目录

[1. 机器学习的范数规则化 2](#_Toc485889367)

1. **机器学习的范数规则化**

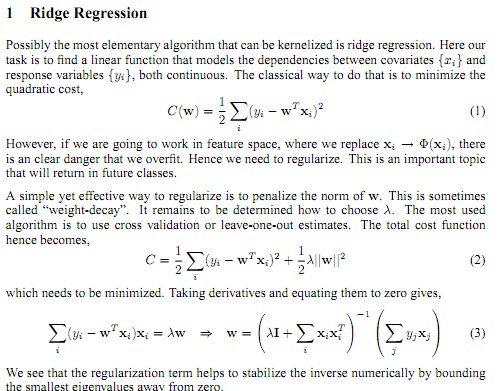
参考网址：http://blog.csdn.net/zouxy09/article/details/24971995

* 监督机器学习——minimize your error while regularizing your parameters，在规则化参数的同时最小化误差，后者是使模型拟合训练数据，前者是防止过度拟合。
* **规则项的作用：**
  + 目标不是训练误差最小，而是模型测试误差小，使得参数的泛化性能好
  + 人为加入先验知识到模型中，得到例如稀疏、低秩、平滑等人们想要的结果。从贝叶斯估计的角度来看，规则化项对应于模型的先验概率
* **目标函数：**



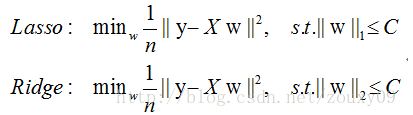
* + 第一项是Loss函数。Square loss就是最小二乘；Hinge loss就是SVM；exp-loss就是Boosting；log-loss就是Logistics Regression
  + 规则项可以是模型参数向量的范数
* **范数正则项**
  + L0：指向量中非0元素的个数，用它来规则化参数，使得参数向量稀疏。但是因为1）L0范数求解是NP难问题2）L1是L0的最优凸近似（@）
  + L1（稀疏规则算子）：
    - 含义是向量各个元素绝对值之和。任何规则化算子，如果它在Wi=0的地方不可微，并且可以分解为”求和“的形式，那么这个规则化算子可以实现稀疏。
    - 稀疏的两点好处：1）在训练数据中无用的参数如果不置0，在测试数据中就会影响性能2）有点降维的含义，便于后续分析
  + L2（回归里面叫“岭回归”）：
    - L2含义是向量各个元素的平方和然后平方根。用L2进行规则化，会使各个参数的值变小，但不会变为0。
    - 三个角度看好处：
      * 实现对模型空间的限制，防止过拟合，提升泛化能力
      * 矩阵求解方程组角度。有助于处理condition number不好情况下矩阵求逆困难的问题，即可以使优化求解变得稳定和快速。

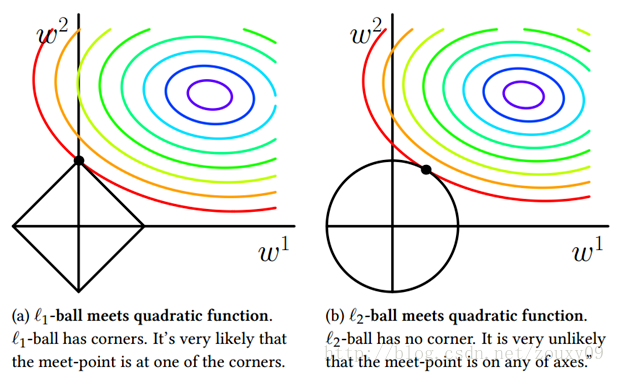
|  |
| --- |
| condition number是一个矩阵的稳定性或者敏感度的对量，当它在1附近时，则矩阵为well-condition，反之则为ill-condition。敏感是指对于一个线性方程组，如果y值有一点点变化，导致解出来的x值变化很大，这是我们不希望看到的。  可以联想到MMSE算法和ZF算法，前者之所以好是因为它是一个well-condition，容忍噪声度更大。 |



* 迭代求解方程组角度。加上规则化会使目标函数变成-strongly convex(强凸)，即在正确解附近的y值不会太相近，函数不会太平坦，这样收敛速度比较快。
  + 为什么L1是稀疏解，L2却不是？

L1和L2规则化的代价函数：





1. **生成模型和判别模型**

* 监督学习分为生成模型和判别模型
  + 生成模型：由数据学习**联合概率分布**，然后求出条件概率分布作为预测的模型。eg：朴素贝叶斯、隐马尔科夫(EM算法)、混合高斯
  + 判别模型：由数据直接学习**决策函数**或者**条件概率分布**作为预测模型。eg：KNN、感知机、决策树、逻辑斯蒂回归、最大熵、svm、提升方法、条件随机场。
* 另一种理解

后验正比于先验乘以似然。先验是输入的分布，似然是参数的分布。如果我们model先验和似然，然后乘起来得到后验，并做决策，这种模型叫生成模型。因为我们是用乘法“生成”了后验分布。

但是如果我们**不管先验和似然，直接假设后验的分布所满足的形式**（例如逻辑回归），这样学习到后验分布的方式就是“判别模型”