解释：L1正则化对参数的的稀疏性效果

“相比L2正则来说，L1正则更可能产生稀疏的参数”。这个情况可以从不同的视角来解释。主要有两种视角：第一种视角是“基于约束优化的图形解释”，第二种视角是“概率解释”。

### 1.基于约束优化的图形解释

目标函数加上（L1或L2）正则化，等价于约束最优化问题。为了简化讨论，假设目标函数是凸函数，且参数只有两维。则L1正则化的约束空间的边界是正方形，L2正则化的约束空间的边界是圆形。如下图所示。

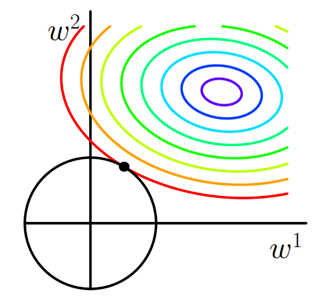
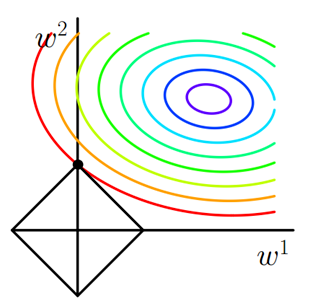


图1 L1和L2正则化的等价约束优化图解

由约束优化的意义可知，最优解必然在目标函数等值线和约束空间边界的相切点处。而L1的形状使得其相切点更容易产生在坐标轴上，也就是或为0。

### 2.概率解释

为了简化讨论，假设目标函数是MSE损失函数。

MSE损失函数可以从两种准则出发进行构建，两种准则殊途同归。第一种准则就是最小化均方误差；第二种准则则是极大似然。

（1）基于极大似然准则构建目标函数

我们回想一下用极大似然准则构建目标函数。假设模型能很好地拟合数据，则模型的预测值与真实值之间的误差服从0均值的高斯分布：

则样本的似然概率为：

自然地，可以用极大似然原则构建目标函数：

可以取对数，化简：

这就得到了MSE损失函数。

（2）高斯（Gauss）先验与拉普拉斯（Laplace）先验

参数的先验分布若为Gauss分布，则有：

参数的先验分布若为Laplace分布，则有：

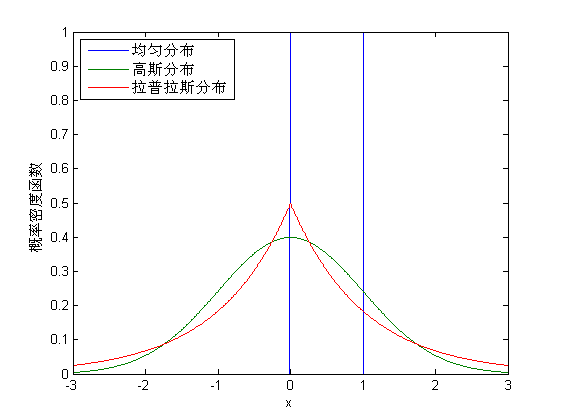


图2 Gauss分布与Laplace分布

其中，表示的分量。

若使用极大后验原则构建目标函数，即在极大似然的基础上，加上参数的先验，则：

然后取对数并化简，对于Gauss先验，得到：

这就是L2正则化。

对于Laplace先验，得到：

这就是L1正则化。

由于目标函数都要求参数尽量小，因此参数会在0附近。又因为Laplace分布在0处比较尖，因此参数在0附近时，等于0的概率比较大；而Gauss分布在0处比较圆滑，因此参数在0附近时，等于0的概率没那么大。

### 3.参考资料

<https://www.cnblogs.com/heguanyou/p/7688344.html>

<https://www.zhihu.com/question/37096933/answer/475278057>

<https://www.zhihu.com/question/37096933/answer/189905987>