局部加权线性回归

局部加权线性回归算法，是一种非参数学习法（non-parametric）：

$$J(\theta)=\sum_{i=1}^{m}w^{(i)}\left[y^{(i)}-\theta^{T}x^{(i)}\right]^{2}\quad 其中：w^{(i)}=e^{\frac{\left(x^{(i)}-x\right)^{2}}{2k^{2}}}$$

对于上述公式的理解是这样的：x为某个预测点，x^((i))为样本点，样本点距离预测点越近，贡献的误差越大（权值越大），越远则贡献的误差越小（权值越小）。关于预测点的选取，在我的代码中取的是样本点。其中k是带宽参数，控制w（钟形函数）的宽窄程度，类似于高斯函数的标准差。

算法思路：假设预测点取样本点中的第i个样本点（共m个样本点），遍历1到m个样本点（含第i个），算出每一个样本点与预测点的距离，也就可以计算出每个样本贡献误差的权值，可以看出w是一个有m个元素的向量（写成对角阵形式），代入上式J(θ)中。

$$w=\begin{bmatrix}w_1 & & & & \\ & \ddots & & & \\ & & w_i & & \\ & & & \ddots & \\ & & & & w_m\end{bmatrix}\qquad \begin{aligned}J(\theta)&=\sum_{i=1}^{m}w^{(i)}\left[y^{(i)}-\theta^{T}x^{(i)}\right]^{2}\\&=y^{T}wy-\theta^{T}x^{T}wy-y^{T}w^{T}x\theta+\theta^{T}x^{T}wx\theta\end{aligned}$$

利用最小二乘法，可以计算出一个θ向量（一个预测点对应一个向量）

$$\nabla_{\theta}J(\theta)=0$$

$$-x^{T}wy-x^{T}wy+2x^{T}wx\theta=0$$

$$\theta=(x^{T}wx)^{-1}x^{T}wy$$

如此下去，i从1取到m，就可以计算出m个θ向量。

岭回归（加入L2范数正则化）

岭回归(Ridge Regression)是在平方误差的基础上增加正则项

$$\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{p}w_jx_{ij}\right)^{2}+\lambda\sum_{j=0}^{p}w_j^{2}\ ,\lambda$$

通过确定λ的值可以使得在方差和偏差之间达到平衡：随着λ的增大，模型方差减小而偏差增大。

对w求导，结果为

$$2X^{T}\left(Y-XW\right)-2\lambda W$$

令其为0，可求得w的值：

$$\hat{w}=\left(X^{T}X+\lambda I\right)^{-1}X^{T}Y$$

最先用来处理特征数多于样本数的情况，现在也用于在估计中加入偏差，从而得到更好的估计。通过引入lambda限制了所有w之和，通过引入该惩罚项，能减少不重要的参数，在统计学中叫缩减（shrinkage）。

Lasso（Least Absolute Shrinkage and Selection Operator）（L1范数正则化）

$$\min_{\boldsymbol{w}}\sum_{i=1}^{m}(y_i-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i)^{2}+\lambda\|\boldsymbol{w}\|_1\ .$$

在lambda足够小的时候，一些系数会因此被迫缩减到0，容易获得“稀疏解”，所以lasso在压缩感知中应用广泛。但是需要二次规划方法来求解

Lasso 回归主要的解法有两种：坐标轴下降法（coordinate descent）和最小角回归法（ Least Angle Regression）。