

ML Homework 2

林大权

ID: 85610653

1 (1) :

$$Pr(Y_i|\mathbf{x}_i) \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\theta}, \sigma^2)$$

$$L_n(\mathbf{x}_i^T \boldsymbol{\theta}; y|x) = \prod_{i=1}^n f_{Y|x}(y_i|\mathbf{x}_i; \mathbf{x}_i^T \boldsymbol{\theta})$$

log MLE is:

$$l_n(\mathbf{x}_i^T \boldsymbol{\theta}; y|x) = \sum_{i=1}^n \ln \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2}{2\sigma^2}\right)$$

$$l_n(\mathbf{x}_i^T \boldsymbol{\theta}; y|x) = -\frac{m}{2} \ln \sigma^2 - \frac{m}{2} \ln(2\pi) - \frac{1}{\sigma^2} \frac{1}{2} \|\mathbf{X}\boldsymbol{\theta} - \mathbf{y}\|^2$$

1 (2) : $\because \log(\cdot)$ is a monotone function.

To find the maximum of $l_n(\mathbf{x}_i^T \boldsymbol{\theta}; y|x)$ is to find the minimum of $\|\mathbf{X}\boldsymbol{\theta} - \mathbf{y}\|^2$

1 (3) : Because of the square, large errors have relatively greater influence on MSE than do the smaller error.

2 (1) : Loss function:

$$J(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{X} + b$$

2 (2) : My termination criterion is 1000 iterates, each iterate use 1 samples. And Learning Rate is 0.001, Regularization Strength is 10^{-5} . $\sigma = 2$ with Huber Loss Function.

2 (3) : Convergence result:

$$y = 0.92x + 0.31$$

Fitted curve in fig.1. the convergence result in fig.2

3 (1) (i) : The linear regression is:

$$Price = -0.17135 Mileage + 24811.9$$

3 (1) (ii) : Use R^2 score:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \frac{1}{n} \sum_{i=1}^n y_i)^2}$$

$$R^2 = 0.0204$$

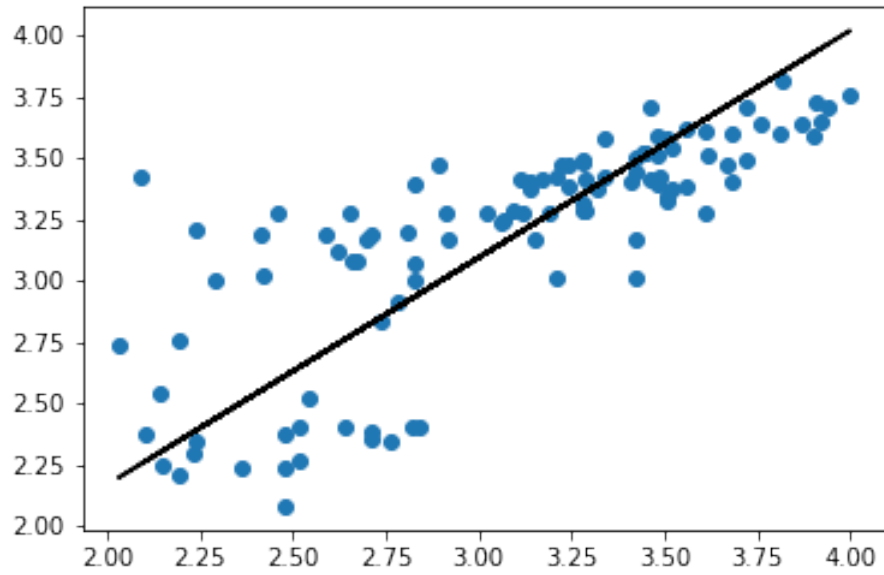


图 1: The fitted curve of GPA.txt in Problem 2.

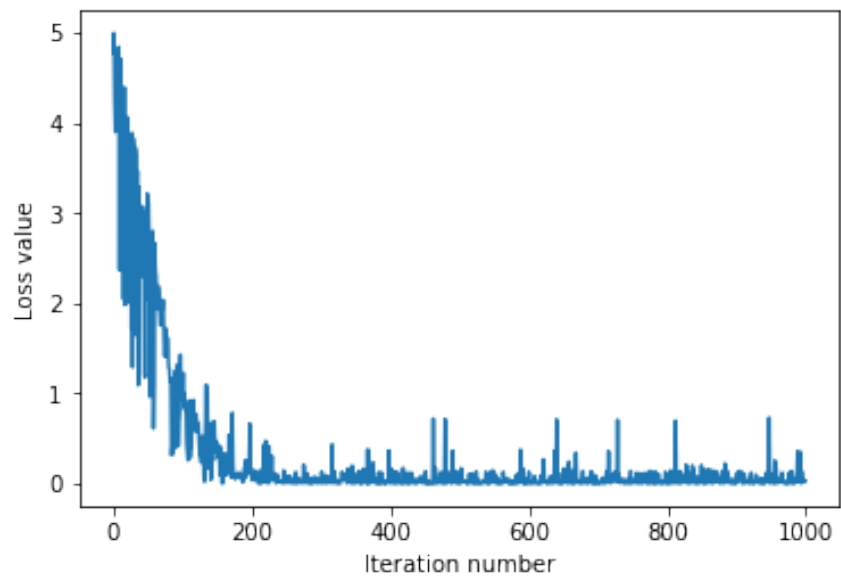


图 2: The fitted curve of GPA.txt in Problem 2.

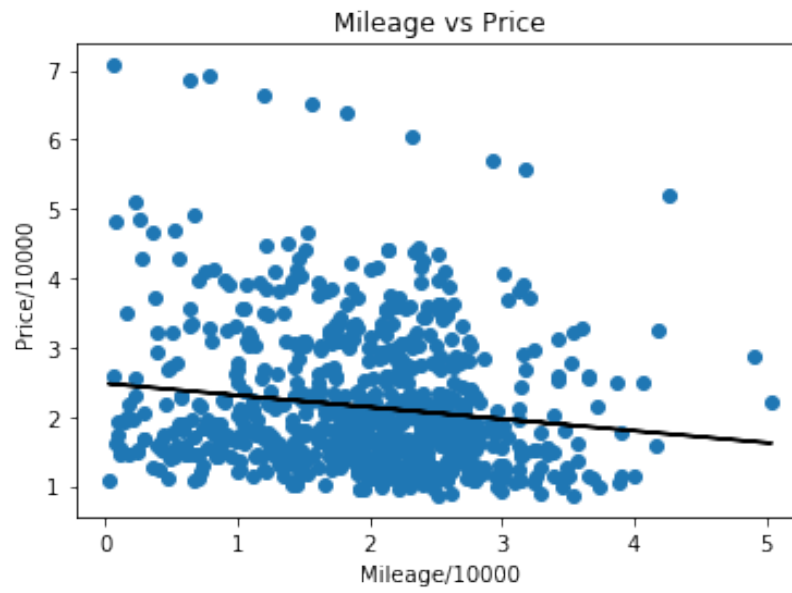


图 3: The fitted curve of Mileage vs Price.

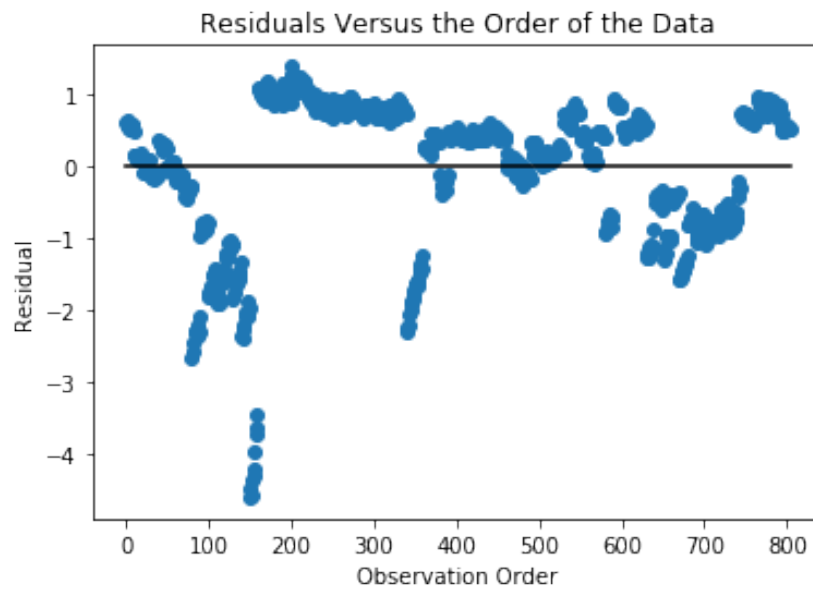


图 4: Residuals($y_p - y$) vs the Order of the Data

3 (1) (iii) : Mileage affect the price, absolutely.

Look at Pic.3, the outliers on the above of the plot are all Cadillac XLR-V8, which is more expensive than average. It shows a strong linear relationship with Mileage, which illustrates an important phenomenon: the strength of the relationship can be highly dependent on the set of cases. Also, it can find in Pic.4, the cars are all cluster together. So, Mileage is important to Price.

3 (2) (i) :

$$Price = -0.17Mileage + 0.39Cylinder - 0.10Liter - 0.15Doors + 0.63Cruise - 0.19Sound + 0.34Leather + 6097 \quad (1)$$

3 (2) (ii) : $R^2 = 0.44$, it's very small than 1, so it doesn't a good fit.

3 (2) (iii) : Use all 97 factors, $R^2 = 0.997$ is the best predictor for price.