MAXIMA & MINIMA

A TOPIC IN MULTIVARIABLE CALCULUS

The problems of maxima and minima are problems of optimization, if one has dealt with basic calculus or basic physics then one has dealt with the problem of maxima and minima. A good use case from data science perspective is the problem of gradient descent where we use the concepts of maxima and minima to find the optimum value of parameters.

Most of us are aware that in case of maxima or minima the first derivative is zero. If the second derivative is negative then it's a maxima and if it's positive then it's minima. But why is that so? Intuition behind it and mathematical proof.

What about function of two variables? How to prove that ?

In the process of answering these questions we would explore the beautiful interconnectedness that lies in mathematics and we would see, that understanding one problem in depth can uncover many concepts.

So the right place to start is from Taylor's approximation of any function.

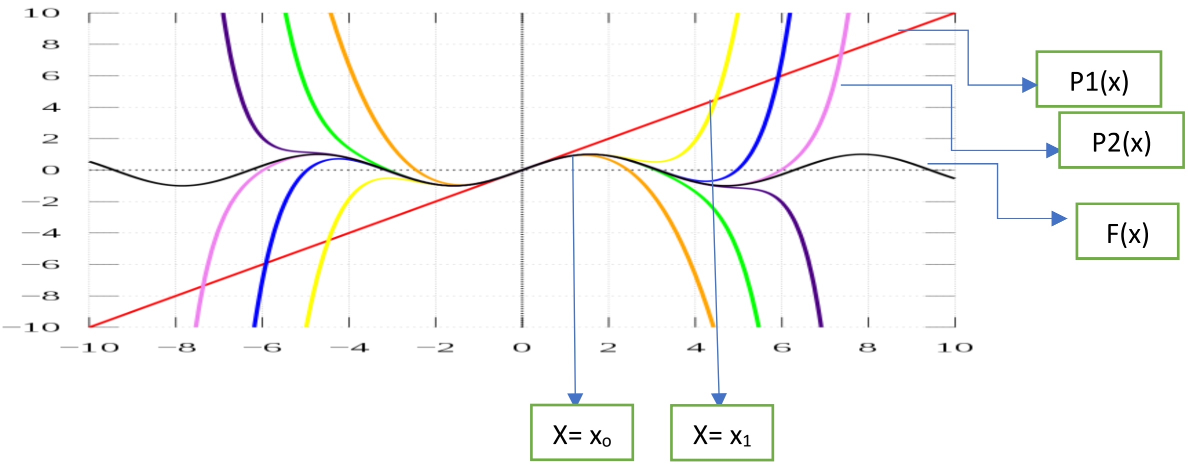
TAYLOR SERIES(ONE VARIABLE)

Taylor's approximation is one the most useful important approximation in mathematics and it's imperative that those who are interested in understanding math especially calculus, should know it. This would be especially useful once we move to two variables.

The gist of Taylor's theorem is that if a function is continuous and differentiable in a domain, then we can approximate the function around a point by a polynomial of n^{th} degree.

The higher the degree of the polynomial, the better the approximation.

What does it mean?



I the above diagram we approximate the curve F(x) at point x_o using the curves P1(x), P2(x), P3(x) at point x_o and so on.

We can see that after a certain number of terms the introduction of new terms add little value to accuracy of the result. We can also see that as we move from X =x_o to say x_1 the approximation fails.

So what are the polynomials P1, P2 etc.

$F(x+h) = F(x) + A_1 h + A_2h^2 + A_3h^3 + \ldots\ldots.\ A_nh^n$ ………..   (i)

We need to determine A_o , A_1 , A_2, A_3, … A_n, to find the nth order polynomial to approximate F(x+h) given that we know F(x).

Assumption being that h is small as f(x+h) is not far from f(x), or else the approximation fails.

<u>In fact, as h tends to zero the approximate becomes accurate.</u>

Differentiating (i) with respect to h

$F'(x) = A1+ 2A_2{*}h + 3A_3{*}h^2..\ \ nA_n{*}h^{n-1}$(putting h = 0)

$F'(x+h) = A1$

Similarly $A2 = F''(x)/2$, $A3 = F'''(x)/3!$, $An =F^n(x)/n!$

So $F(x+h) = F(x) + F'(x)\ h + F''(x)\ h^2/2 + F'''(x)\ h^3/3! + \ldots\ldots.\ F^n(x)\ h^n /n!$…….. (ii)

We can see that since h is small, h^2, h^3 etc become very small. So in most cases terms upto <u>h^2 .i.e. 2^{nd} term is good enough</u>.

We know that maxima or minima the F’(x) term is zero. F’(x) signifies that function is increasing or decreasing, and if function is increasing or decreasing at a point then it can’t be maxima.

Proof Analytical

For example if at point x_0 the $F'x_0$ is negative then Function can’t be maxima as value of F(x) just before x_0, would be greater than it’s value at xo.

Similarly it can’t be minima as it’s value just after x_0, would be lower than it’s value at x_0.

Written in forms of equation:-

$F(x_0+h) = F(x_0) + vF'(x)$..(linear approximation)

$F'(x) < 0$

$\Rightarrow F(x_0) > F(x_0+h)$ (so not minimum)

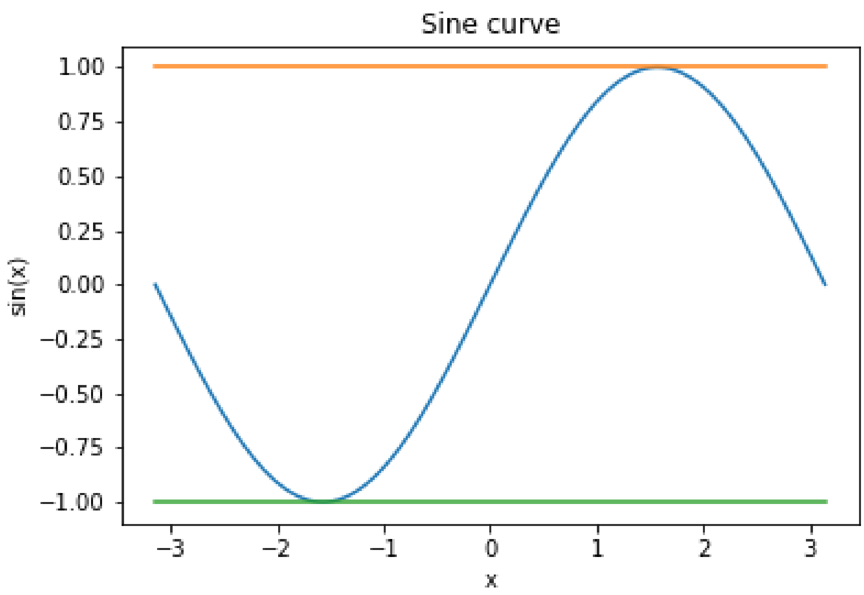
Can it be maximum

If $h < 0$

$V_x < 0$

- ⇨ *$V_x F'(x) > 0$*
- ⇨ *$F(x_0+h) > F(x_0)$ so $F(x_0)$ can’t be maximum also.*

Proof geometric



F(x) = sin(x)

We can see that only at maxima and minima that slope of tangent , that is rate of change of F(x) .i.e. F'(x) is zero.

Why second derivative negative mean maxima?

From equation (ii)

$F(x+h) = F(x) + F'(x)\, h + F''(x)\, h^2 /2$

$F'(x) = 0$

So it reduces to , $F(x+h) = F(x) + F''(x)\, h^2 /2$

If F(x) is maxima then it's greater than F(x+h) , a point in its neighborhood.

So, $F(x+h) - F(x) = F''(x)\, h^2 /2$

$F''(x)\, h^2/2$ = '-' *ve*

h^2 *is always positive implying F''(x) is negative.*

The reverse is true in case of minima.

TAYLOR SERIES(TWO VARIABLE)

$F(x+h,y+k) =?$

Writing it in parametric form $x = a + ht$, $y = b + kt$

$dx/dt = h$, $dy/dt = k$

So now x and y become function of t.

$Z = F(x,y) = g(t)$

$Z' = F_x(x,y)*dx/dt + F_y(x,y)* dy/dt$ ……(iii)

$= F_x(x,y)*h + F_y(x,y)*k$

$Z'' = F_{xx}(x,y)*h*h + F_{yy}(x,y)*k*k + 2hkF_{xy}(x,y)$………(iV)

$= F_{xx}(x,y)*h^2 + + F_{yy}(x,y)*k^2 + 2hkF_{xy}(x,y)$

$F(x+h,y+k) = Z + Z' + Z''/2!$

$= F(x,y) + F_x(x,y)*h + F_y(x,y)*k +(F_{xx}(x,y)* h^2 + + F_{yy}(x,y)*k^2 + 2hkF_{xy}(x,y))/2$..........(V)

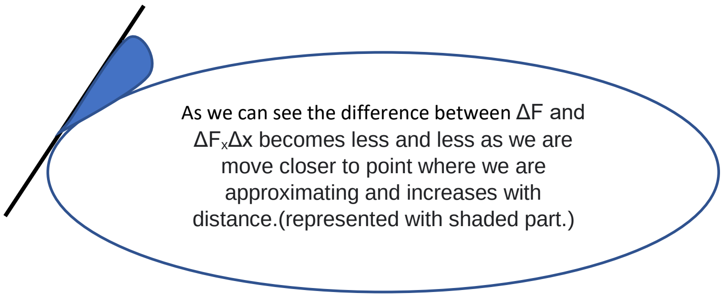
In case you are wondering how differentiation resulted in (iii)…refer chain rule

Let F be a function of (x,y) and x,y be function of t

For small changes $\Delta F \approx \Delta F_x \Delta x + \Delta F_y \Delta y$ (linear approximation)

Dividing everything by Δt, we get
$\Delta F/ \Delta t = \Delta F_x \Delta x/ \Delta t + \Delta F_y \Delta y/ \Delta t$
For smaller and smaller values we get more accurate results, and at dx,dy it becomes accurate(provided function is continuous and differentiable).

Hence the result.
$F'(x,y) = F_x(x,y)*dx/dt + F_y(x,y)* dy/dt$
The following figure would clarify it further(though it's only one variable)



So, now we are left with equation V.. we have neglected higher order terms.

$F(x+h,y+k)- F(x,y)= F_x(x,y)*h + F_y(x,y)*k +(F_{xx}(x,y)*h^2+ + + F_{yy}(x,y)*k^2 + 2hkF_{xy}(x,y))/2$

$F_x(x,y),= F_y(x,y) = 0$. For function to be minima or maxima(local or global) it's slope has to be zero and incase of two variables this has to be zero with respect to both the variables.

$F(x+h,y+k)- F(x,y) =$ $\underline{(F_{xx}(x,y)*h^2 + + F_{yy}(x,y)*k^2 + 2hkF_{xy}(x,y))/2}$

So Δxy depends solely on 2nd order terms(generally) in case of points whose 1st order diffrentials are zero.

$(F_{xx}(x,y)*h^2 + + F_{yy}(x,y)*k^2 + 2hkF_{xy}(x,y))$ *is of form* $(ax^2+ by^2+2cxy)$

$a = F_{xx}(x,y) ,b = F_{yy}(x,y), c = F_{xy}$

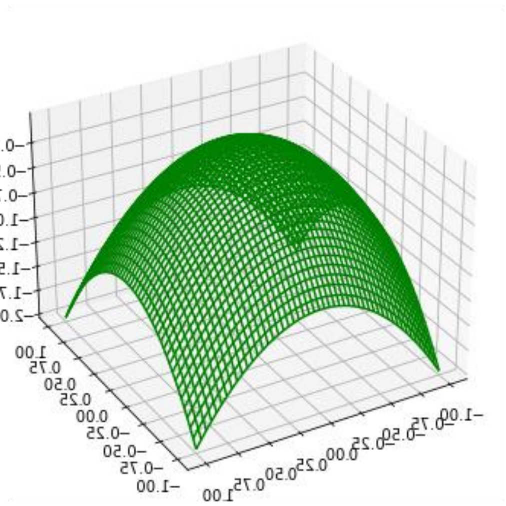
ax2 + by2+2cxy can be written as $1/a(a^2x^2 + 2acxy + aby^2)$

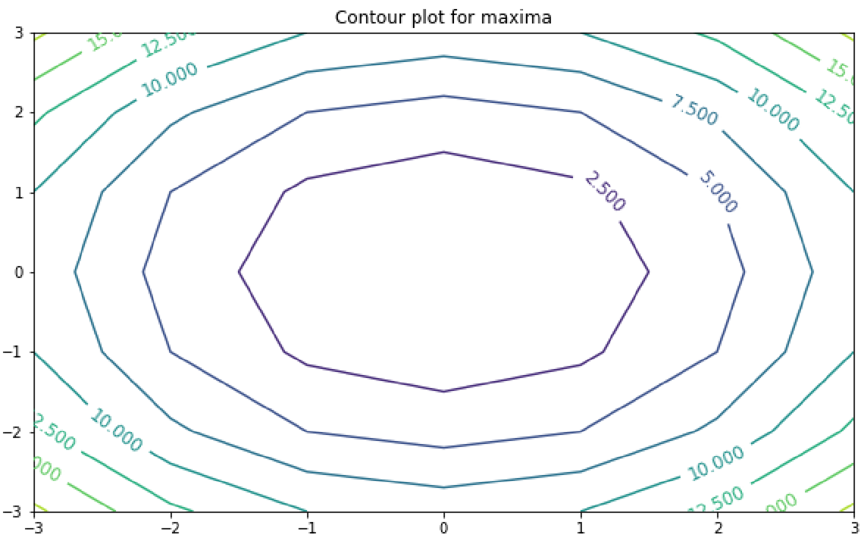
⇨ $1/a((ax+cy)^2 + (ab-c2)y^2)$

If $ab - c^2$ is positive, then the sign of above term depends solely on a and we can say with certainty that function is going to be maxima or minima.

The case of maxima.

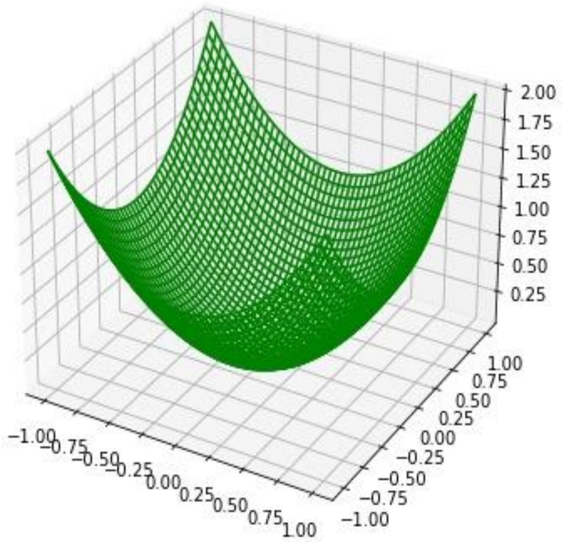
MAXIMA



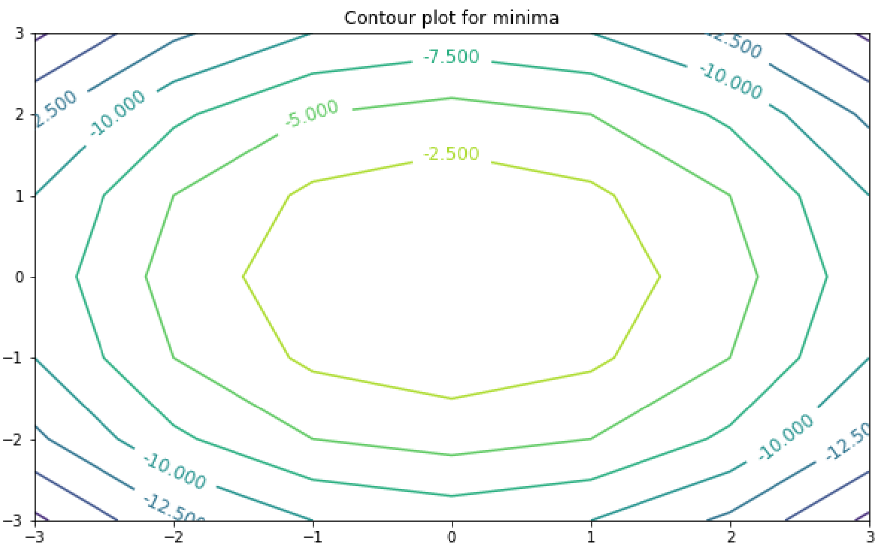


a or (F_{xx}) is negative. Function is maximum at the point.

MINIMA

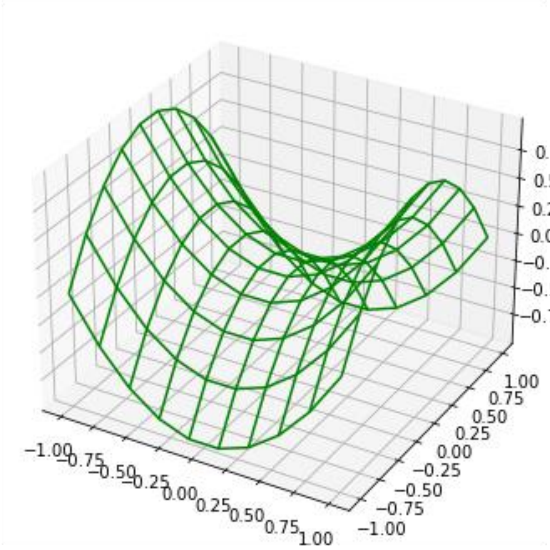


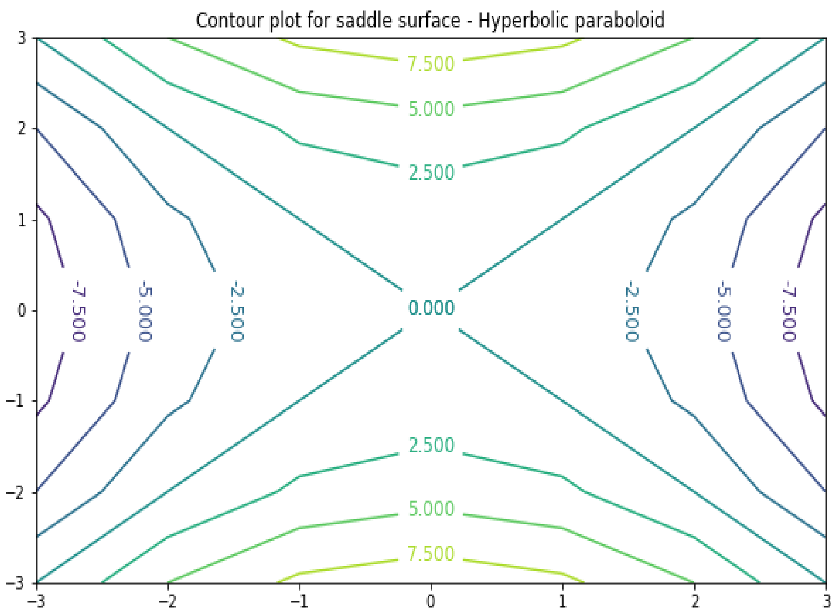
This specific plot is of least squares and we can see that we don’t have to worry about local minima or maxima in that case



a or F_{xx} is positive. Function is minimum at the point.

SADDLE

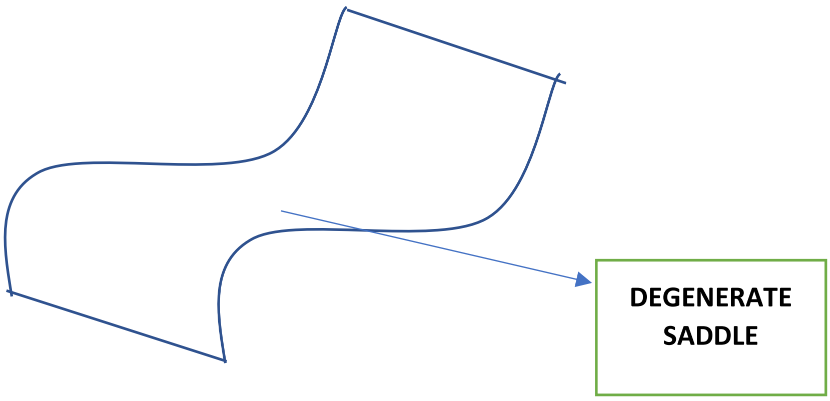




We cam see that as we move along x, we reach the minima and along Y we reach maxima.

If $F_{xy}^2 - F_{xx}*F_{yy}$ is negative, then we cannot really say about sign of above term as it can be both positive or negative, it's a saddle point in which the direction we move determine the value of F(x,y)

What about ab – c2 =0, in that case , it's the case of degenerate saddle, where though Fx, Fy is zero the function can not be classified as maxima or minima and we need higher order differentials to determine ΔF. However if Fxx is positive we can rule out it being maxima and if Fxx is negative then we can rule out it being minima.



HESSIAN MATRIX AND MORE THAN 2 VARIABLES

So for function to be maxima or minima $F_{xy}^2 - F_{xx} * F_{yy}$ has to be positive, if negative it's not definite.

We got here from the expression $ax^2 + by^2 + 2cxy$.

Any term of the form $ax^2 + by^2 + 2cxy$ or pure quadratic (2^{nd} order terms from Taylor approximation)can be expressed as.

$|x \quad y| \begin{matrix} a & c \\ c & b \end{matrix} \begin{vmatrix} x \\ y \end{vmatrix}$ *.i.e. X^TAX*

X^TAX is the 2^{nd}order term in Taylor series approximation.

The matrix A is called as Hessian matrix

The expression X^TAX is applicable for n terms too. Thus we can now deal with maxima and minima in n variables.

Hessian matrices are symmetric matrices and thus has real eigen values.

The matrix A has to be positive definite for Minima (concave up) or negative definite for maxima (concave down)to exist.

Positive definite means all eigen values are positive which means X and AX are in same direction, which means X^TAX is positive, which means $\Delta_{(x1,x2,x3..xn)}$ is positive implying minima.

Similarly negative definite implies maxima..

F has a minimum when the pure quadratic = $x^T A x$ is positive.

If A has both positive and negative eigen values then we can't really say and is a saddle point and if A is singular then we have degenerate case.

More on eigen values and vector when I deal with PCA.

Note : -

F' is 1st order differential, F'' is 2nd order.

F_x is 1st order differential with respect to x

Code for plots is uploaded at https://github.com/ML-NOOB