(c)

$$\begin{aligned}\ell(\theta) &= -\sum_{i=1}^{m} \log p(y^{(i)}|x^{(i)};\theta) \\ &= \sum_{i=1}^{m} -\log b(y^{(i)}) - \theta^T x^{(i)} y^{(i)} + a(\theta^T x^{(i)})\end{aligned}$$

$$\frac{\partial \ell(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} [a'(\theta^T x^{(i)}) - y^{(i)}] x_j^{(i)}$$

$$H_{jk} = \frac{\partial^2 \ell(\theta)}{\partial \theta_j \theta_k} = \sum_{i=1}^{m} a''(\theta^T x^{(i)}) x_j^{(i)} x_k^{(i)}$$

$$\begin{aligned}z^T H z &= \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} a''(\theta^T x^{(i)}) x_j^{(i)} x_k^{(i)} z_j z_k \\ &= \sum_{i=1}^{m} a''(\theta^T x^{(i)}) [(x^{(i)})^T z]^2\end{aligned}$$

$$a''(\theta^T x) = Var[Y|X;\theta] \geq 0 \;\Rightarrow\; z^T H z \geq 0$$

5.

(a)

i.

$$W \in \mathbb{R}^{m \times m}$$

$$W_{ij} = \begin{cases} \frac{1}{2} w^{(i)} & i = j \\ 0 & i \neq j \end{cases}$$

ii.

$$\begin{aligned}\nabla_\theta J(\theta) &= \nabla_\theta (X\theta - y)^T W (X\theta - y) \\ &= \nabla_\theta (\theta^T X^T - y^T) W (X\theta - y) \\ &= \nabla_\theta (\theta^T X^T W X \theta - y^T W X \theta - \theta^T X^T W y + y^T W y) \\ &= \nabla_\theta (\theta^T X^T W X \theta - 2 y^T W X \theta) \\ &= 2 X^T W X \theta - 2 X^T W y\end{aligned}$$

$$\nabla_\theta J(\theta) = 0 \;\Rightarrow\; \theta = (X^T W X)^{-1} X^T W y$$

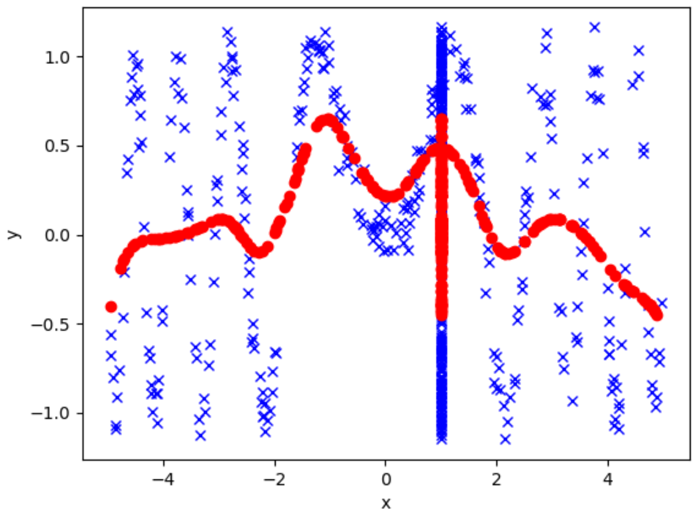
iii.

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{m} \log p(y^{(i)}|x^{(i)};\theta) \\ &= \sum_{i=1}^{m} -\log(\sqrt{2\pi}\sigma^{(i)}) - \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\end{aligned}$$

$$w^{(i)} = -\frac{1}{(\sigma^{(i)})^2}$$

$$\frac{\partial \ell(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} \frac{y^{(i)} - \theta^T x^{(i)}}{(\sigma^{(i)})^2} x_j^{(i)}$$

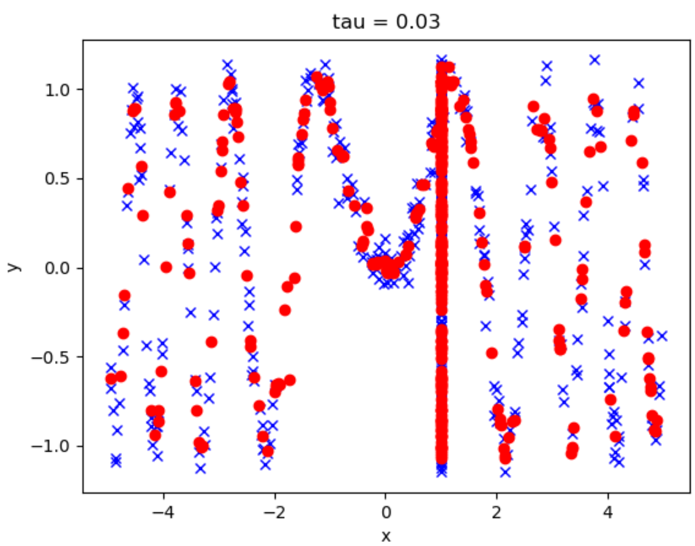
(b)

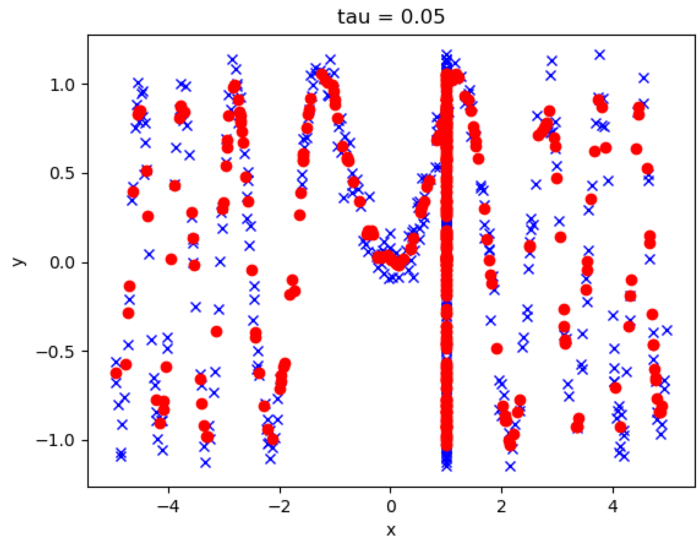


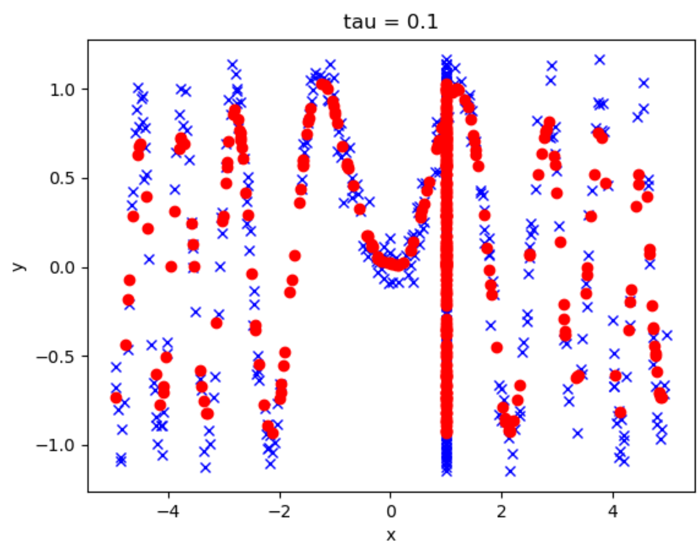
MSE=0.331.

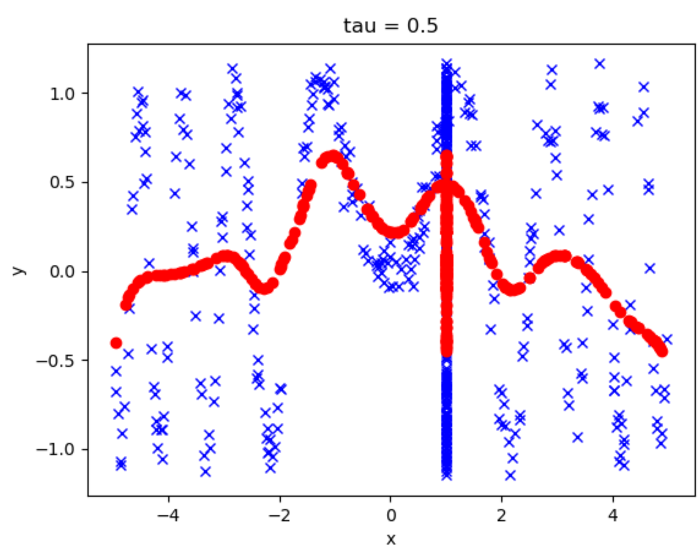
The model seems to be underfitting.

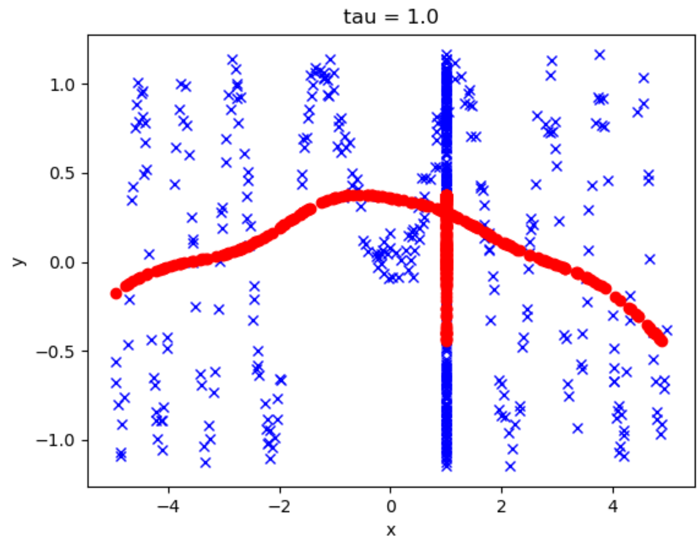
(c)

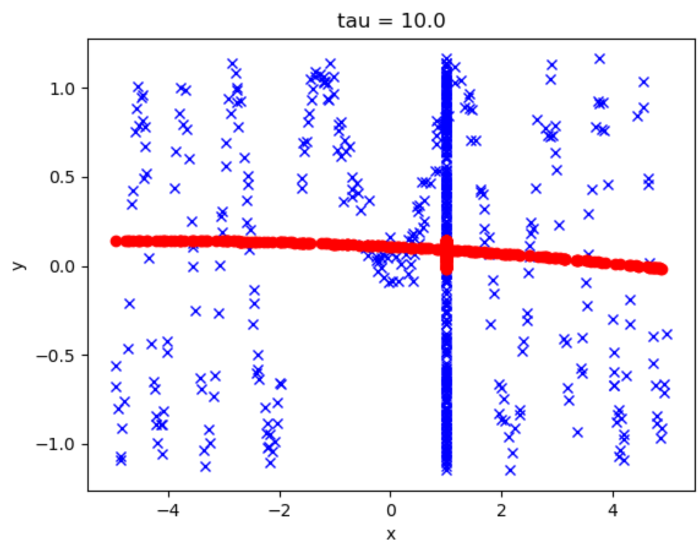



tau = 0.05
y
x
1.0
0.5
0.0
−0.5
−1.0
−4
−2
0
2
4


tau = 0.1
y
x
1.0
0.5
0.0
−0.5
−1.0
−4
−2
0
2
4


tau = 0.5
y
x
1.0
0.5
0.0
−0.5
−1.0
−4
−2
0
2
4





$\tau = 0.05$ achieves the lowest MSE on the valid set.

MSE=0.012 on the valid set, MSE=0.017 on the test set.