

# 正则化原因和L\_1, L\_2正则化异同

## 正则化原因

减小参数的值，防止参数波动太大（拟合一些异常的点），造成过拟合现象。

## L\_1, L\_2正则化的异同

- 相同：都可以减小参数值，防止过拟合
- 差异：L\_1公式为

$$||x||_1 = |x_1| + |x_2| + \dots + |x_n|$$

L\_1范数优化趋向于产生少量特征，即减少非零元个数

L\_2公式为

$$||x||_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

L\_2范数优化趋向于减小每个参数的值，获得小值参数。

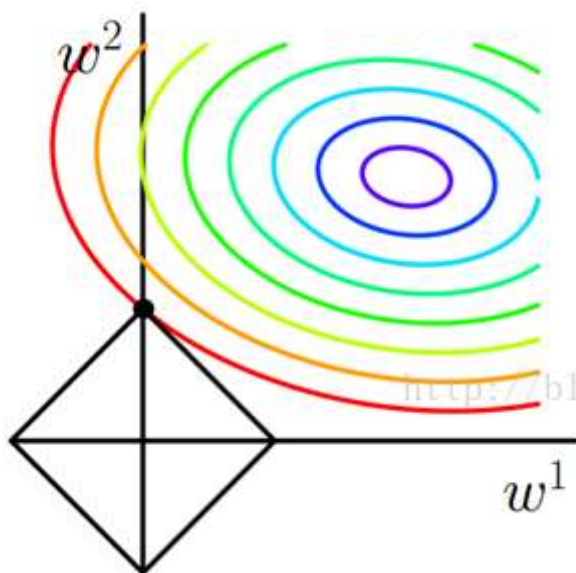
## 例子

优化目标函数

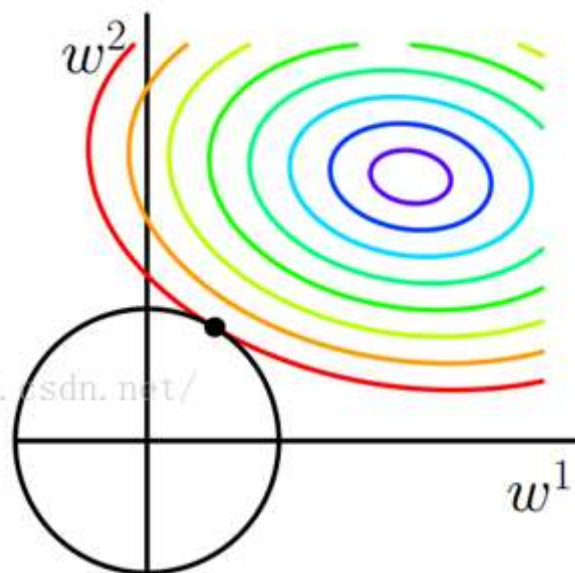
$$Lasso: \min_w \frac{1}{n} \|y - Xw\|^2, \quad s.t. \|w\|_1 \leq C$$

$$Ridge: \min_w \frac{1}{n} \|y - Xw\|^2, \quad s.t. \|w\|_2 \leq C$$

最优值获得在第一次交点处



(a)  $\ell_1$ -ball meets quadratic function.  
 $\ell_1$ -ball has corners. It's very likely that the meet-point is at one of the corners.



(b)  $\ell_2$ -ball meets quadratic function.  
 $\ell_2$ -ball has no corner. It is very unlikely that the meet-point is on any of axes.

**分析：**可以看出L<sub>1</sub>范数极大可能在尖点，也就是在与坐标轴交点处取得最优值，这种地方往往某些维度的特征取值为0，而L<sub>2</sub>范数则倾向靠近各个分量特征都比较小的取值