Solution

1 Problem 1

1.1

1) 网站的安全防护，用机器学习的方法分析网络访问流量模式，当出现异常网络流量访问时，对其进行安全防护处理。

2) 电商的风控系统，当出现恶意的订单行为 - 比如恶意薅羊毛行为，对其进行报告，提醒及时应对风险。

1.2

Suppose move a sufficient small step αd along direction d, $\alpha > 0$ and $\alpha \to 0$. Using Taylor's expansion, we have:

$$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^T d + o(\alpha)$$

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

Since $\lim_{\alpha \to 0} \frac{o(\alpha)}{\alpha} = 0$, there exist $\bar{\alpha}$, s.t. $\forall \alpha \in (0, \bar{\alpha})$, we have $\frac{|o(\alpha)|}{\alpha} < \frac{1}{2}|\nabla^T f(x) d|$.

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

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

$\square$

2 Problem 2

2.1

If the Hessian $\mathbf{H}$ of f is positive semidefinite, such that:

$$\forall x \qquad x^T \mathbf{H} x \geq 0$$

As the key hint tells us, combine it with the property of PSD, we have,

$$f(x) \geq f(x\prime) + \nabla f(x\prime)^T (x - x\prime)$$

Such that, $\forall x, y \in \mathbf{dom(f)}$, denote $z = \lambda x + (1 - \lambda y)$ where $\lambda \in [0, 1]$

$$f(x) \geq f(z) + \nabla f(z)^T (x - z) \tag{1}$$
$$f(y) \geq f(z) + \nabla f(z)^T (y - z) \tag{2}$$

Then, combine equation (1) and (2) together with the form of $(1) * \lambda + (2) * (1 - \lambda)$

$$\begin{aligned}
\lambda f(x) + (1 - \lambda) f(y) &\geq \lambda(f(z) + \nabla f(z)^T (x - z)) + (1 - \lambda)(f(z) + \nabla f(z)^T (y - z)) \\
&= f(z) + \nabla f(z)^T (\lambda x + (1 - \lambda) y - z) \\
&= f(z) \\
&= f(\lambda x + (1 - \lambda) y)
\end{aligned}$$

We can conclude that $f(x)$ if convex. $\square$

2.2

Since f is convex; by definition we have $\lambda f(y) + (1 - \lambda) f(x) \geq f(\lambda y + (1 - \lambda) x)$, where $x, y \in \mathbf{dom(f)}$ and $\lambda \in [0, 1]$.

After rewriting the above inequality equation, we have:

$$\begin{aligned}
f(y) - f(x) &\geq \frac{f(x + \lambda(y - x)) - f(x)}{\lambda(y - x)} (y - x) \\
&= \nabla f(x)^T (y - x)
\end{aligned} \tag{3}$$

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

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

Observe that

$$g\prime(\alpha) = (y - x^*)^T \nabla f(x^* + \alpha(y - x^*))$$

$$\implies g\prime(0) = (y - x^*)^T \nabla f(x^*) < 0 \tag{4}$$

This implies that

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

$$\implies f(x^* + \alpha(y - x^*)) < f(x^*), \forall \alpha \in (0, \delta) \tag{5}$$

But this contradicts the optimality of x^*. $\square$

3 Problem 3

3.1

Linear function of x, where x is age of drivers, $h_\theta(x)$ is the distance at the age x.

$$h_\theta(x) = \theta_0 + \theta_1 x, \theta = [\theta_0, \theta_1]^T$$

Hypothesis

$$\mathcal{H} = \{\theta_0 + \theta_1 x | \theta_0, \theta_1 \in \mathbb{R}\}$$

Cost function, square error.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

3.2

Please check it in the code.

4 Problem 4

Cost function, huber cost function.

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=0}^{m} \begin{cases} \frac{1}{2}(y - h_\theta(x))^2 & \text{if } |y - h_\theta(x)| < \sigma \\ \sigma(|y - h_\theta(x)| - \frac{1}{2}\sigma) & \text{otherwise} \end{cases}$$

Run the code on **5-traningdata.txt** with square error function and huber cost function respectively.

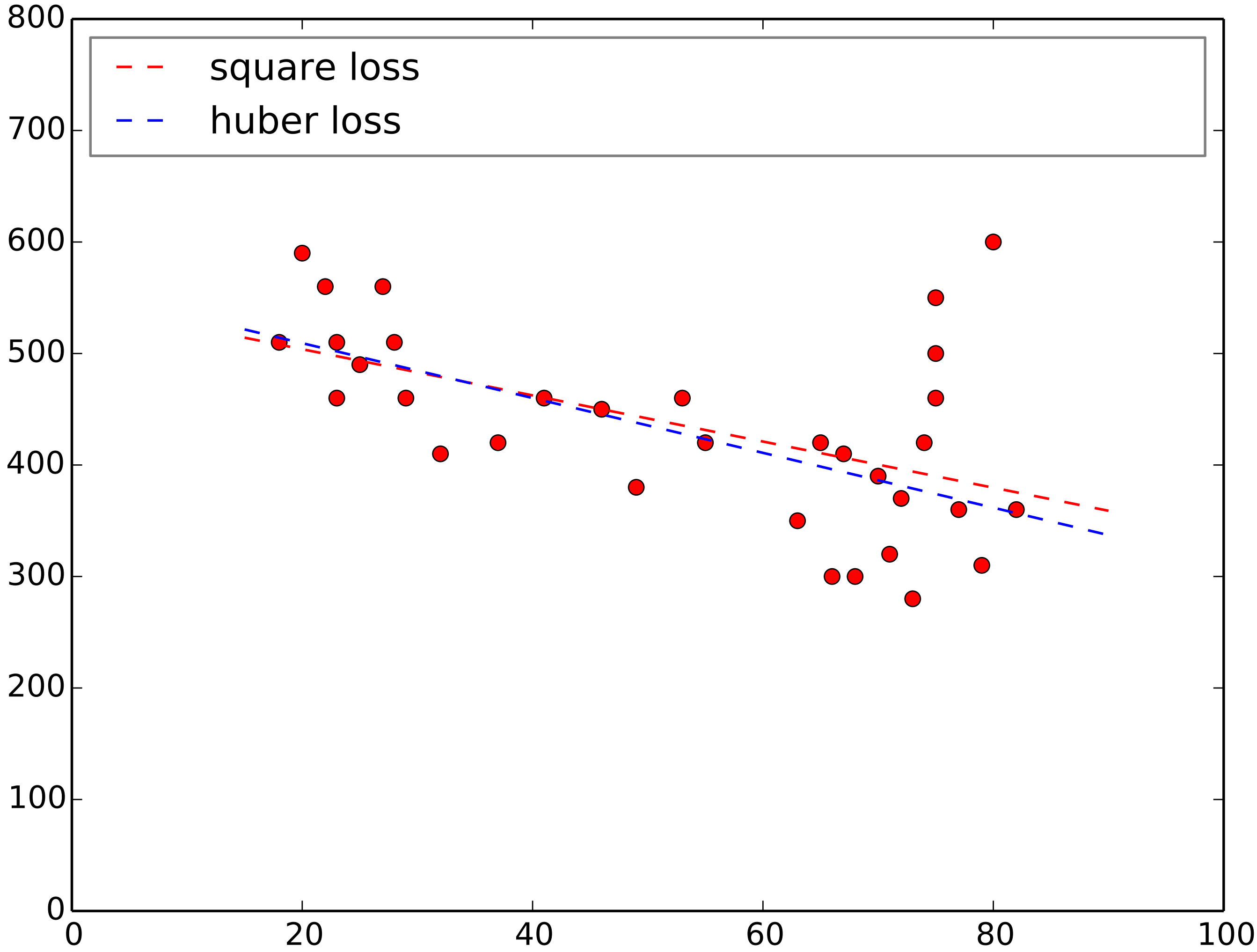
$$\theta_a = \theta^T_{square\ cost\ function} = (\theta_0, \theta_1) = (545.236509046, -2.06992700152)$$

$$\theta_b = \theta^T_{huber\ cost\ function} = (\theta_0, \theta_1) = (558.451008291, -2.45927631313)$$

Run the code with the above results, get the loss

$$J^{\theta_a}_{square\ loss} = 930.622443336$$

$$J^{\theta_b}_{square\ loss} = 626.494535827$$


square loss
huber loss
0
100
200
300
400
500
600
700
800
0
20
40
60
80
100