一、L1和L2正则法则

1、L1范式（Lasso回归）和L2范式（Ridge回归）为什么可以看做损失函数的惩罚项？

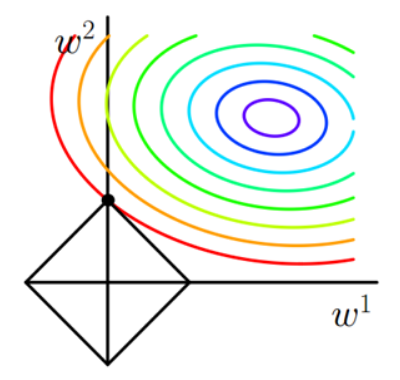
所谓“惩罚”是指对损失函数中胡某些参数做一些限制。

2、为什么L1正则化可以产生稀疏权值矩阵，即产生一个稀疏模型，可以用来特征选择，L2正则化可以防止模型过拟合（overfitting），一定程度上，L1也可以防止过拟合？

为什么L1正则化可以产生稀疏模型（L1是怎么让系数等于零的）？

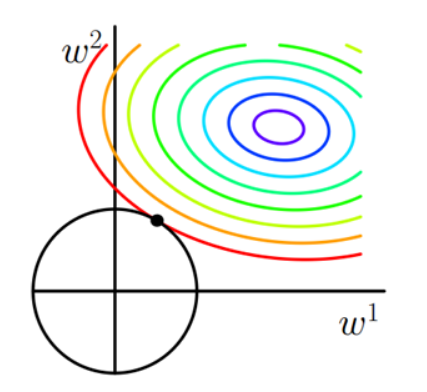
这里其实可以想象一个很极端的情况，只有两个参数，加号后面的部分可以绘制成一个图形，一个顶点在坐标轴上的矩阵，而前面部分是由无数条等直线组成，当等值线与矩阵相交时，即为最优值，而这个交点与在顶点的概率会远大于与其它部位接触的机率

而前面的参数用来控制的大小，当越大，就越小，甚至可以控制在原点附近。



为什么L2正则化可以防止过拟合？





与L1相比，L2加号后半部分是一个圆，所以最优点在坐标轴上的概率大大减小，这也是为什么L2不具备稀疏性。其中λ就是正则化参数。从上式可以看到，与未添加L2正则化的迭代公式相比，每一次迭代，θj都要先乘以一个小于1的因子，从而使得θj不断减小，因此总得来看，θ是不断减小的。

（为什么倾向于让参数比较小，想象一下数据偏移一点点，结果就变化很大，抗扰动能力，也在一定程度上避免了过拟合现象）