Intégrale

constante : x

x : x2/2

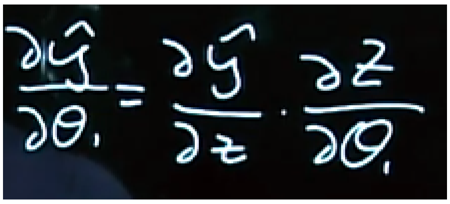
x^n : x^(n+1)/ (n+1)

e^(a*x+b) : (1/a)* e^(a*x+b)

a/x :ln(ax)

Dérivée

Chain rule :



x^n -> n*x^(n-1)

x : 1

e^x : x’ e^x

ln(x) : 1/x

ln(f) : f’/f

Dérivée produit h(x) = f(x)*g(x) :   h’(x) = f’(x)*g(x) + f(x)*g’(x)

Dérivée de la somme: somme des dérivées

Calculators

https://www.symbolab.com/solver/

https://www.derivative-calculator.net/

https://www.integral-calculator.com/

Inverse of a function

Remplacer y par x et solve pour x

Famous inverses :

logB(x) : b^x

ln(x): e^x

Si j’applique l’inverse dans une inegalité, le sens de l’inégalité change

Limits

Lim n infinity (1 + x/n)^n → e^x

L’Hospital rule to check limits

x->c

f(x), g(x) -> inf

lim as x-> c of f(x) / g(x) eq. lim as x-> c of f’(x) / g’(x)

Inequalities

e^(-x) >= 1 - x

e^x >= 1 + x

Exp

Exp(-inf)=0

<u>Series expansion for the exponential</u>

Exp peut s’exprimer en fonction de serie avec des factorielles :

$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\ldots\ldots\ldots.$

$e^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \ldots\ldots\ldots\ldots.$ [coefft. Of x^n ix $(-1)^n/n!$]

by adding & subtracting above two series , we get

$\frac{e^x+e^{-x}}{2} = 1 + x^2/2! + x^4/4! + \ldots\ldots\ldots$ and $\frac{e^x-e^{-x}}{2} = x + x^3/3! + x^5/5! + \ldots\ldots\ldots$

$\frac{e^1+e^{-1}}{2} = \sum_{n=0}^{\infty}\frac{1}{(2n)!} = \sum_{n=1}^{\infty}\frac{1}{(2n-2)!}$ and $\frac{e^1-e^{-1}}{2} = \sum_{n=0}^{\infty}\frac{1}{(2n+1)!} = \sum_{n=1}^{\infty}\frac{1}{(2n-1)!}$

Some important points should be noticed

(i) $\sum_{n=0}^{\infty}\frac{1}{(n)!} = e = \sum_{n=0}^{\infty}\frac{1}{(n-1)!}$

(ii) $\sum_{n=1}^{\infty}\frac{1}{(n)!} = e - 1$, $\sum_{n=2}^{\infty}\frac{1}{(n)!} = e - 2$

(iii) $\sum_{n=0}^{\infty}\frac{1}{(n+1)!} = e - 1$, $\sum_{n=0}^{\infty}\frac{1}{(n+2)!} = e - 2 = \sum_{n=1}^{\infty}\frac{1}{(n+1)!}$

Ln

-ln(x)= ln(1/x)

log(x/y) = log(x) – log(y)

sum of the logs is the log of the products

log of the products is sum of the logs

log(x^k) = k*log(x)

ln(a*x) = ln(a) + ln(x)

ln(e^a) = a (ln is inverse of exp)

General calculous

x^a + x^b = x^(a+b)

(x^a)^b = x^ab

Series

1+2+3+… + n = n(n+1)/2

$$\frac{1}{6}n(n+1)(2n+1)$$

$$\overbrace{(0^2+1^2+2^2+\dots+n^2)}^{n}$$

Geometric series |x| > 1, Sum (from 0 to inf) of x^n = 1/(1-x)

Sets

Bonferroni's inequality.

(a) Prove that for any two events A_1 and A_2, we have

$$\mathbf{P}(A_1 \cap A_2) \geq \mathbf{P}(A_1) + \mathbf{P}(A_2) - 1.$$

(b) Generalize to the case of n events $A_1, A_2, \dots, A_n$, by showing that

$$\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) \geq \mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots + \mathbf{P}(A_n) - (n-1).$$

Countable: discrete sets, including infinite discrete sets, like N or Rational numbers. By def, can be arranged in a sequence indexed by a positive integer.

Uncountable: continuous sets, like R or place

Trigo

cos(x)^2 + sin(x)^2 = 1

Symmetric functions

|x|, x2, … are symmetric functions

Vectors

<u>Norm</u>

Norm of a vector ||v ||: square root of the sum of (each element squared).

Dérivée de la norme d’un vecteur v = Dérivée de v

$$\|\mathbf{Ax}\|^2 = (\mathbf{Ax})^T(\mathbf{Ax}) = \mathbf{x}^T\mathbf{A}^T\mathbf{Ax}$$

<u>Orthogonal vector V (angle between then is pi/2)</u>

vi’.vj = 0

<u>Orthonormal vector</u>

It is an orthogonal vector with norm 1: vi’.vi = 1

<u>Dot product of u.v</u>

u.v = Sum over i (ui.vi)

u.v = norm(u).norm(v).cos(angle of u,v)

<u>Projection of a vector u onto a vector v</u>

How much of u goes in the direction of v

Proj onto v (u) = (u.v/norm^2(v)) . v

Planes

Defining a plane

Theta is orthogonal to the plane, has the same dimension as X

X.Theta + Theta0 = 0

Signed distance between the hyperplane and x

$$\frac{trans\left(x\right)\cdot\theta+\theta_0}{||\theta||}$$

Matrices

Symmetric

A’ = A

Spectral theorem for symmetric matrix A:

A = V L V’ with V orthogonal and L diagonal

A.vi = lambdai.vi with lambda eigenvalue of A

Orthogonal matrix V

VV’ = I = V’V

||Vx || = ||x ||

Matrix trace tr(A)

Sum of the elements of the diagonal.

If symmetric matrix, it is the sum of the eigenvalues of the matrix.

Negative semi definite

If the eigenvalues are negative, the matrix is negative semi definite.

Invertible

Invertible matrix ⇔ Determinant IS NOT 0

Invertible matrix ⇔ matrix n*p has full rank p (p<n)

Determinant IS NOT 0 ⇔ matrix n*p has full rank p

<u>Matrix determinant det(A)= |A|</u> : scalar rep volume scaling factor of the linear transformation described by the matrix.

Matrix determinant det(A)= product of the eigenvalues of the matrix.

For a matrix a b, det(A)= ad – bc (a b)

(c d)

An invertible matrix has (multiplicatively) invertible determinant.

If the determinant is invertible, then so is the matrix itself.

<u>Rank</u>

Nb of non zero linerarly independent vectors = dimension of the span of vectors

Column space = Row space

"Every rank-1 matrix can be written as an outer product. Conversely, every outer product $\mathbf{u}\mathbf{v}^T$ is a rank-1 matrix."

<u>Inverse</u>

To find the inverse of a 2x2 matrix: swap the positions of a and d, put negatives in front of b and c, and divide everything by the determinant (ad-bc).

A.A^-1 = Identity matrix

<u>Inverse of a diagonal matrix</u>

Take the inverse of each diag element

Hint: The inverse of a diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ where $a, b \neq 0$ is the diagonal matrix $\begin{pmatrix} 1/a & 0 \\ 0 & 1/b \end{pmatrix}$.

<u>Projection matrix</u>

H^2 = H

<u>Orthogonal projection matrix</u>

H^2 = H

It is symmetric H=H’

Its eigenvalues are 0 or 1.

In R

mymatrix <- **matrix(***vector*, **nrow=***r*, **ncol=***c*, **byrow=***FALSE*,
dimnames=list(*char_vector_rownames*, *char_vector_colnames***))**

```
A <- matrix( c(5, 1, 0,
                   3,-1, 2,
                   4, 0,-1), nrow=3, byrow=TRUE)
   det(A)
## [1] 16
```

`det(A) != 0`, so inverse exists

library(matlib)

```
   (AI  <- inv(A))
```

Matrix multiplication

The commutative property of multiplication does not hold!	$AB \neq BA$
Associative property of multiplication	$(AB)C = A(BC)$
Distributive properties	$A(B+C) = AB + AC$
	$(B+C)A = BA + CA$
Multiplicative identity property	$IA = A$ and $AI = A$
Multiplicative property of zero	$OA = O$ and $AO = O$
Dimension property	The product of an $m \times n$ matrix and an $n \times k$ matrix is an $m \times k$ matrix.

<u>In R</u>

Element wise multiplication: *

Vector multiplication. %*%

Vector transpose: t(vector)

<u>In Python</u>

@ (alias for numpy.matmul())

Quadratic function calculator

https://www.symbolab.com/solver/quadratic-equation-calculator

Convolution

Mathematically speaking it is a weighted average with one function(discrete or analog) constituting the weights and another the function to be averaged.

To maximize $\log P(D|\theta)$ subject to the contraint $\sum_{w\in W}\theta_w = 1$, we use the Lagrange multiplier method.

Method of Lagrange Multipliers

Problem

Let f be a function from $\mathbb{R}^N$ to $\mathbb{R}$. Find the (local) maxima/minima of f **subject to a given constraint** $g = 0$, where g is a function $\mathbb{R}^N$ to $\mathbb{R}$.

A two dimensional example is: Find the local extrema of $f(x,y) = x^2$ subject to the constraint $x^2 + y^2 = 1$ i.e. optimize the function f on the unit circle.

Method of Lagrange Multipliers

Without the constraint, the optimization problem can be solved as usual by setting the gradient of f to zero i.e.

$$\nabla f = 0.$$

With the constraint, we can solve the following equation instead:

$$\nabla f = \lambda \nabla g$$

where λ is a constant scalar. Geometrically, for $\lambda \neq 0$, a solution to the equation above is a point in $\mathbb{R}^N$ where the gradient of f is "parallel" to the gradient of g, or equivalently, where the gradient of f is perpendicular to the tangent of the curve defined by $g = k$ for some k. In other words, at a solution point, the directional derivative of f is zero along the direction tangent to the curve $g = k$ for some constant k, and hence f is stationary as we travel along $g = k$.

Finally, we impose the constraint $g = 0$ to find the local extrema of f on $g = 0$.

Since the equation $\nabla f = \lambda \nabla g$ is equivalent to $\nabla L = 0$ where $L = f - \lambda g$, the problem of optimizing f subject to $g = 0$ can be reformulated as optimizing the function L along with the constraint $g = 0$. We call the function L the **Lagrangian function** , and the scalar λ the **Lagranage multiplier** .

Note that we can equally define $L = f + \lambda g$, since λ is an unknown scalar we will solve for.

Example

Find the local extrema of $f(x,y) = x^2$ subject to the constraint $x^2 + y^2 = 1$. Geometrically, the function f is a parabolic cylinder, i.e. f is a parabolic in the x direction with constant values in the y direction. The constraint is a unit circle.

Solution:
First, solve the equation

$$\begin{aligned}
\nabla f &= \lambda \nabla g \qquad \text{where } g(x,y) = x^2 + y^2 - 1\\
\Longleftrightarrow \quad \begin{bmatrix} 2x \\ 0 \end{bmatrix} &= \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix}\\
\Longleftrightarrow \quad \begin{bmatrix} (1-\lambda)\,2x \\ \lambda(2y) \end{bmatrix} &= 0
\end{aligned}$$

The set of all possible solutions to the equation above are $(\lambda = 1, y = 0)$, or $(\lambda = 0, x = 0)$, or $(x = y = 0)$.

Finally, impose the constraint $x^2 + y^2 - 1 = 0$ to further pin down the local extrema. Subject to $x^2 + y^2 = 1$, $f(x,y) = x^2$ is at local maximum or mininum at $(x = 0, y = \pm 1)$ and $(y = 0, x = \pm 1)$. At $(x = 0, y = \pm 1)$, we have $\lambda = 0$ and $\nabla f = 0$. Since f has only local minima, these two points remain local minima of f on the unit circle. At $(y = 0, x = \pm 1)$, we have $\lambda = 1$ and hence $\nabla f = \nabla g$. Equivalently, the directional derivative ∇f is zero along the tangent direction of the circle at this point. Visualizing or computing second derivatives will allow us to see that these two points are local maxima of f along the unit circle.

Hide

Define the Lagrange function:

$$L = \log P(D|\theta) + \lambda\left(\sum_{w\in W}\theta_w - 1\right)$$