正则化的作用及 L_1，L_2 正则化的区别

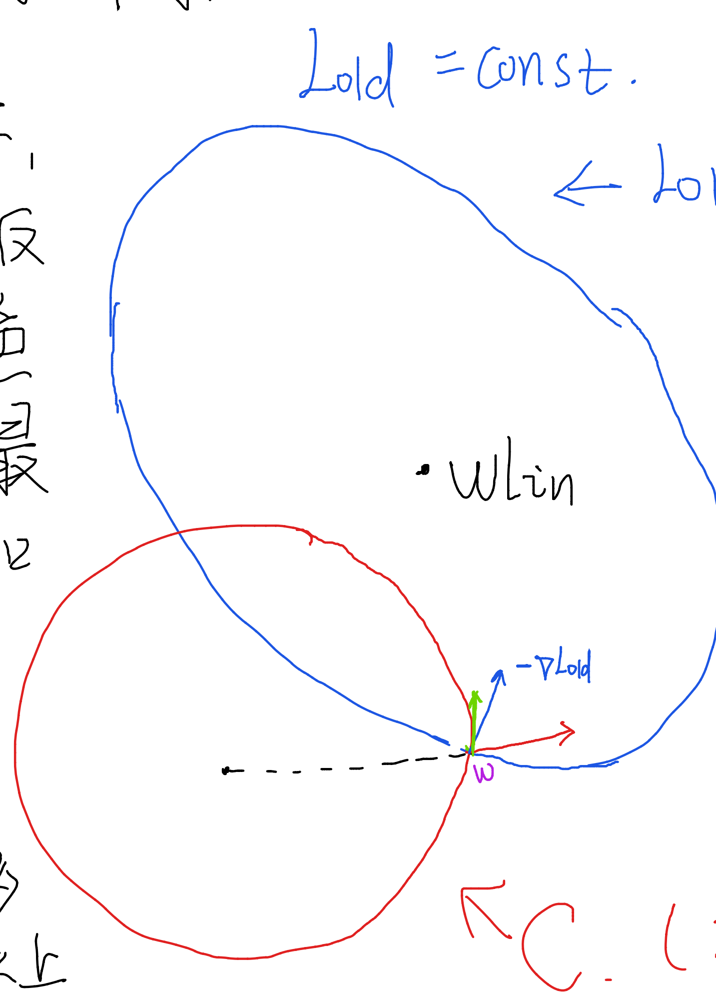
正则化的作用：限制参数过多/过大，避免模型更加复杂。将某些高阶 w 限制为0，最直观的做法就是限制 w 的个数，寻找比限制 w 更宽松的条件：

$$\sum_j w_j^2 \le C \quad / \quad \sum_j |w_j| \le C.$$

L_2 正则化： $[L(w) = L_{old}(w) + \lambda \sum_j w_j^2]$.

Loss function： $L = L_{old} + \lambda \sum_j w_j^2$.

$\iff$ 在约束 $\sum_j w_j^2$ 下，使 L_{old} min，使用拉格朗日乘子法，就等价为 L.



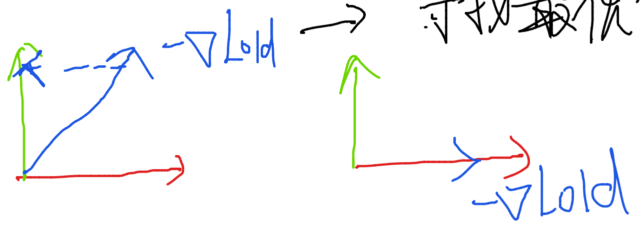
在不加约束的条件下，w 会沿着 L_{old} 梯度反方向 $-\nabla L_{in}$ 在蓝色区域中寻找全局最优值 w_{lin}。但加上约束后，w 只能在红色区域内进行搜索，最多只能在红圆的边缘上沿着切线进行变化.（绿色箭头）.

[正则化约束 w 的解空间，使不会太大，也不会太小，更可以产生 w=0 了.

C.（对 w 进行约束的空间，限制 w 的变化不能离开红色空间）

w 运动的方向与 w 的方向（红色箭头）垂直。若在约束下，找到 w 的最优值，也就是 w 不再变化，也就是蓝色箭头方向与红色箭头平行。不然蓝色会在绿色方向有分量，则 w 会继续运动

最优.



寻找最优解的过程

$$-\nabla L_{old} + \lambda w = 0 \iff \nabla L_{old} + \lambda w = 0$$

$L'(w) =$

$\Leftarrow$ 为 $f'(\frac{\lambda}{2} w^2)$.

$$\iff \boxed{L = L_{old} + \frac{\lambda}{2} w^2}$$

（L_2 正则的由来）.

https://blog.csdn.net/wfei101/article/details/82530591