CS229 Problemset 01

KOH-

February 2025

1 Problem 01

$$
\begin{aligned}
(a): \frac{\partial g(\theta^T x)}{\partial \theta_j} &= g(\theta^T x) \cdot \big(1 - g(\theta^T x)\big) \cdot x_j^{(i} \\
J(\theta) &= -\frac{1}{m} \sum_{i=1}^{m} \Big[y_i \, \log\big(g(\theta^{\mathsf{T}} x_i)\big) + (1 - y_i) \, \log\big(1 - g(\theta^{\mathsf{T}} x_i)\big) \Big]. \\
\frac{\partial J(\theta)}{\partial \theta_j} &= -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \cdot \big(1 - g(\theta^T x^{(i)})\big) \cdot x_j^{(i)} + (y^{(i)} - 1) \cdot g(\theta^T x^{(i}) \cdot x_j^{(i)} \\
&= \frac{1}{m} \sum_{i=1}^{m} \big(g(\theta^T x^{(i)}) - y^{(i)}\big) \cdot x_j^{(i)} = \Phi(j) \\
\nabla_\theta J(\theta) &= \sum_{i=1}^{m} \big(g(\theta^{\mathsf{T}} x_i) - y_i\big) \, x_i, \\
H_{jk} &= \frac{\partial \Phi(j)}{\partial \theta_k} \\
&= \frac{1}{m} \sum_{i=1}^{m} x_j^{(i)} x_k^{(i)} \cdot g(\theta^T x^{(i)}) \cdot (1 - g(\theta^T x^{(i)})) \\
H &= \frac{1}{m} x^T W x, (W_{ii} = g(\theta^T x^{(i)}) \cdot (1 - g(\theta^T x^{(i)}) \geq 0)
\end{aligned}
$$

Obviously, W is PSD. then $\frac{1}{m} x^T W x = H$ is PSD too.

(b): Coding Problem

$$
\begin{aligned}
(c): P(x|y=1) &= \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1)) \\
P(y=1|x) &= \frac{P(x|y=1)P(y=1)}{P(x)} = \frac{P(x|y=1)P(y=1)}{P(x|y=0)P(y=0) + P(x|y=1)P(y=1)} \\
&= \frac{1}{1 + \exp(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1}(x-\mu_0) + \frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1) + \ln\frac{1-\phi}{\phi})} \\
&= \frac{1}{1 + \exp(-1(\cdot(\Sigma^{-1}(\mu_1 - \mu_0))^T x + \frac{1}{2}(\mu_0^T \Sigma^{-1}\mu_0 - \mu_1 \Sigma^{-1}\mu_1) + \ln\frac{\phi}{1-\phi}))}
\end{aligned}
$$

So $\theta = \Sigma^{-1}(\mu_1 - \mu_0), \theta_0 = \frac{1}{2}(\mu_0^T \Sigma^{-1}\mu_0 - \mu_1 \Sigma^{-1}\mu_1) + \ln\frac{\phi}{1-\phi}$

$$(d): P(y=0)=1-\phi, P(y=1)=\phi$$

$$\ell(\theta)=\sum_{i=1}^{m}\ln(P(x^{(i)}|y^{(i)}))+\ln(P(y^{(i)}))$$

$$P(x^{(i)}|y^{(i)};\mu_0,\mu_1,\Sigma)=\frac{1}{\sqrt{(2\pi\Sigma)}}\exp(-\frac{x^{(i)}-\mu^{(i)}}{2\Sigma})$$

$$\frac{\partial\ell}{\partial\phi}=\frac{\partial\sum_{i=1}^{m}\ln(\phi^{1\{y^{(i)}=1\}}+(1-\phi)^{1-1\{y^{(i)}=1\}})}{\partial\phi}$$

$$=\frac{1}{\phi}\sum_{i=1}^{m}1\{y^{(i)}=1\}-\frac{1}{1-\phi}\sum_{i=1}^{m}1-1\{y^{(i)}=1\}=0$$

$$\phi=\sum_{i=1}^{m}1\{y^{(i)}=1\}$$

$$\frac{\partial\ell}{\partial\mu_0}=\frac{\partial\sum_{i=1}^{m}1\{y^{(i)}=0\}\cdot(\frac{-(x^{(i)}-\mu^{(i)})^2}{2\Sigma})}{\partial\mu_0}=0$$

$$\mu_0=\frac{\sum_{i=1}^{m}1\{y^{(i)}=0\}\cdot x^{(i)}}{\sum_{i=1}^{m}1\{y^{(i)}=0\}}$$

$$for\ the\ same\ reason:$$

$$\mu_1=\frac{\sum_{i=1}^{m}1\{y^{(i)}=1\}\cdot x^{(i)}}{\sum_{i=1}^{m}1\{y^{(i)}=1\}}$$

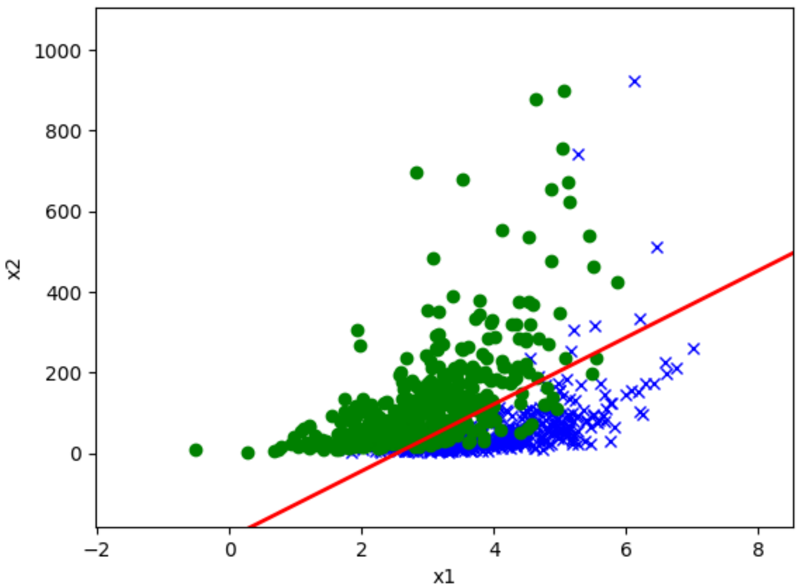
$$\frac{\partial\ell}{\partial\Sigma}=\frac{\partial\sum_{i=1}^{m}\ln(\frac{1}{\sqrt{2\pi\Sigma}})-\frac{(x^{(i)}-\mu^{(i)})^2}{2\Sigma}}{\partial\Sigma}=0$$

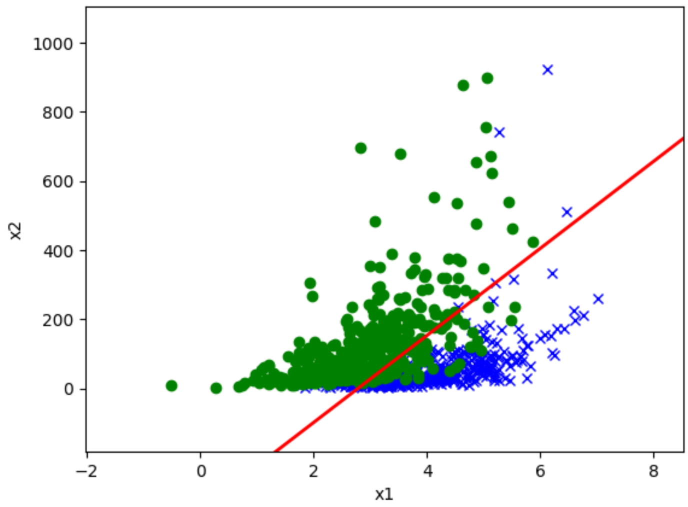
$$\sum_{i=1}^{m}-\frac{1}{\Sigma^{1/2}}+\frac{(x^{(i)}-\mu^{(i)})^2}{\Sigma^{3/2}}=0$$

$$\Sigma=\frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu^{(i)})^2=\sigma^2$$

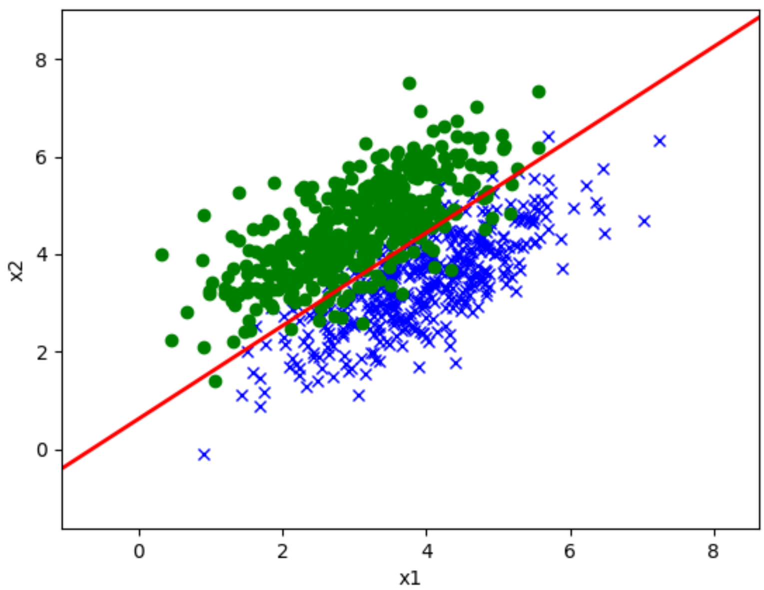
(e) Coding Problem

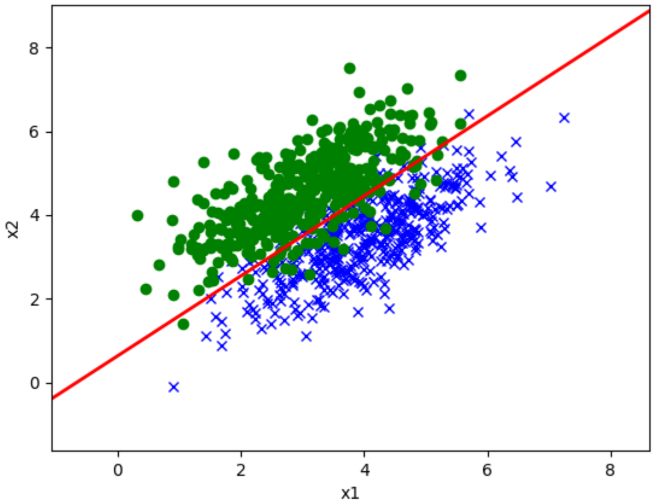
(f):





(g)





on dataset 1 GDA perform worse than logistic regression. Maybe because the data under the condition of $y = 1/0$ doesn't follow the Gaussian distribution.

(h): log or sqrt or Box-Cox transformation.

2 Problem 02

(a)

$$P(y^{(i)}=1|t^{(i)}=1,x^{(i)})=\frac{P(y^{(i)}=1,t^(i)=1,x^(i))}{P(t^{(i)}=1,x^{(i)})}=\frac{P(y^{(i)}=1,t^(i)=1,x^(i))}{P(t^{(i)}=1|x^{(i)})P(x^{(i)})}$$

$$P(t^{(i)}=1|t^{(i)}=1,x^{(i)})=\frac{P(t^{(i)}=1,y^(i)=1,x^(i))}{P(y^{(i)}=1,x^{(i)})}=\frac{P(t^{(i)=1},y^(i)=1,x^(i))}{P(y^{(i)}=1|x^{(i)})P(x^{(i)})}$$

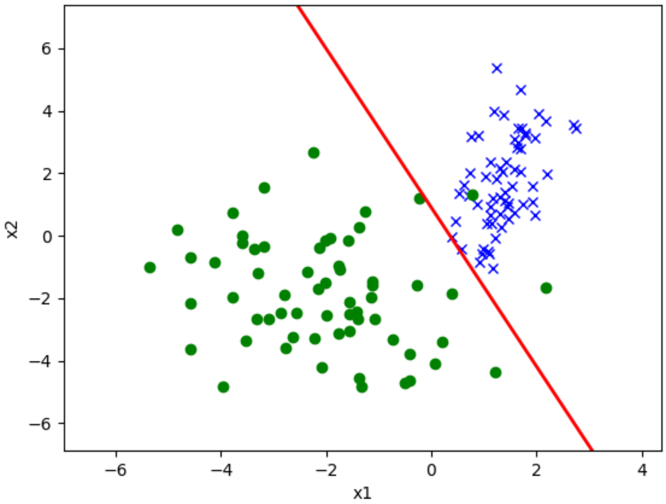
$$\alpha=\frac{P(y^{(i)}=1|x(i))}{P(t^{(i)}=1|x(i))}=\frac{P(y^{(i)}=1|t^{(i)}=1,x^{(i)})}{P(t^{(i)}=1|t^{(i)}=1,x^{(i)})}=P(y^{(i)}=1|t^{(i)}=1)$$

(b)

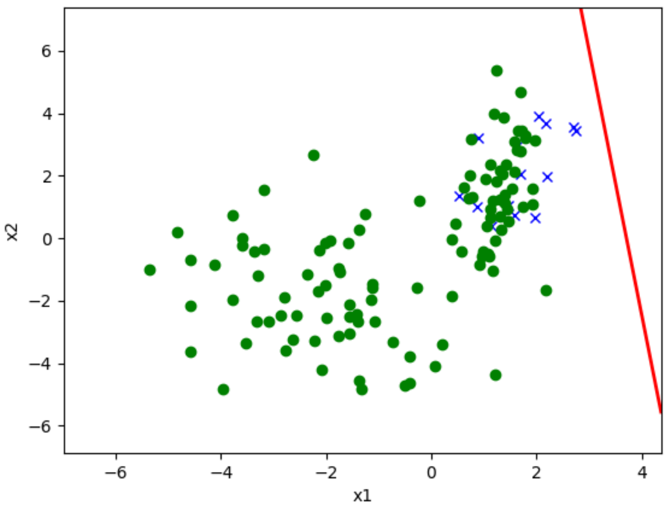
$$h(x^{(i)})\approx P(t^{(i)}=1|x^{(i)})$$

$$\alpha=\frac{P(y^{(i)}=1|x(i))}{P(t^{(i)}=1|x(i))}\approx P(y^{(i)}=1|x^{(i)})=h(x^{(i)})(x^{(i)}\in V_+)$$

(c)



(d)

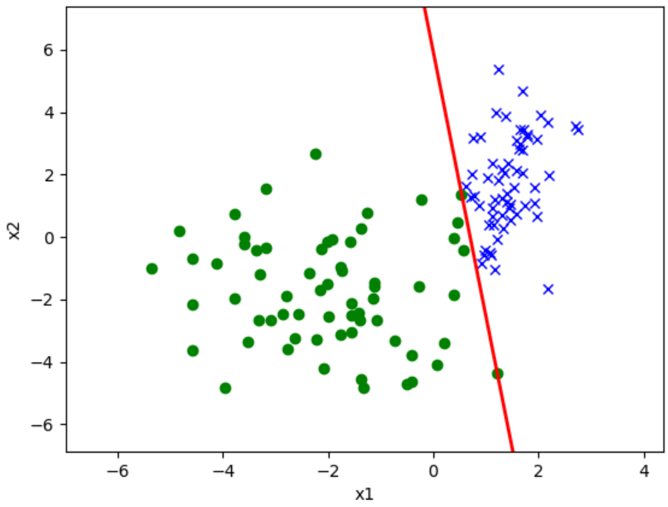


(e)

$$\theta_0 \cdot correctness + \theta_1 x_1 + \theta_2 x_2 = 0$$

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 = \ln(\frac{\alpha}{2-\alpha})$$

$$correctness = 1 + \ln(\frac{2-\alpha}{\alpha})/\theta_0$$



3 Problem 03

(a)

$$P(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!} = b(y)\exp(\eta^T T(y) - \alpha(\eta))$$

$$\begin{cases} b(y) = \dfrac{1}{y!} \\ \eta = \ln\lambda \\ T(y) = y \\ \alpha(\eta) = e^{\eta} \end{cases}$$

(b)

$g(\eta) = E(T(y)|x;\eta) = \lambda = e^{\theta^T x}$

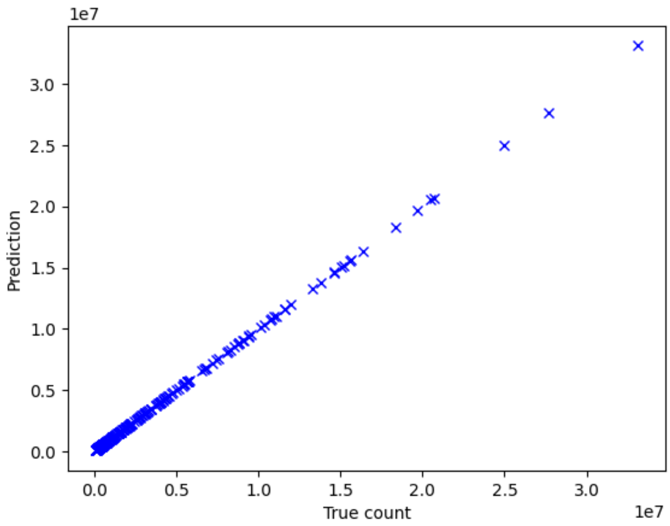
(c)

$$l(\theta) = -\lambda + y\ln\lambda - \sum_{i=1}^{y}\ln(y) = -e^{\theta^T x^{(i)}} + y\theta^T x^{(i)} - \sum_{i=1}^{y}\ln(y)$$

$$\frac{\partial l(\theta)}{\partial\theta_j} = (y - e^{\theta^T x^{(i)}})x_j^{(i)}$$

$$\theta_j' := \theta_j + \alpha\cdot(y - e^{\theta^T x^{(i)}})x_j^{(i)}$$

(d)



Due to the problem of overflow, I modified the learning rate to 1e-10.(Default value is 1e-7)

4 Problem 04

(a)

$$
\begin{aligned}
\int P(y;\eta)dy &= 1 \\
\frac{\partial 1}{\partial \eta} &= 0 \\
\int \frac{\partial P(y;\eta)}{\partial \eta}dy &= \int (y-\alpha'(\eta))P(y;\eta)dy = E(Y) - \alpha'(\eta) \\
so\ E(Y) &= \alpha'(\eta)
\end{aligned}
$$

(b)

$$
\begin{aligned}
\frac{\partial E(\eta)}{\partial \eta} &= \frac{\partial}{\partial \eta}\int yP(y;\eta)dy = \int y\frac{\partial P(y;\eta)}{\partial \eta}dy \\
&= \int y(y-\alpha'(\theta^T x^{(i)}))P(y;\eta)dy = E(Y^2) - [E(Y)]^2 = Var(Y) \\
\alpha''(\eta) &= Var(Y)
\end{aligned}
$$

(c)

$$
\begin{aligned}
\ell(\theta) &= -\sum_{i=1}^{m} P(y^{(i)}|x^{(i)};\theta) \\
&= -\sum_{i=1}^{m} \log(b(y)) + \theta^T x^{(i)} y^{(i)} - \alpha(\theta^T x^{(i)}) \\
\frac{\partial \ell(\theta)}{\partial \theta_j} &= -\sum_{i=1}^{m} (y^{(i)} - \alpha'(\theta^T x^{(i)}) x_j^{(i)}) \\
H_{jk} = \frac{\partial^2 \ell(\theta)}{\partial \theta_j \partial \theta_k} &= \sum_{i=1}^{m} (x_j^{(i)} \cdot x_k^{(i)} \cdot \alpha''(\theta^T x^{(i)})) \\
H &= X^T \Sigma X (\Sigma_{ii} = \alpha''(\theta^T x^{(i)}) = Var(y^{(i)} \geq 0) \\
H\ is\ PSD.
\end{aligned}
$$

5 Problem 05

(i)

$W_{ii} = \frac{1}{2}w^{(i)}$

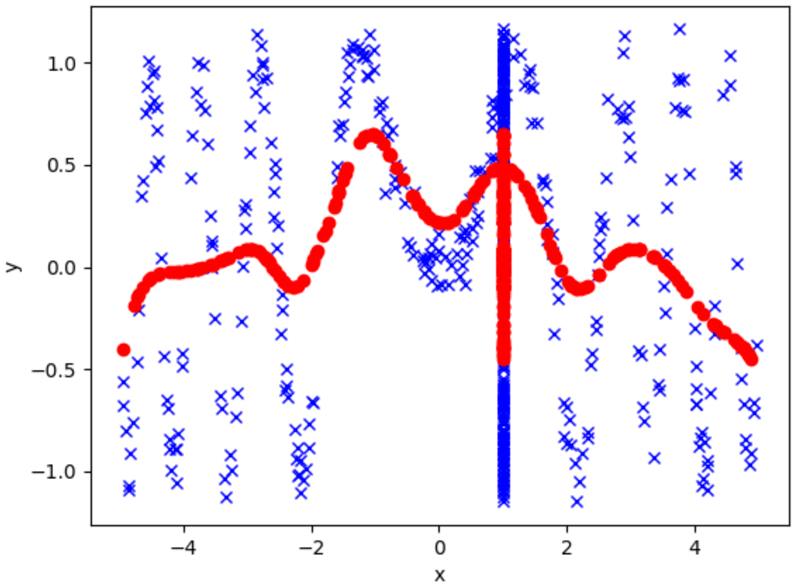
(ii)

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} w^{(i)} \cdot 2(\theta^T x^{(i)} - y^{(i)})x_j^{(i)}$$
$$\nabla_\theta J(\theta) = 2X^T w(X\theta - Y)$$
$$\theta = (X^T W X)^{-1} X^T Y$$

(iii)

$$\ell(\theta) = \sum_{i=1}^{m} (\log \frac{1}{\sqrt{2\pi}\sigma^{(i)}} - \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^{(i)}})$$
$$w^{(i)} = \frac{1}{(\sigma^{(i)})^2}$$

(b)



It's underfitting.

(c)

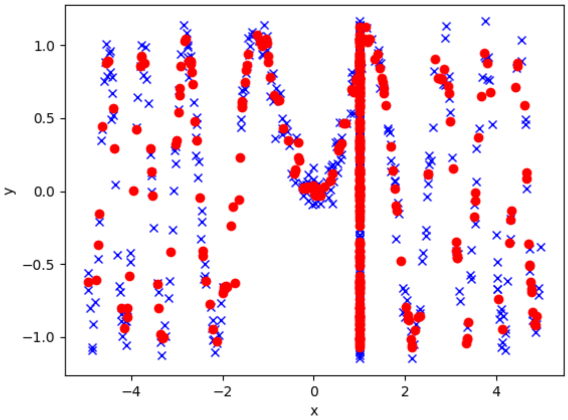


Figure 1: tau=0.02,MSE=0.018

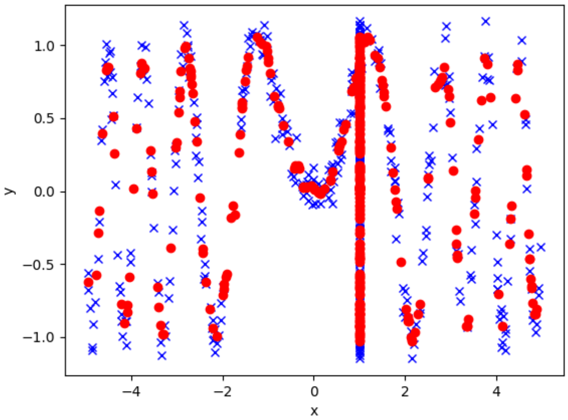


Figure 2: tau=0.05,MSE=0.012

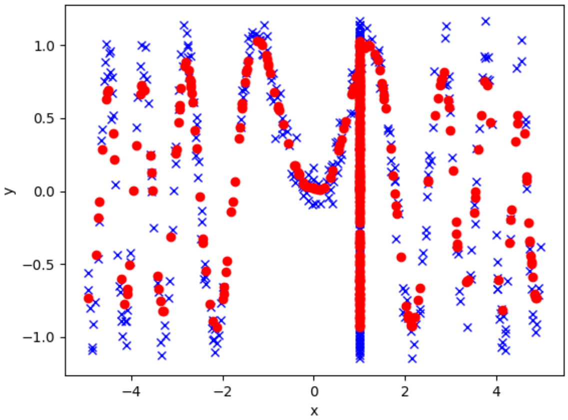


Figure 3: tau=0.1,MSE=0.24

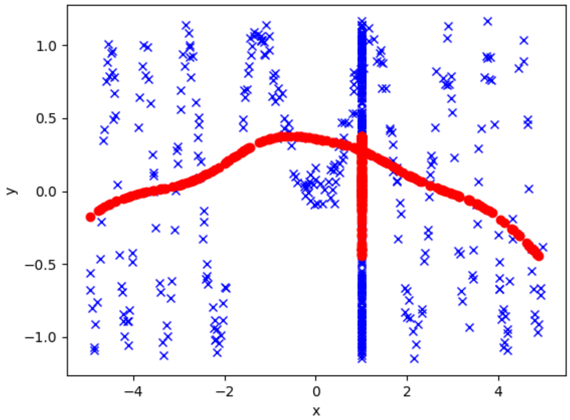


Figure 4: tau=1,MSE=0.4

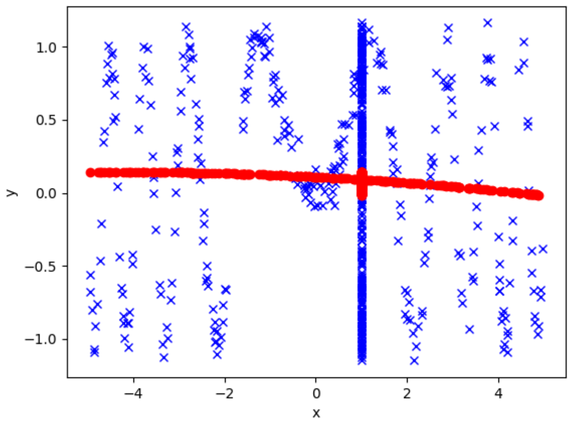


Figure 5: tau=10,MSE=0.43

When τ=0.05, it achieves the lowest MSE with 0.012.