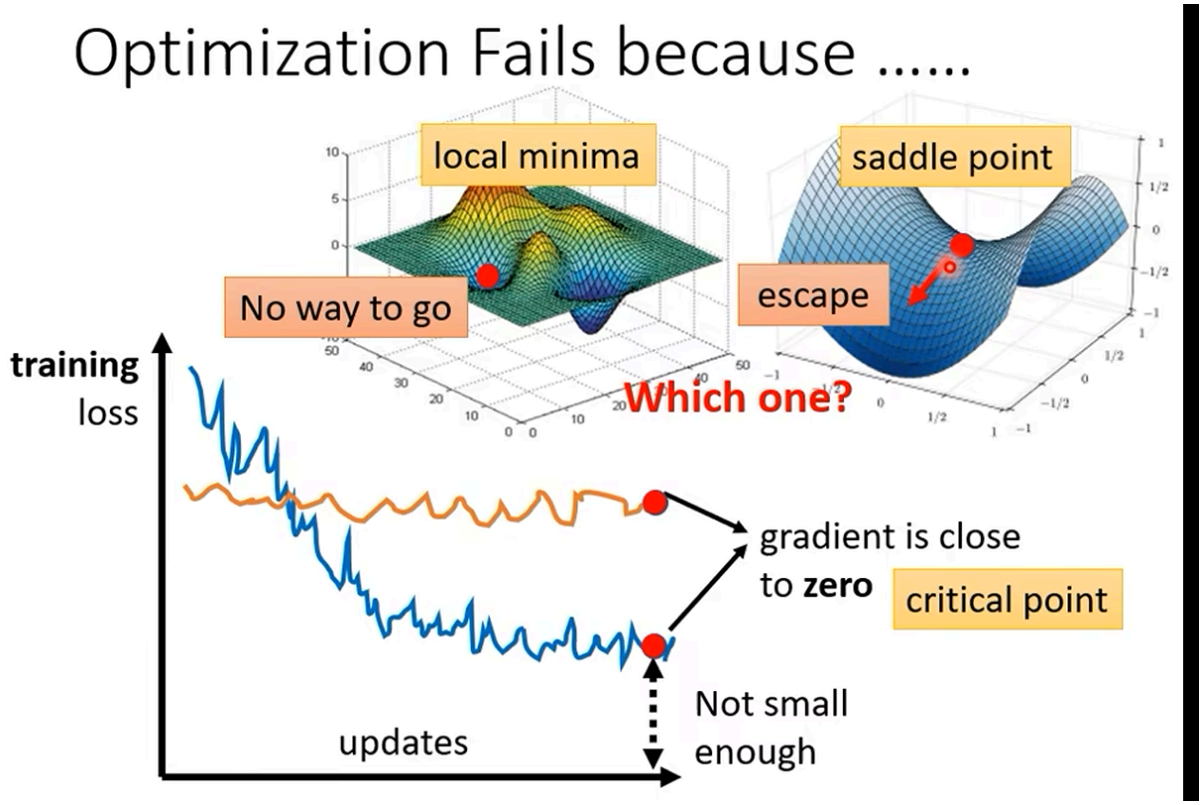
类神经网络训练不起来怎么办（一）局部最小值和鞍点

optimization失败的原因：

loss不再变化不一定是局部最小值，也可能是saddle Point（鞍点），梯度为0的点统称为critical Point。loss不变是卡在了critical Point。

local minima 是无路可走的，而saddle Point是可以逃离的，可以使你的loss变低。



根据hessian判断当前的critical Point是什么类型：

Hessian

At critical point:

$$L(\boldsymbol{\theta}) \approx L(\boldsymbol{\theta}') + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}')^T H(\boldsymbol{\theta}-\boldsymbol{\theta}')$$

$\boldsymbol{v}^T H \boldsymbol{v}$

For all $\boldsymbol{v}$

$\boldsymbol{v}^T H \boldsymbol{v} > 0$ ➡ Around $\boldsymbol{\theta}'$: $L(\boldsymbol{\theta}) > L(\boldsymbol{\theta}')$ ➡ **Local minima**

= H is positive definite = All eigen values are positive.

For all $\boldsymbol{v}$

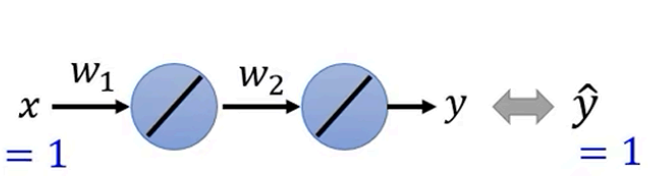
$\boldsymbol{v}^T H \boldsymbol{v} < 0$ ➡ Around $\boldsymbol{\theta}'$: $L(\boldsymbol{\theta}) < L(\boldsymbol{\theta}')$ ➡ **Local maxima**

= H is negative definite = All eigen values are negative.

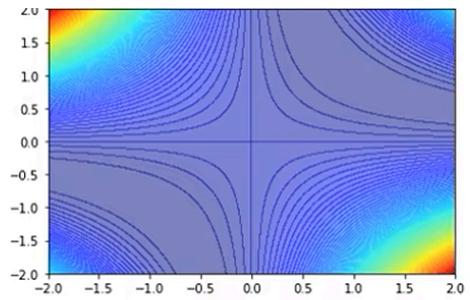
Sometimes $\boldsymbol{v}^T H \boldsymbol{v} > 0$, sometimes $\boldsymbol{v}^T H \boldsymbol{v} < 0$ ➡ **Saddle point**

Some eigen values are positive, and some are negative.

举例：



$$L = (\hat{y} - w_1 w_2 x)^2 = (1 - w_1 w_2)^2$$



$$g = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_2} \end{bmatrix} = \begin{bmatrix} 2(1-w_1 w_2)(-w_2) \\ 2(1-w_1 w_2)(-w_1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Critical point: $w_1 = 0, w_2 = 0$

$$H = \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix} \quad \lambda_1 = 2, \lambda_2 = -2$$

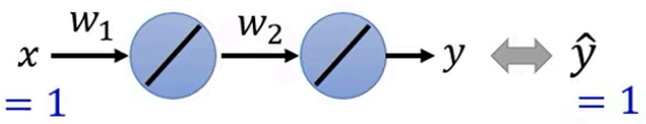
Saddle point

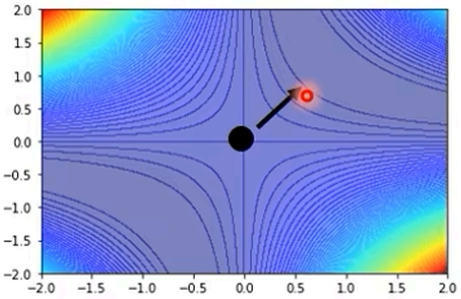
$$H = \begin{bmatrix} \frac{\partial^2 L}{\partial w_1^2} & \frac{\partial^2 L}{\partial w_1 \partial w_2} \\ \frac{\partial^2 L}{\partial w_2 \partial w_1} & \frac{\partial^2 L}{\partial w_2^2} \end{bmatrix} = \begin{bmatrix} 2(-w_2)(-w_2) & -2 + 4w_1 w_2 \\ -2 + 4w_1 w_2 & 2(-w_1)(-w_1) \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix}$$

这里的g（梯度）就是一个向量（绿色的）

这里的H就是由二次偏导组成的，由H有正有负，所以（0,0）为鞍点。

那么如何逃出鞍点：





$$L = (\hat{y} - w_1 w_2 x)^2 = (1 - w_1 w_2)^2$$

$$\frac{\partial L}{\partial w_1} = 2(1 - w_1 w_2)(-w_2)$$

$$\frac{\partial L}{\partial w_2} = 2(1 - w_1 w_2)(-w_1)$$

Critical point: $w_1 = 0, w_2 = 0$

$$H = \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix} \quad \lambda_1 = 2, \lambda_2 = -2$$

Saddle point

$\lambda_2 = -2$ Has eigenvector $\boldsymbol{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Update the parameter along the direction of $\boldsymbol{u}$

You can escape the saddle point and decrease the loss.

求H的特征值，找到负的特征值对应的特征向量，沿着特征向量的方向就可以减少loss，但是H的计算很复杂，实践中几乎没有人用这个方式逃离鞍点。