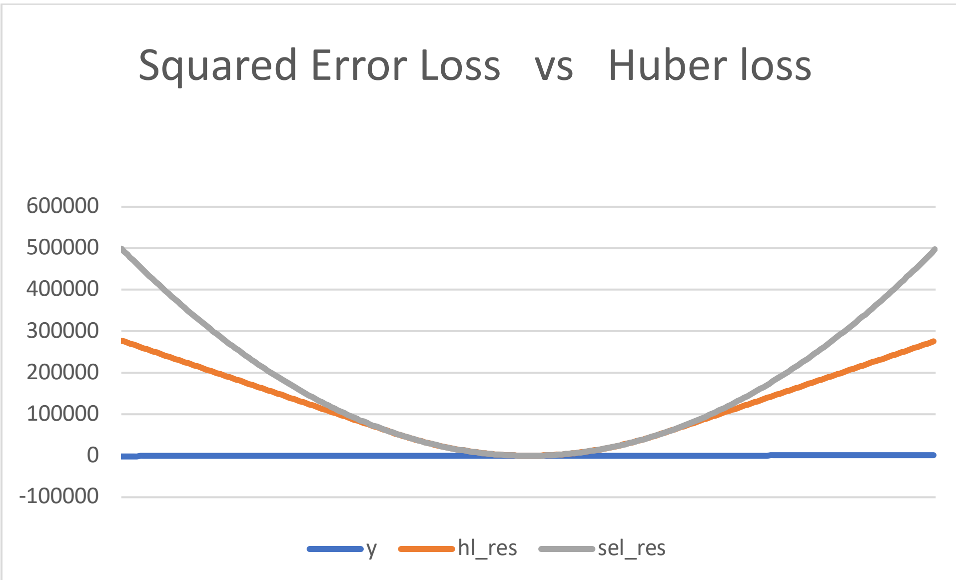
Q1

(a)

When the value y is outlier, compared to Squared Error Loss, the Huber loss is much less which is shown in below plot. ($\delta = 333,\ y's\ interval\ is\ [-999, 999]$)



(b)

(b) $$H_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{if } |a| \le \delta \\ \delta(|a| - \frac{1}{2}\delta) & \text{if } |a| > \delta \end{cases}$$

$$\Rightarrow H'_\delta(a) = \begin{cases} a & \text{if } |a| \le \delta \\ \delta & \text{if } a > \delta \\ -\delta & \text{if } a < -\delta \end{cases} \qquad ①$$

$$y = w^T x + b = \sum_{1}^{d} w_i x_i + b$$

$$\frac{\partial L_\delta}{\partial w} = H'_\delta(y-t) \cdot \frac{\partial (y-t)}{\partial w} = H'_\delta(y-t) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

where we can apply formula ①

$$\frac{\partial L_\delta}{\partial b} = H'_\delta(y-t) \cdot \frac{\partial (y-t)}{\partial b} = H'_\delta(y-t) \cdot 1 = H'_\delta(y-t)$$

where we can apply formula ①

Q2

(a)

$$J = \frac{1}{2}\sum_{i=1}^{N} a^{(i)}(y^{(i)} - w^T x^{(i)})^2 + \frac{\lambda}{2}\|w\|^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} a^{(i)}\left({y^{(i)}}^2 - 2y^{(i)}w^T\mathbf{x}^{(i)} + (w^T x^{(i)})^2\right) + \frac{\lambda}{2}\|w\|^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} a^{(i)}{y^{(i)}}^2 - \sum_{i=1}^{N} a^{(i)}y^{(i)}w^T\mathbf{x}^{(i)} + \frac{1}{2}\sum_{i=1}^{N} a^{(i)}(w^T\mathbf{x}^{(i)})^2 + \frac{\lambda}{2}\|w\|^2$$

$$= \frac{1}{2}y^TAy - w^TX^TAy + \frac{1}{2}w^TX^TAXw + \frac{\lambda}{2}w^Tw$$

Let $\nabla_w J = 0$

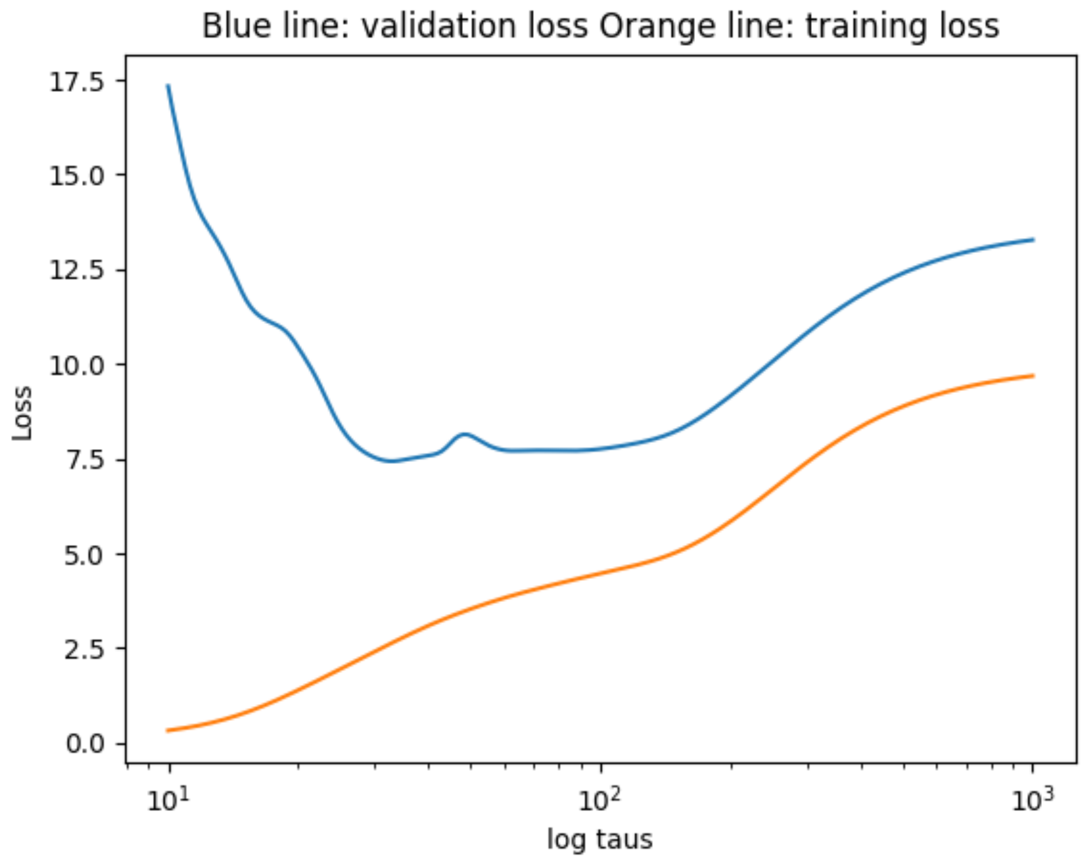
$$\Rightarrow -X^TAy + X^TAXw + \lambda w = 0$$

$$(X^TAX + \lambda I)w = X^TAy$$

$$\Rightarrow w = (X^TAX + \lambda I)^{-1}X^TAy$$

$$\Rightarrow w^* = (X^TAX + \lambda I)^{-1}X^TAy$$

(c)



(d)

When γ -> ∞, each training example is equally weighted, so the model is close to standard regression with L^2 penalty.

When γ -> 0, the test point becomes dramatic driver to the model which could lead poor generalization.