Homework VI: Linear Model

自 71 游晶 2017011542

November 4, 2019

1 Problem I

Given a Gaussian linear regression model, Maximum likelihood estimation of under Gaussian noise assumption is equivalent to the least square loss minimization.

Prove:

Let

$$y_n | \boldsymbol{w}, \boldsymbol{x}_n \sim \mathcal{N}\left(\boldsymbol{w}^T \boldsymbol{x}_n, \sigma^2\right)$$

So for every $(\boldsymbol{x}_n, y_n)$

$$p\left(y_n | \boldsymbol{w}, \boldsymbol{x}_n\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2\right\}$$

And the log-likelihood on the whole dataset is

$$\begin{aligned} \log p\left(\mathscr{D}_n; \boldsymbol{w}\right) &= \textstyle\sum_{n=1}^N \log p\left(y_n | \boldsymbol{w}, \boldsymbol{x}_n\right) \\ &= \tfrac{N}{2} \log \tfrac{1}{2\pi\sigma^2} - \tfrac{1}{2\sigma^2} \textstyle\sum_{n=1}^N \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2 \end{aligned}$$

So maximum likelihood estimation of $\boldsymbol{w}$ is

$$\begin{aligned} &\max_{\boldsymbol{w}} \frac{N}{2} \log \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2} \sum_{n=1}^N \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2 \\ &= \max_{\boldsymbol{w}} - \frac{1}{2\sigma^2} \sum_{n=1}^N \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2 \\ &= \min_{\boldsymbol{w}} \sum_{n=1}^N \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2 \end{aligned}$$

2 PROBLEM II

2. Given a Laplacian linear regression model, Maximum likelihood estimation of under Laplacian noise assumption is equivalent to the least absolute loss (L1 loss) minimization.

Prove:

Let

$$y_n|\boldsymbol{w},\boldsymbol{x}_n \sim \text{Laplace}\left(\boldsymbol{w}^T\boldsymbol{x}_n, b\right)$$

So for every $(\boldsymbol{x}_n, y_n)$

$$p\left(y_n|\boldsymbol{w},\boldsymbol{x}_n\right) = \frac{1}{2b}\exp\left\{-\frac{\left|y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right|}{b}\right\}$$

And the log-likelihood on the whole dataset is

$$\begin{aligned}&\log p\left(\mathscr{D}_n;\boldsymbol{w}\right) = \textstyle\sum_{n=1}^N \log p\left(y_n|\boldsymbol{w},\boldsymbol{x}_n\right)\\ &= N\log\tfrac{1}{2b} - \tfrac{1}{b}\textstyle\sum_{n=1}^N\left|y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right|\end{aligned}$$

So maximum likelihood estimation of $\boldsymbol{w}$ is

$$\max_{\boldsymbol{w}}\sum_{n=1}^N \log p\left(y_n|\boldsymbol{w},\boldsymbol{x}_n\right) = \min_{\boldsymbol{w}}\sum_{n=1}^N\left|y_n - \boldsymbol{w}^T\boldsymbol{x}_n\right|$$

3 PROBLEM III

3.1 Ridge Regression

Tikhonov Form:

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}\in\mathbb{R}^d}{\arg\min}\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{w}^T\boldsymbol{x}_i - y_i\right)^2 + \lambda\|\boldsymbol{w}\|_2^2$$

$$\text{while } \lambda \geq 0$$

Ivanov Form:

$$\hat{\boldsymbol{w}} = \underset{\|\boldsymbol{w}\|_2^2\leq r}{\arg\min}\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{w}^T\boldsymbol{x}_i - y_i\right)^2$$

$$\text{while } r \geq 0$$

3.2 Lasso Regression

Tikhonov Form:

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}\in\mathbb{R}^d}{\arg\min}\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{w}^T\boldsymbol{x}_i - y_i\right)^2 + \lambda\|\boldsymbol{w}\|_1$$

$$\text{while } \lambda \geq 0$$

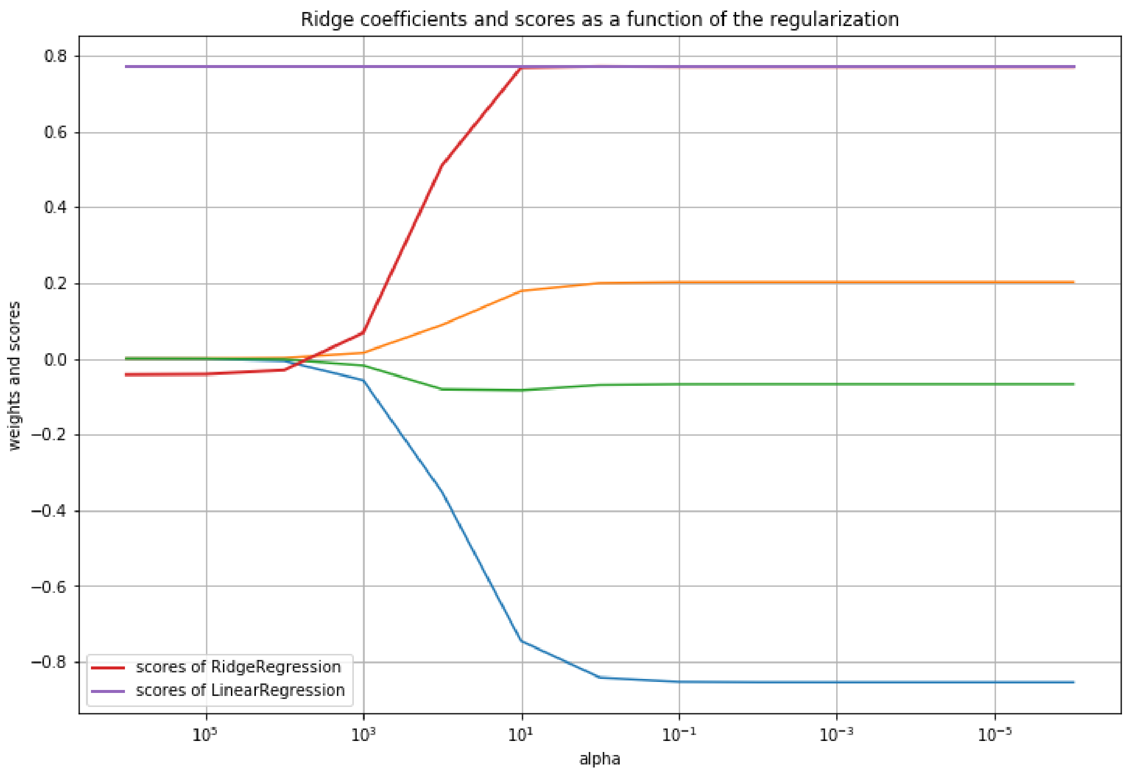
Ivanov Form:

$$\hat{\boldsymbol{w}} = \underset{\|\boldsymbol{w}\|_1\leq r}{\arg\min}\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{w}^T\boldsymbol{x}_i - y_i\right)^2$$

$$\text{while } r \geq 0$$

4 PROBLEM IV

画出岭回归和线性回归两种模型的 R2_score 与参数 alpha 的图像如下（详见.ipynb 文件）：



由于线性回归与 alpha 无关，故为一条水平线。

对于岭回归模型，从图中可以看到，当 alpha 较大时，对权重的约束较大，导致权重接近于零，模型欠拟合，预测效果不理想。

当 alpha 较小时，对权重的约束相当于没有，模型退化为线性回归模型。

当 alpha 在这两种极端情况之间时，预测的效果却并没有显著的提升，这是因为我们的模型太简单，特征只有三个，对于线性回归模型来说甚至没有出现欠拟合的现象，加入正则项来防止欠拟合更是无从谈起，因此岭回归就算是选取了适当的超参数也无法提升预测的效果。

所以我们可以看到，岭回归模型并没有取得更好的预测效果，或者说并没有获得更小的泛化误差。