

$\frac{dy}{dx} = 1 + 2x$ (using the usual rules)

But let's try it w/ single variable chain rule:

$u_1(x) = x^2$

$y = u_2(x) = x + u_1$

$\Rightarrow \frac{du_2}{du_1} = 0 + 1 = 1$

$\frac{du_1}{dx} = 2x$

after fixing

$$\frac{dy}{dx} = \frac{\partial u_2}{\partial x} + \frac{\partial u_2}{\partial u_1} \cdot \frac{\partial u_1}{\partial x}$$

$$= \frac{\partial}{\partial x}(x + u_1) + \frac{\partial}{\partial u_1}(x + u_1) \cdot \frac{\partial}{\partial x}(x^2)$$

$= 1 + (1)(2x)$

$= 1 + 2x \checkmark$

Now solve old case: $y = x + x^2$

$x \rightarrow +$

$x \rightarrow$ square $\rightarrow +$

Multiple paths

But why?

for single

→ direct + indirect

pa... But why?

$\frac{du_1}{dx} = 2x$

Law of total derivative.

direct + indirect

"To compute $\frac{dy}{dx}$, we need to sum up all the contribution from x to y."

for single var. ch. rule, we only had a single path of data flow.

$\Rightarrow \frac{dy}{dx} = \frac{du_2}{du_1} \cdot \frac{du_1}{dx}$

$= (1)(2x) = 2x$ ✗

can we try w/ PD?

$\frac{du_1(x)}{dx} = 2x$

$\frac{\partial u_2(u_1,x)}{\partial u_1} = \frac{\partial}{\partial u_1}(x+u_1) = 1$

$\frac{\partial u_2(u_1,x)}{\partial x} = \frac{\partial}{\partial x}(x+u_1) = 1$

we're assuming one variable is constant. So no work.

$$\frac{dy}{dx} = \frac{\partial f(x)}{\partial x} = \frac{\partial u_2(x,u_1)}{\partial x}$$

$$= \frac{\partial u_2}{\partial x} \cdot \frac{\partial x}{\partial x} + \frac{\partial u_2}{\partial u_1} \cdot \frac{\partial u_1}{\partial x}$$

Now, w/ more variables:

$$\frac{\partial f(x,u_1,\dots u_n)}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u_1} \cdot \frac{\partial u_1}{\partial x} + \dots \frac{\partial f}{\partial u_n} \cdot \frac{\partial u_n}{\partial x}$$

$$\underbrace{\frac{\partial f(x,u_1,\dots u_n)}{\partial x}}_{\text{partial derivative}} = \underbrace{\frac{\partial f}{\partial x} + \sum_{i=1}^{n} \frac{\partial f}{\partial u_i} \times \frac{\partial u_i}{\partial x}}_{\text{total derivative}}$$

can we simplify it further?

temporary variable for x: $\boxed{u_{n+1} = x}$ don't forget how we got here!

$$\Rightarrow \frac{\partial f(x,u_1,\dots u_n)}{\partial x} = \sum_{i=1}^{n+1} \frac{\partial f}{\partial u_i} \times \frac{\partial u_i}{\partial x}$$

$\rightarrow$ this looks like a dot product now: $\frac{\partial f}{\partial \bar{u}} \cdot \frac{\partial \bar{u}}{\partial x}$

$\bar{u}$ is a vec

here we're expanding $\bar{u}$ into its constituents

$[\bar{u} = (u_1 + u_2 + u_3 + \dots)^T = (\sum_{i=1}^{n} u_i)^T]$

Ex: $f(x) = \sin(x + x^2)$

① intermediate variables.

notice how we're nesting!

- $u_1(x) = x^2$
- $u_2(x,u_1) = x + u_1$
- $u_3(u_2) = \sin(u_2)$

like a russian doll but w/ f^ns.

② Taking partials:

$\frac{\partial}{\partial x} u_1(x) = 2x$

$\frac{\partial}{\partial x} u_2(x,u_1) = \frac{\partial u_2}{\partial x} + \frac{\partial u_2}{\partial u_1} \cdot \frac{\partial u_1}{\partial x}$

w/ this formula

just OPEN it up along each of its constituents.

$= 1 + (1 \cdot 2x)$

$= 1 + 2x$

$\frac{\partial}{\partial x} u_3(u_2) = \frac{\partial u_3}{\partial x} + \frac{\partial u_3}{\partial u_2} \cdot \frac{\partial u_2}{\partial x}$

first diff wrt whoever came first outside

then diff wrt the awkward outsider but make sure it looks like $\partial f/\partial x$ @ the end

Eg: $\frac{\partial u_3}{\partial u_2} \cdot \frac{\partial u_2}{\partial x} = \frac{\partial u_3}{\partial x}$ w/ extra steps to include u_2

$= 0 + \cos(u_2) \cdot (1 + 2x)$

③ Put it all together:

$\frac{\partial f}{\partial x} = \cos(u_2) \cdot (1+2x)$

$= (1+2x) \cos(x + u_1)$

$= \underline{(1+2x) \cos(x+x^2)}$

To remember: Total derivative ALWAYS sums

Eg: $y = x \cdot x^2$

① $u_1(x) = x^2$

$u_2(x, u_1) = x \cdot u_1$

data flow: $x \rightarrow u_2$, $u_1 \rightarrow u_2$, $x \rightarrow u_1$

"u_2 is a f^n of $x + u_1$. u_1 is a f^n of x."

② $\frac{\partial u_1}{\partial x} = 2x$

$\frac{\partial}{\partial x} u_2(x,u_1) = \frac{\partial u_2}{\partial x} + \frac{\partial u_2}{\partial u_1} \cdot \frac{\partial u_1}{\partial x}$

$= u_1 + x \cdot 2x$

$= u_1 + 2x^2$

③ $\frac{\partial y}{\partial x} = u_1 + 2x^2$

$= x^2 + 2x^2$

$= 3x^2$ ✓

Why learn it though?

- simplify chain rule
- AUTOMATIC DIFN IN NN (well, mimics it)

Vector Chain Rule.

Deals w/ vectors of variables + functions!

$\bar{y} = \bar{F}(x)$ $\rightarrow$ $\bar{y}$ is a vector $\rightarrow$ function is a vector.

$$\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}$$

$\rightarrow$ here, vectors are like accordions

$\rightarrow$ Let $\bar{y} = \bar{f}(x)$

Namita Achyuthan

$$\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}$$

Assumption: I/P is a vector of f's of a single variable.

$\rightarrow$ Let $\bar{y} = \bar{f}(x)$

$$\underbrace{\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}}_{\text{All single variable.}} = \underbrace{\begin{bmatrix} \ln(x^2) \\ \sin(3x) \end{bmatrix}}_{\text{vector of functions}}$$

① We introduce two intermediate variables, $g_1 + g_2$. $\bar{g} = [g_1 \; g_2]^T$ cuz col^m vector

How?

$$\bar{y} = \bar{f}(x) = \bar{f}(\bar{g}(x))$$

$$\Rightarrow \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} = \begin{bmatrix} x^2 \\ 3x \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} f_1(\bar{g}) \\ f_2(\bar{g}) \end{bmatrix} = \begin{bmatrix} \ln(g_1) \\ \sin(g_2) \end{bmatrix}$$

★ Derivative of vector $\bar{y}$ wrt scalar x

$\downarrow$

COLUMN VECTOR

$$\frac{\partial \bar{y}}{\partial x} = \begin{bmatrix} \partial y_1/\partial x \\ \partial y_2/\partial x \\ \vdots \\ \partial y_n/\partial x \end{bmatrix}$$

② $$\frac{\partial \bar{y}}{\partial x} = \begin{bmatrix} \partial f_1(\bar{g})/\partial x \\ \partial f_2(\bar{g})/\partial x \end{bmatrix}$$

★ Components of vector $\bar{y}$ are computed using single var total deriv chain rule.

$$\frac{\partial \bar{y}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} f_1(g_1, g_2) \\ \frac{\partial}{\partial x} f_2(g_1, g_2) \end{bmatrix}$$

$$\frac{\partial \bar{y}}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial g_1} \cdot \frac{\partial g_1}{\partial x} + \frac{\partial f_1}{\partial g_2} \cdot \frac{\partial g_2}{\partial x} \\ \frac{\partial f_2}{\partial g_1} \cdot \frac{\partial g_1}{\partial x} + \frac{\partial f_2}{\partial g_2} \cdot \frac{\partial g_2}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{1}{g_1} \cdot 2x + 0 \\ 0 + \cos(g_2) \cdot 3 \end{bmatrix} = \begin{bmatrix} 2x/x^2 \\ 3\cos(3x) \end{bmatrix} = \begin{bmatrix} 2/x \\ 3\cos(3x) \end{bmatrix}$$

BUT WE NO DONE!

To note: Vector Rule

$$\boxed{\frac{\partial}{\partial x} \bar{f}(\bar{g}(x)) = \frac{\partial \bar{f}}{\partial \bar{g}} \cdot \frac{\partial \bar{g}}{\partial x}}$$

vector, vector

We can write this as a matrix-vector oper^n !!

We have $\bar{f}(\bar{g}(x))$

do $\bar{g}$ wrt x first.

$$\frac{\partial \bar{y}}{\partial x} = \frac{\partial \bar{f}(\bar{g}(x))}{\partial x} = \begin{bmatrix} \frac{\partial f_1(\bar{g})}{\partial x} \\ \frac{\partial f_2(\bar{g})}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial g_1} & \frac{\partial f_1}{\partial g_2} \\ \frac{\partial f_2}{\partial g_1} & \frac{\partial f_2}{\partial g_2} \end{bmatrix} \begin{bmatrix} \frac{\partial g_1}{\partial x} \\ \frac{\partial g_2}{\partial x} \end{bmatrix} = \frac{\partial \bar{f}}{\partial \bar{g}} \cdot \frac{\partial \bar{g}}{\partial x}$$

and then multiply $\bar{f}$ wrt $\bar{g}$

Generalization Time!

Assumptions: $\bar{x}, \bar{g}, \bar{f}$ are vectors.

$|\bar{x}| = n$ $\quad$ $|\bar{g}| = k$ $\quad$ $|\bar{f}| = m$

$\bar{f} = \bar{f}(\bar{g})$

$\bar{g}: \mathbb{R}^m \rightarrow \mathbb{R}^k$

$\bar{f} = \bar{f}(\bar{x})$ thro' $\bar{g}$

$\bar{f}: \mathbb{R}^k \rightarrow \mathbb{R}^n$

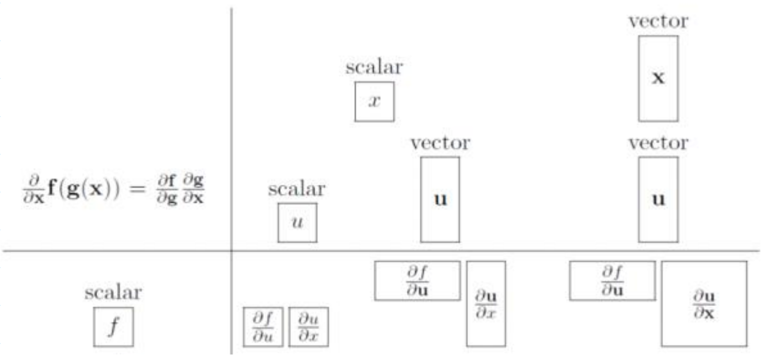
vector in $\rightarrow$ $\boxed{\text{Jacobian}}$ $\rightarrow$ vector out

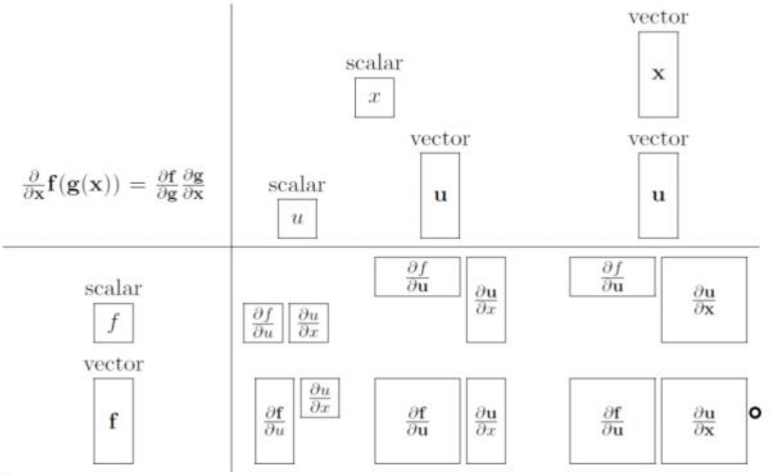
$(m \times k) \times (k \times n)$

ie [dim^n of outermost × dim^n of innermost]

Jacobian = $m \times n$ matrix for $\frac{\partial}{\partial \bar{x}} \bar{f}(\bar{g}(\bar{x}))$

$$\frac{\partial}{\partial \bar{x}} \bar{f}(\bar{g}(\bar{x})) = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial g_1} & \frac{\partial f_1}{\partial g_2} & \cdots & \frac{\partial f_1}{\partial g_k} \\ \vdots & & & \\ \frac{\partial f_m}{\partial g_1} & \frac{\partial f_m}{\partial g_2} & \cdots & \frac{\partial f_m}{\partial g_k} \end{bmatrix}}_{m \times k} \underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & & \\ \frac{\partial g_k}{\partial x_1} & \frac{\partial g_k}{\partial x_2} & \cdots & \frac{\partial g_k}{\partial x_n} \end{bmatrix}}_{k \times n}$$




$\frac{\partial}{\partial \mathbf{x}}\mathbf{f}(\mathbf{g}(\mathbf{x})) = \frac{\partial \mathbf{f}}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{x}}$
scalar
vector
scalar
vector
vector
scalar
vector