

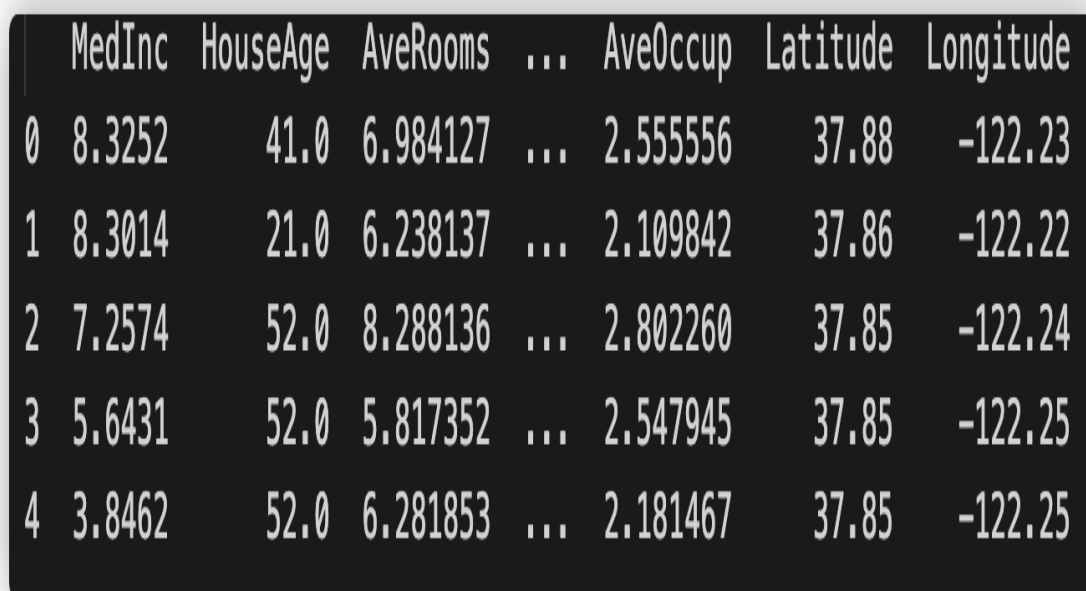
Homework 7

张思源 21110850018

November 11, 2021

1 Ex1

Sol 1.1 首先导入数据, 将列名更改为特征名称, 并打印前 5 行可以得到结果如下图:



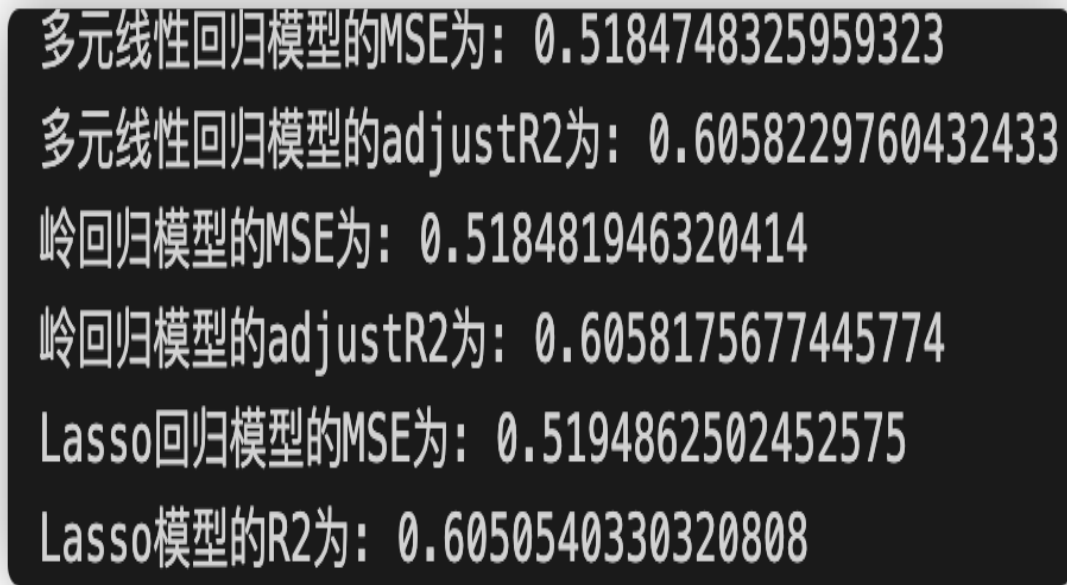
	MedInc	HouseAge	AveRooms	...	AveOccup	Latitude	Longitude
0	8.3252	41.0	6.984127	...	2.555556	37.88	-122.23
1	8.3014	21.0	6.238137	...	2.109842	37.86	-122.22
2	7.2574	52.0	8.288136	...	2.802260	37.85	-122.24
3	5.6431	52.0	5.817352	...	2.547945	37.85	-122.25
4	3.8462	52.0	6.281853	...	2.181467	37.85	-122.25

Figure 1: Head of the data

然后, 打印房价 (即 *target*) 的最值, 可以发现其最大值为 5.00001, 最小值为 0.14999. 进而计算模型的 *VIF*(方差膨胀系数): 可以得到 8 个特征的 *VIF* 系数依次为: 11.51, 7.20, 45.99, 43.59, 2.94, 1.10, 559.87, 633.71, 可以看出第 1 个、第 3 个、第 4 个特征的 *VIF* 系数超过了 10, 第 7 个和第 8 个特征的 *VIF* 系数更是超过了 100, 说明数据存在多重共线性.

对数据集分别利用多元线性回归、岭回归和 *Lasso* 回归模型, 可以得到模型的 *MSE* 和 *adjustR²* 分别如下图所示, 其中

$$\text{adjust}R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

Figure 2: MSE and $adjustR^2$ of 3 models

最后, 比较不同的正则化项可得下表:

	<i>Lasso</i> (<i>L1</i> 正则项)	<i>Ridge</i> (<i>L2</i> 正则项)
相同点	都可以用来解决过拟合问题	
不同点	可以用来做 <i>feature selection</i> , 更容易使得权重变为 0	不可以做 <i>feature selection</i> , 更容易使得权重接近 0

2 Ex2

Sol 2.1 首先, 计算得:

$$\begin{aligned}
 \frac{\partial E}{\partial w_{kj}} &= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}} \\
 &= (o_k - d_k) f\left(\sum_{j=0}^m w_{kj} y_j\right) \left(1 - \left(\sum_{j=0}^m w_{kj} y_j\right)\right) y_j \\
 &= (o_k - d_k) o_k (1 - o_k) y_j
 \end{aligned}$$

所以 $\Delta w_{kj} = -\eta(o_k - d_k) o_k (1 - o_k) y_j$. 并且可以计算得:

$$\begin{aligned}
 \frac{\partial E}{\partial v_{ji}} &= \sum_{k=1}^l \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial net_k} \frac{\partial net_k}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial v_{ji}} \\
 &= \sum_{k=1}^l (o_k - d_k) f\left(\sum_{j=0}^m w_{kj} y_j\right) \left(1 - f\left(\sum_{j=0}^m w_{kj} y_j\right)\right) w_{kj} f\left(\sum_{i=0}^n v_{ji} x_i\right) \left(1 - f\left(\sum_{i=0}^n v_{ji} x_i\right)\right) x_i \\
 &= \sum_{k=1}^l (o_k - d_k) o_k (1 - o_k) w_{kj} y_j (1 - y_j) x_i
 \end{aligned}$$

所以 $\Delta v_{ji} = -\eta \sum_{k=1}^l (o_k - d_k) o_k (1 - o_k) w_{kj} y_j (1 - y_j) x_i$.

References

- [1] 李航. 统计学习方法 [M]. 清华大学出版社, 2012.
- [2] Goodfellow, Ian, et al. Deep Learning[M]. MIT Press, 2016.