# 课程1-2

## 线性回归惩罚项的引入

课上不是通过“抑制线性拟合系数变化过大”引入惩罚项，而是通过在样本集上求解问题的解析解时，矩阵可逆的必要条件是矩阵必须可逆来引入的。

### 1.1 一般线性回归问题

具体来说，对于一个回归问题，假设有个样本，其中，那么如果使用特征的线性组合（线性回归）来拟合：



再以平方损失最为代价，单个样本代价为：



在整个数据集上的代价，用矩阵可以表示为：



其中：，

对关于参数向量求梯度，令梯度为0，解方程：



展开后得到：



如果是可逆的，那么就可以得到参数的解析解：



### 1.2 加入正则项的回归问题

样本矩阵由于这里矩阵是维矩阵因此可逆的必要条件是。那么如果样本数量少于样本特征数，这时可以通过引入正则项来使方程中原来不可逆的部分变得可逆。为了解决这个问题，在线性拟合函数中引入二次正则项，有



任然对加入正则项后的关于向量求梯度，令梯度为0解方程：



由于一定是半正定矩阵，且是正定矩阵，因此一定是正定矩阵（依据：正定矩阵加上一个半正定矩阵得到的和一定是正定矩阵），而正定矩阵一定是可逆矩阵，因此方程左侧的系数矩阵是可逆的，于是



### 1.3 术语

使用二次正则项（2-范数）的线性回归称之为脊回归或岭回归（ridge regression）；

使用绝对值正则项（1-范数）的线性回归称之为LASSO回归。

## 线性回归中的两个原理

假设回归问题：



其中代表损失函数（Loss），代表参数惩罚项（Penalty）。

**原理1**（指数打分原理）. 可以通过类似Softmax的变换将损失函数变为一个似然函数，例如：。

**原理2**（惩罚对偶原理）. 即参数惩罚项一定对应于参数极大后验估计中的某种参数概率先验。

由于存在上述两个原理，因此有时也把上面的G的表达式称之为正则化框架（regularized framework）