ML Homework 1

林大权

ID: 85610653

1 (1) ：1) 反欺诈：一些恶意商户为了在某个电商平台或者推荐平台上，刷自己旗下的商品的交易量或访问量，这时平台的安全人员可以用机器学习算法进行数据分析，找到离群点，很快发现欺诈数据。

2）交通拥堵预测：分析以往数据预测未来道路交通情况，及时规避拥堵路段。

2 (2) ：

证明. Suppose a sufficiently small step $\alpha\boldsymbol{d}$, $\alpha > 0$.

We have $f(\boldsymbol{x}+\alpha\boldsymbol{d}) = f(\boldsymbol{x}) + \alpha\nabla f(\boldsymbol{x})^T\boldsymbol{d} + o(\alpha||\boldsymbol{d}||)$, where $o(\alpha||\boldsymbol{d}||) \to 0$ as $\alpha \to 0$, since $||\boldsymbol{d}||$ is 1.

$\Rightarrow$

$$\frac{f(\boldsymbol{x}+\alpha\boldsymbol{d}) - f(\boldsymbol{x})}{\alpha} = \nabla f(\boldsymbol{x})^T\boldsymbol{d} + \frac{o(\alpha||\boldsymbol{d}||)}{\alpha}$$

Since $\lim_{\alpha\to 0}\frac{o(\alpha||\boldsymbol{d}||)}{\alpha}$,

then $\exists\bar{\alpha}$, s.t. $\forall\alpha \in (0,\bar{\alpha})$,

we have $\frac{o(\alpha||\boldsymbol{d}||)}{\alpha} < \frac{1}{2}|\alpha\nabla f(\boldsymbol{x})^T\boldsymbol{d}|$

Since $\alpha\nabla f(\boldsymbol{x})^T\boldsymbol{d} < 0$ by assumption, we conclude that $\forall\alpha \in (0,\bar{\alpha})$,

$f(\boldsymbol{x}+\alpha\boldsymbol{d}) - f(\boldsymbol{x}) < \frac{1}{2}\alpha\nabla f(\boldsymbol{x})^T\boldsymbol{d} < 0$

$\square$

2 (1) ：

证明. $\because$ the Hessian of f : $\boldsymbol{H}$is PSD, and $\forall\boldsymbol{x} \neq \boldsymbol{0}$, we have $\boldsymbol{x^T H x} \geq 0$

$\therefore$ From the key, we can get $f(\boldsymbol{x}) \geq f(\boldsymbol{x'}) + \nabla f(\boldsymbol{x'})^T(\boldsymbol{x} - \boldsymbol{x'})$

$\therefore \forall x, y \in \boldsymbol{dom}(f)$, Suppose $z = \lambda x + (1-\lambda)y$

we have

$$f(\boldsymbol{x}) \geq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{x} - \boldsymbol{z}) \tag{1}$$

$$f(\boldsymbol{y}) \geq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{y} - \boldsymbol{z}) \tag{2}$$

then $(1) * \lambda + (2) * (1-\lambda)$, we have:

$$\lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}) \geq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} - \boldsymbol{z}) = f(\boldsymbol{z}) = f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})$$

$\therefore f$ is convex.

$\square$

2 (2) :

证明. $\because f$ is convex, be definition:

$$\lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x}) \geq f(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x}), \forall \lambda \in [0,1], x, y \in \boldsymbol{dom}(f)$$

$\Rightarrow$

$$f(\boldsymbol{x}) + \lambda(f(\boldsymbol{y}) - f(\boldsymbol{x})) \geq f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x}))$$

$\Rightarrow$

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda(\boldsymbol{y} - \boldsymbol{x})}(\boldsymbol{y} - \boldsymbol{x})$$

$\Rightarrow$

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x})$$

Suppose $\boldsymbol{x}^*$ is the global minimizer, $\boldsymbol{x}^*$ doesn't satisfy $\nabla f(\boldsymbol{x}^*) = 0$, we can move $\boldsymbol{x}^*$ along direction $-\nabla f(\boldsymbol{x}^*)$ with non-zero distance to $\boldsymbol{y}$; such that $\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{y} - \boldsymbol{x}^*) < 0$.

Consider $\phi(\alpha) = f(\boldsymbol{x}^* + \alpha(\boldsymbol{y} - \boldsymbol{x}^*))$, because f is convex, such that $\forall \alpha \in [0,1]$ $\boldsymbol{x}^* + \alpha(\boldsymbol{y} - \boldsymbol{x}^*) \in \boldsymbol{dom}(f)$.
Observe that

$$\phi\prime(\alpha) = (\boldsymbol{y} - \boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^* + \alpha(\boldsymbol{y} - \boldsymbol{x}^*))$$

$\Rightarrow$

$$\phi\prime(0) = (\boldsymbol{y} - \boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*) < 0$$

This implies that

$$\exists \delta > 0, s.t.\ \phi(\alpha) < \phi(0), \forall \alpha \in (0, \delta)$$

$\Rightarrow$

$$f(\boldsymbol{x}^* + \alpha(\boldsymbol{y} - \boldsymbol{x}^*)) < f(\boldsymbol{x}^*), \forall \alpha \in (0, \delta)$$

But this contradicts the optimality of $\boldsymbol{x}^*$.

□

3 : The cost function is:

$$y = a\boldsymbol{x} + b$$

$$J(a,b) = \frac{1}{2N}\sum_{i=0}^{N}(y_i - ax_i - b)^2$$

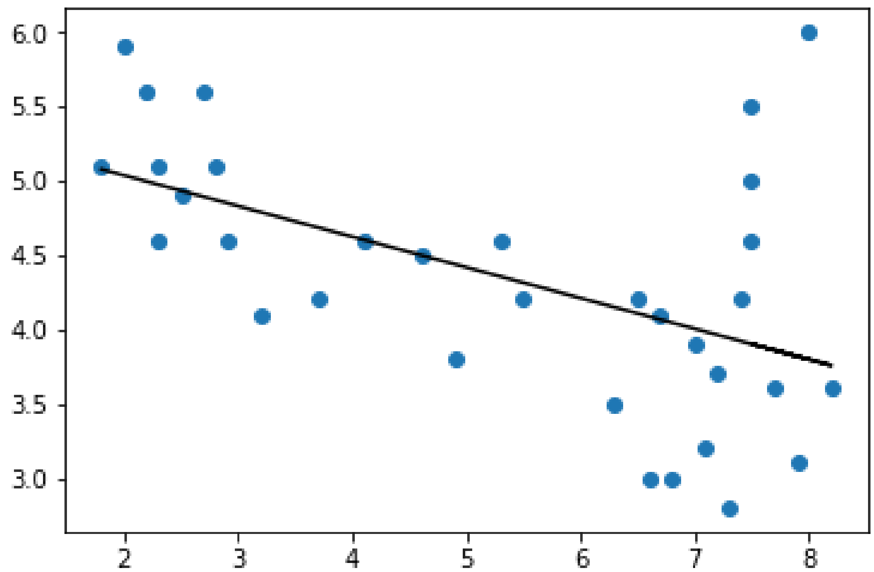

图 1: The result of 4-trainingdata.txt

Termination criterion is number of iteration, here I set 50000.

Due to I rescale the input data, so the result:

$$loss = 0.26617$$

$$a = -0.206369981511$$

$$b = 5.44851173492$$

Shown in Pic.1

4 : The Huber cost function is:

$$L_\sigma(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & |y - f(x)| \leq \sigma \\ \sigma|y - f(x)| - \frac{1}{2}\sigma^2 & otherwise \end{cases}$$

In 5-trainingdata.txt, use the cost function in Problem 3 is shown in Pic.2.

And loss And the Huber cost funcion is Pic.3, a little improve than old one.

The tess loss for The least-square method and Huber method are 0.08336623884690787 and 0.08333359874471885, respectively.

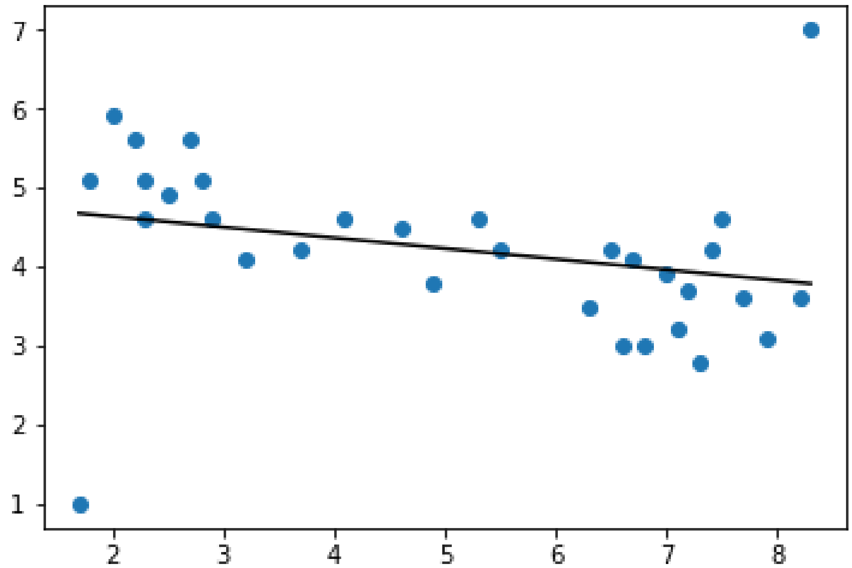

图 2: The result of 5-trainingdata.txt in Problem 3 cost function.

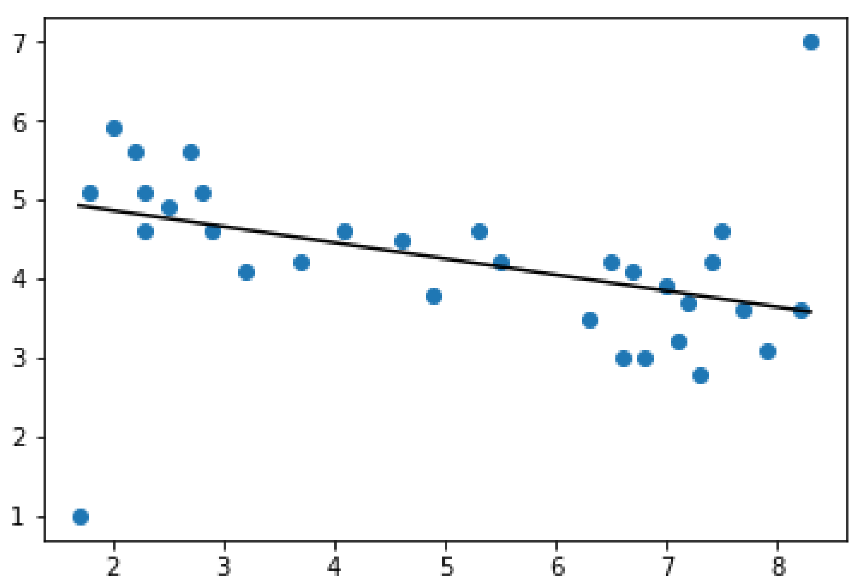

图 3: The result of 5-trainingdata.txt in Huber cost function.