HW11

PB20111689 蓝俊玮

T1

> 结合图 11.2，试举例说明 L_1 正则化在何种情形下不能产生稀疏解

L_1 正则化之所以可以更容易获得稀疏解，是因为它求得的 $\boldsymbol{w}$ 具有更少的非零分量。而这种情况一般出现在 L_1 范数平方误差项等值线与正则化项等值线交点出现在某个坐标轴上。

在图 11.2 中可以看出，L_1 范数时平方误差项等值线与正则化项等值线交点出现在 w_1 轴上，同时这个交点也是它们之间唯一的交点。如果 L_1 范数时平方误差项等值线与正则化项等值线交点不再出现再坐标轴上，而是某个象限中的一点，那么 L_1 将不会产生稀疏解 $\boldsymbol{w}$。

T2

> 试述直接求解 L_0 范数正则化会遇到的困难

L_0 范数正则化求的是 $\boldsymbol{w}$ 中非零分量的个数。但是 L_0 范数是不连续的，且它并不是凸优化问题，所以它无法通过优化直接求解，必须采用遍历的方式，因此导致这个问题是个 NP Hard 问题。

T3

> 证明回归和对率回归的损失函数的梯度是否满足 L-Lipschitz 条件，并求出 L

回归问题的损失函数梯度为：

$$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}} = 2\boldsymbol{X}^T(\boldsymbol{X}\hat{\boldsymbol{w}} - \boldsymbol{y})$$

那么任意取 $\forall \boldsymbol{w}, \boldsymbol{w}'$，可以得到：

$$\begin{aligned}||\nabla f(\boldsymbol{w}') - \nabla f(\boldsymbol{w})|| =& 2||\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{y}) - \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y})|| \\ =& 2||\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{w}' - \boldsymbol{w})|| \\ \leq& 2||\boldsymbol{X}^T\boldsymbol{X}|| \cdot ||\boldsymbol{w}' - \boldsymbol{w}||\end{aligned}$$

其中由于 $2||\boldsymbol{X}^T\boldsymbol{X}||$ 是大于 0 的，因此回归问题是满足 L-Lipschitz 条件的，可以令 $L = 2||\boldsymbol{X}^T\boldsymbol{X}||$。

而对率回归问题的损失函数梯度为：

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{m} \hat{\boldsymbol{x}}_i(y_i - p_1(\hat{\boldsymbol{x}}_i; \boldsymbol{\beta}))$$

那么任意取 $\forall \boldsymbol{\beta}, \boldsymbol{\beta}'$，可以得到：

$$||\nabla l(\boldsymbol{\beta}') - \nabla l(\boldsymbol{\beta})|| = ||\sum_{i=1}^{m} \hat{\boldsymbol{x}}_i(p_1(\hat{\boldsymbol{x}}_i; \boldsymbol{\beta}') - p_1(\hat{\boldsymbol{x}}_i; \boldsymbol{\beta}))||$$

因为有 $p_1(\hat{\boldsymbol{x}}; \boldsymbol{\beta})) = \frac{e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x}}}}{1+e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x}}}}$，所以对其整体求导可以得到 $p_1'(\hat{\boldsymbol{x}}; \boldsymbol{\beta})) = \frac{e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x}}}}{(1+e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x}}})^2} \leq \frac{1}{4}$，所以可以利用拉格朗日中值定理：

$$\begin{aligned}||\nabla l(\boldsymbol{\beta}') - \nabla l(\boldsymbol{\beta})|| &= ||\sum_{i=1}^{m} \hat{\boldsymbol{x}}_i (p_1(\hat{\boldsymbol{x}}_i;\boldsymbol{\beta}') - p_1(\hat{\boldsymbol{x}}_i;\boldsymbol{\beta}))|| \\ &= p_1'(\hat{\boldsymbol{x}};\boldsymbol{\beta}))||\sum_{i=1}^{m} \hat{\boldsymbol{x}}_i (\boldsymbol{\beta}'^T \hat{\boldsymbol{x}}_i - \boldsymbol{\beta}^T \hat{\boldsymbol{x}}_i)|| \\ &\leq \frac{1}{4}||\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\beta}' - \boldsymbol{\beta})|| \\ &\leq \frac{1}{4}||\boldsymbol{X}^T\boldsymbol{X}|| \cdot ||\boldsymbol{\beta}' - \boldsymbol{\beta}||\end{aligned}$$

其中由于 $\frac{1}{4}||\boldsymbol{X}^T\boldsymbol{X}||$ 是大于 0 的，因此对率回归问题是满足 L-Lipschitz 条件的，可以令 $L = \frac{1}{4}||\boldsymbol{X}^T\boldsymbol{X}||$。