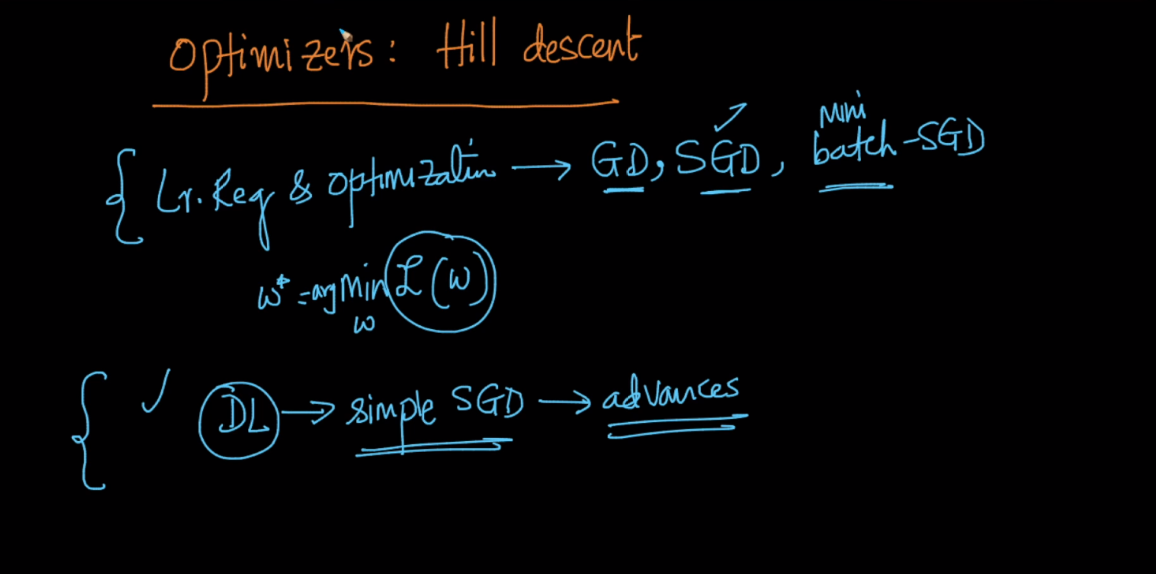
**Optimizers:Hill-descent analogy in 2D**

In logistic regression & optimization we use Gradient descent, SGD, mini batch-SGD

And using this we find optimal w which minimizes loss

But in Deep learning we use more advances techniques.

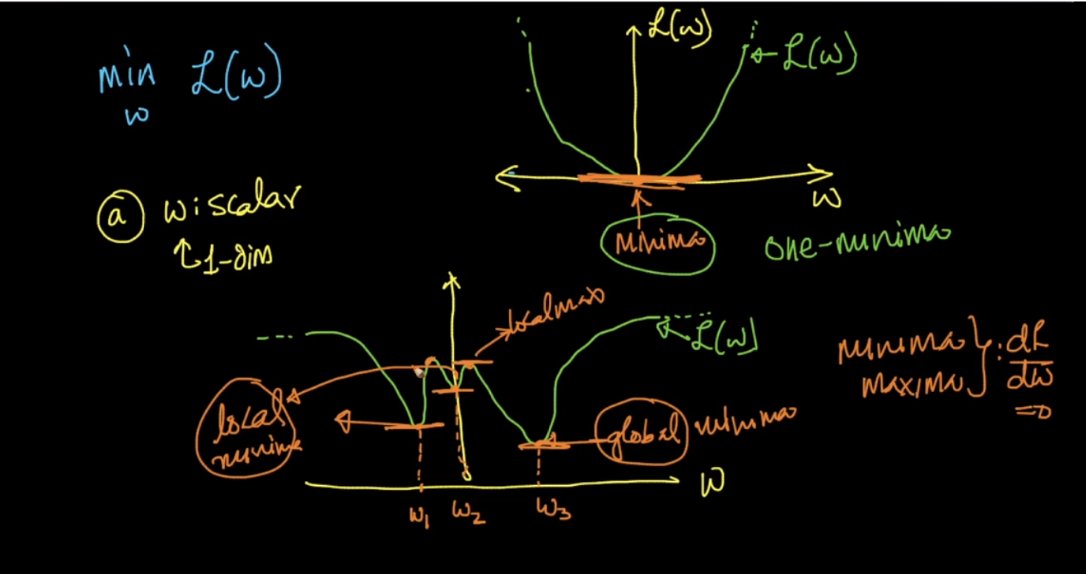


Below image describe about local minima and global minima by graph which have w at x-axis and loss l(w) at y-axis.

In first graph there is only one minima therefore here local minima = global minima.

In another graph there is multiple local minima(w1, w2) and only one global minima w3.

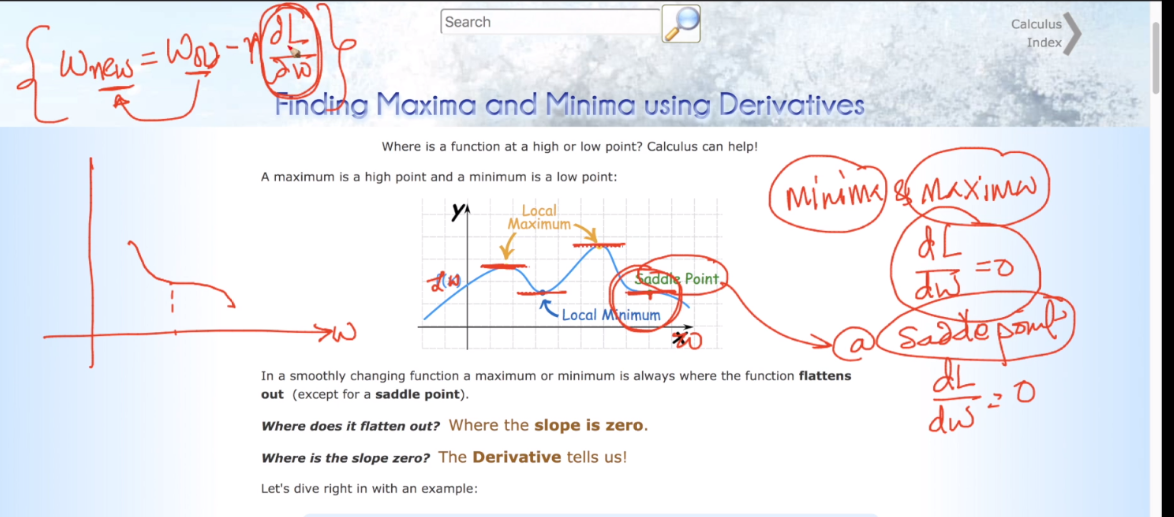
So how can we find that minima and maxima : we can find that by tangent tangent of slope=0 i.e derivative/gradient of loss with respect to w at that point is 0 then it is minima or maxima



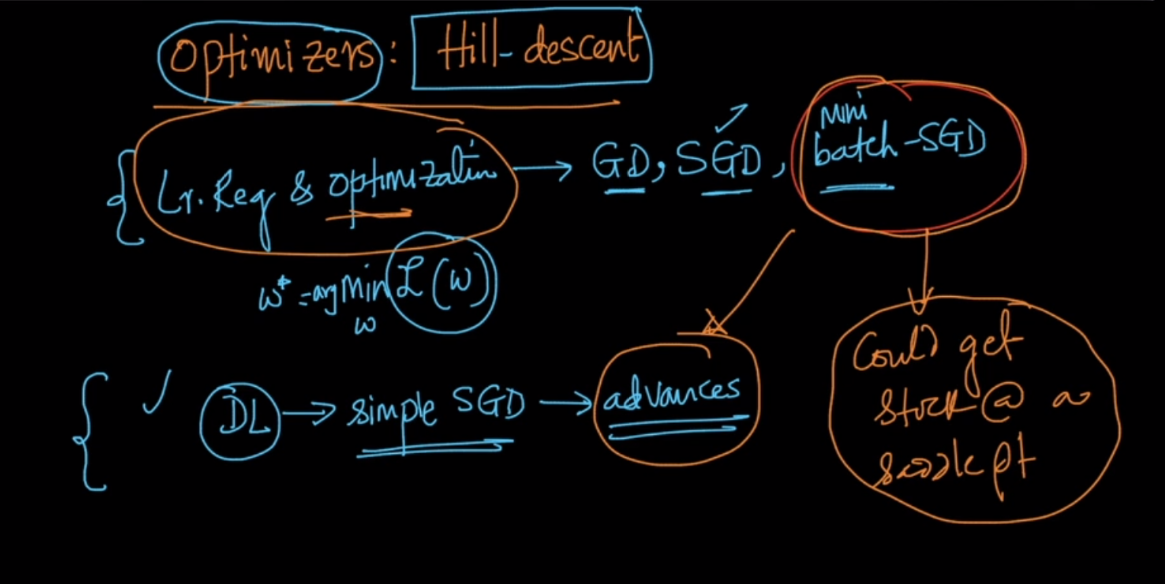
In below image it shows local minima, maxima and saddle point as well.

So like minima and maxima derivative d\_L/d\_w = 0 at saddle point also

And by this we can’t update w as w\_new = w\_old and we stuck here and we can find global minima.



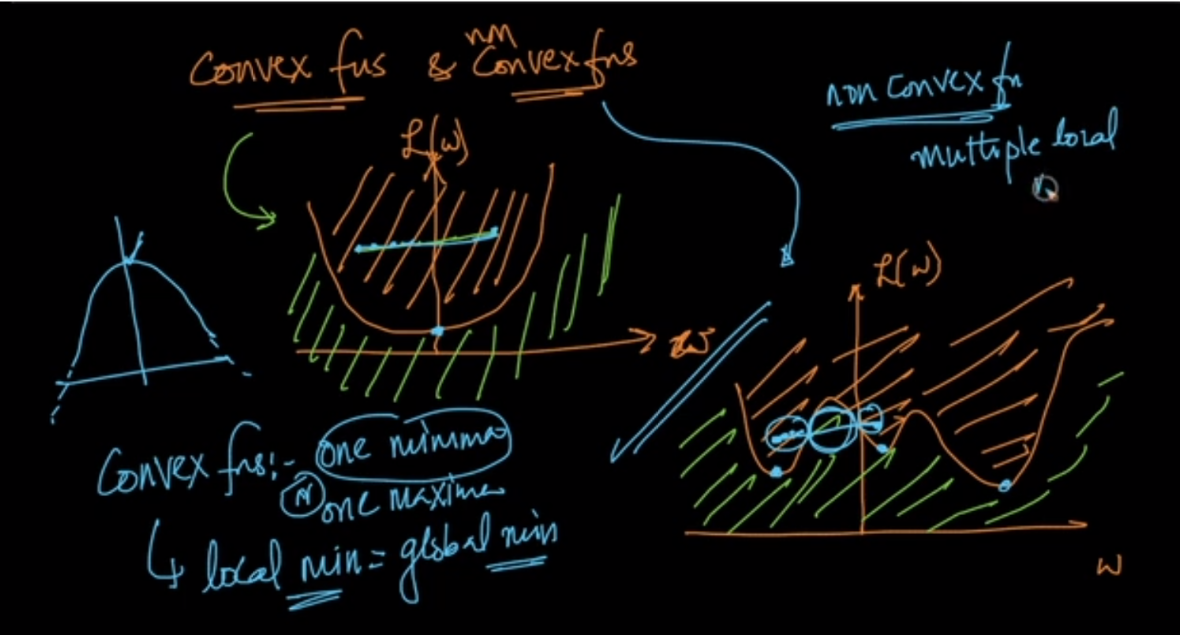
As gd, sgd, mini batch-sgd could get stuck at a saddle point therefore we use advances for deep learning.



Convex function : First graph in below image shows convex function because if we take any two point in any region of this graph (here we take in orange region) then the points on shortest line connecting this two points are also in the same (orange) region then it is convex function. Convex function have either one minima or one maxima and also local minima = global minima

Non-convex function : second graph showing non-convex because as we take two points in two regions but the points in the shortest line connecting this two points are not only in orange region they are also in green region therefore this function is non-convex function.

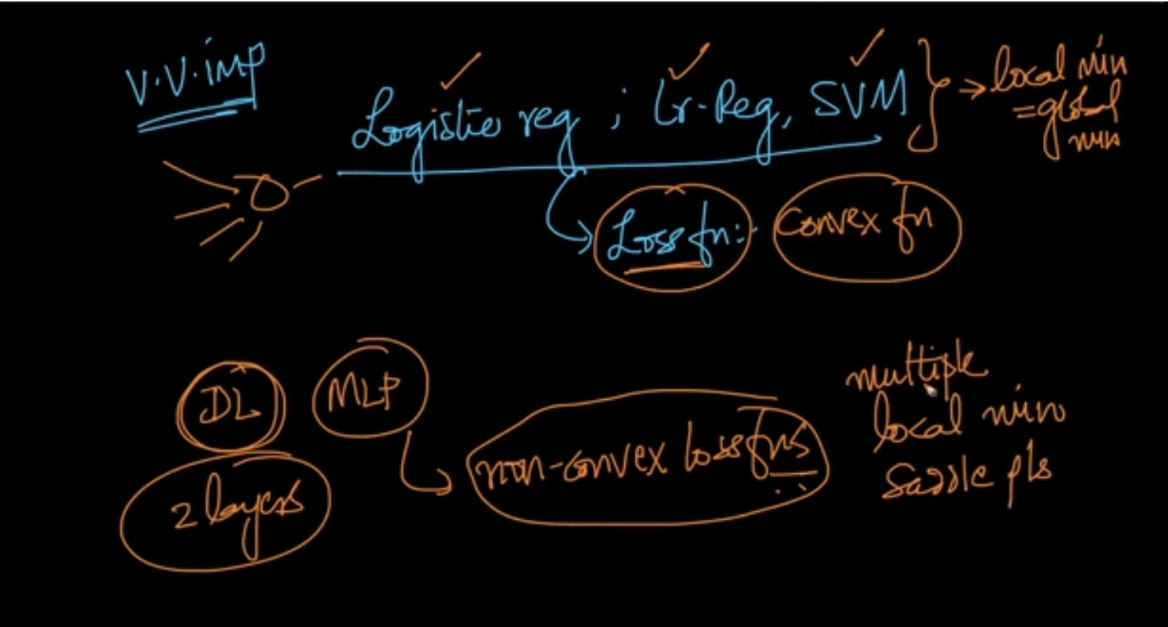
Non-convex function have multiple local minima.



Logistic reg. linear reg. SVM’s loss function are convex function therefore it have only one minima i.e local minima = global minima therefore we could not get stuck in local minima or saddle points.

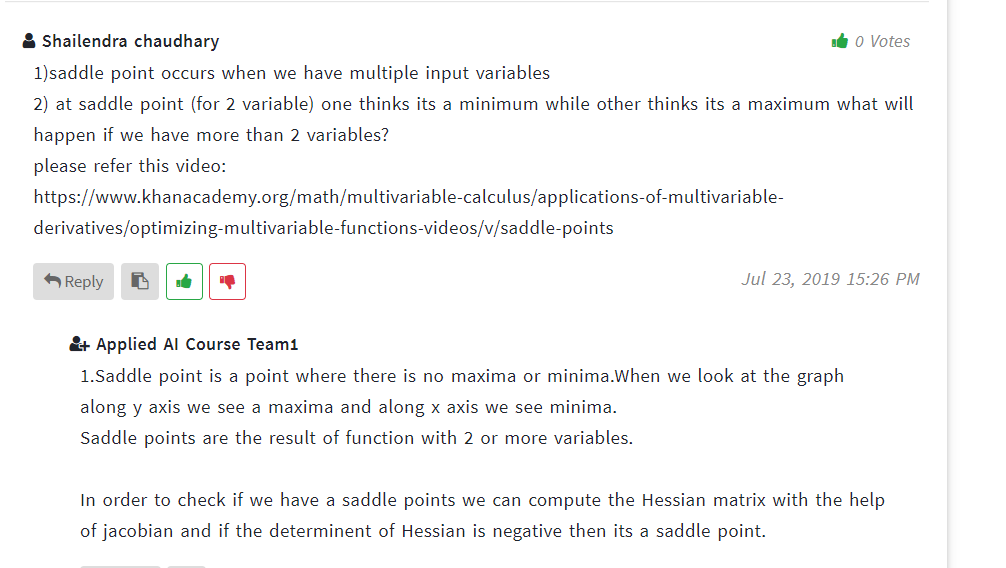
Logistic regression can also be an neural network but with single neuron only.

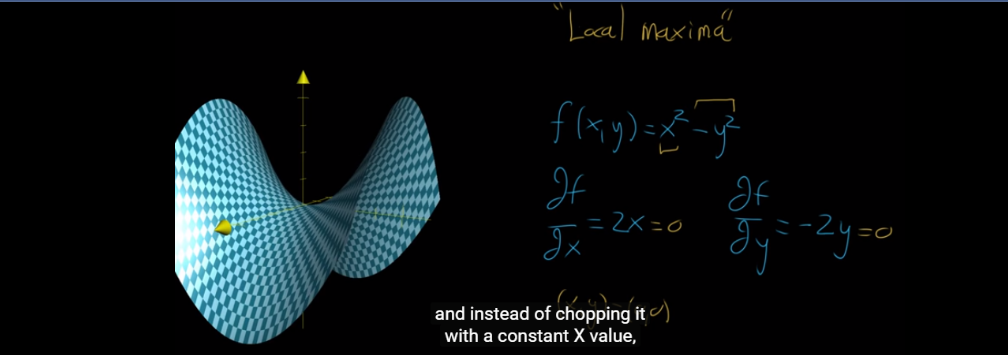
But neural networks of 2 or more layer like MLP have non-convex loss function therefore it have multiple local minima or saddle points.

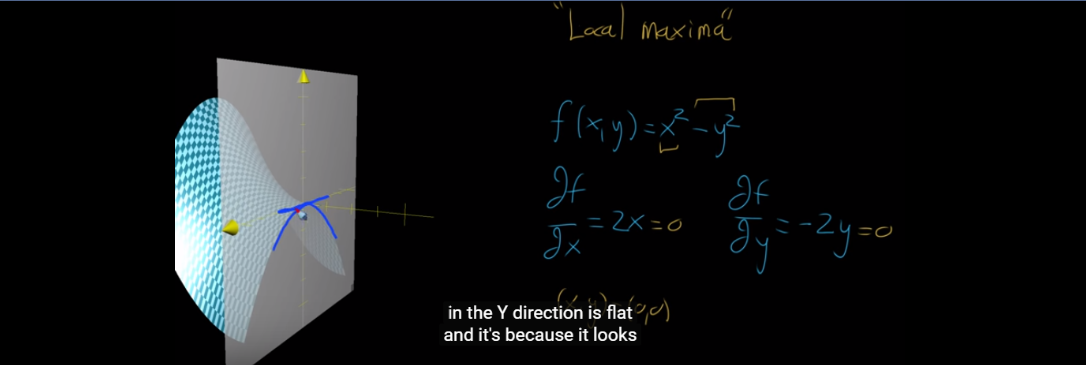


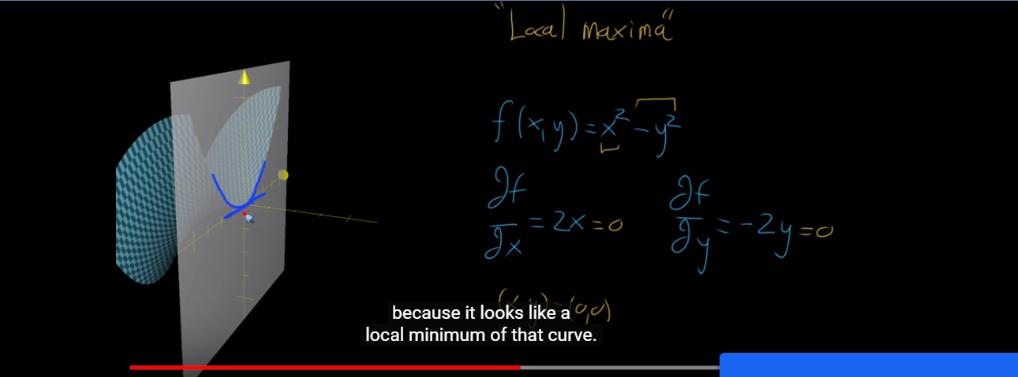
Saddle point :

<https://www.khanacademy.org/math/multivariable-calculus/applications-of-multivariable-derivatives/optimizing-multivariable-functions-videos/v/saddle-points>









Comments :

