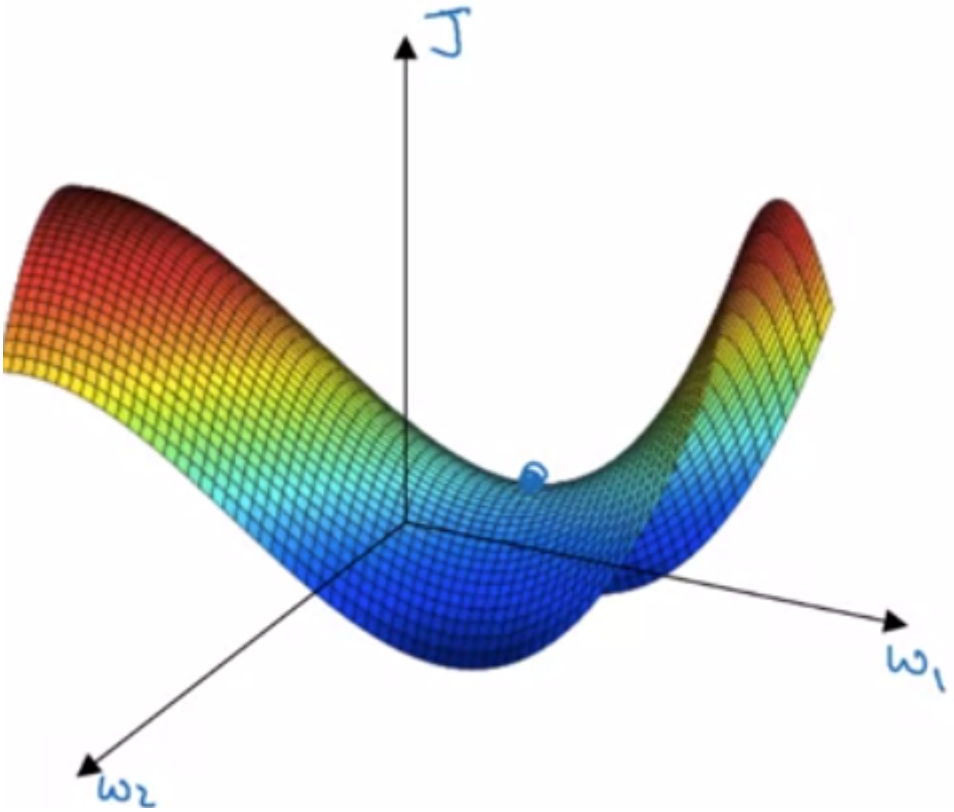
Week 2-4 Local Optima and Saddle Points

笔记本:	DL 2 - Deep NN Hyperparameter Tunning, Regularization & Optimization		
创建时间:	2021/1/9 11:41	**更新时间:**	2021/1/9 11:47

It turns out if you create a neural network, most points of zero gradients are not local optima like points like this. Instead most points of zero gradient in a cost function are saddle points.



In mathematics, a saddle point or minimax point is a point on the surface of the graph of a function

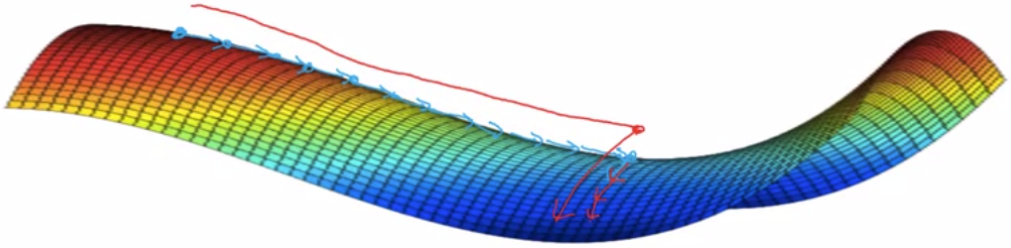
where the slopes (derivatives) in orthogonal directions are all zero (a critical point), but which is not a local extremum of the function.

The name derives from the fact that the prototypical example in two dimensions is a surface that **curves up in one direction**, and **curves down in a different direction**, resembling a riding saddle or a mountain pass between two peaks forming a landform saddle

Saddle points are more rare in high dimensions, so for a NN with many parameters, it is unlikely to fall into saddle points.

But the problem of plateaus affect the training rate a lot

Problem of plateaus



- Unlikely to get stuck in a bad local optima
- Plateaus can make learning slow