ECE 595 HW2

Ruijie Song

Feb.17.2021

Exercise 1: Loading Data via Python

```
['index', 'male_bmi', 'male_stature_mm']
['0', '3.0', '1.679']
['1', '2.56', '1.586']
['2', '2.42', '1.773']
['3', '2.7399999999999998', '1.816']
['4', '2.59', '1.809']
['5', '2.5300000000000002', '1.662']
['6', '2.27', '1.829']
['7', '2.54', '1.686']
['8', '3.41', '1.761']
['9', '3.34', '1.797']

['index', 'female_bmi', 'female_stature_mm']
['0', '2.82', '1.563']
['1', '2.2199999999999998', '1.716']
['2', '2.71', '1.484']
['3', '2.81', '1.651']
['4', '2.55', '1.548']
['5', '2.3', '1.665']
['6', '3.56', '1.564']
['7', '3.1100000000000003', '1.676']
['8', '2.46', '1.69']
['9', '4.3', '1.704']
```

Figure 1. Exercise 1 result

Exercise 2: Build a Linear Classifier via Optimization

a.

2. a)

$\cancel{\hat{\theta} = (X^T X)^{-1} X^T y}$ $\qquad g_\theta = \theta^T x$

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{n=1}^{N} (y_n - g_\theta(x_n))^2 \qquad D = \{(x_n, y_n)\}_{n=1}^{N}$$

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{\|y - X\theta\|^2}_{\mathcal{E}_{train}(\theta)}$$

~~Over-determined~~ X: $\hat{\theta} = (A^T A)^{-1} A^T y$. $\qquad$ Full row rank

2 Technique: 1. Regularization

2. Pseudo-inverse

$$\therefore \hat{\theta} = (X^T X)^{-1} X^T y$$

b.

theta = [-1.070e+01; -1.233e-01; 6.674e+00]

c.

theta = [-1.070e+01; -1.233e-01; 6.674e+00]

d.

d)

$$\mathcal{E}(\theta) = \|y - X\theta\|^2 = \|X\theta - y\|^2$$

$$\cancel{\mathcal{E}\nabla} \quad \nabla \mathcal{E}_{train} = 2X^T(X\theta - y) = \underbrace{2X^T X}_{A}\theta - \underbrace{2X^T y}_{b}$$

$$\mathcal{E}(\theta^k - \alpha^k \nabla\mathcal{E}(\theta^k)) = \mathcal{E}(\theta^k + \alpha d^k)$$

$$= \|X(\theta^k + \alpha d^k) - y\|^2$$

$$\nabla_\alpha \mathcal{E}(\theta^k + \alpha d^k) = \frac{d}{d\alpha}\Big((\theta + \alpha d)^T X^T X(\theta + \alpha d) - 2y^T X(\theta + \alpha d) + \|y\|^2\Big)$$

$$= \frac{d}{d\alpha}\Big(\cancel{\theta X^T X\theta} + (\alpha d)^T X^T X\theta + \theta X^T X\alpha d + (\alpha d)^T X^T X \alpha d - \cancel{2y^T X\theta} - 2y^T X\alpha d + \cancel{\|y\|^2}\Big)$$

$$= 2\theta^T X^T X d + 2\alpha\|Xd\|^2 - 2y^T X d = 0$$

$$\alpha\|Xd\|^2 = y^T X d - \theta^T X^T X d$$

$$\therefore \alpha = \frac{y^T X d - \theta^T X^T X d}{\|Xd\|^2}$$

e.

theta = [-1.070e+01; -1.233e-01; 6.674e+00]

f.

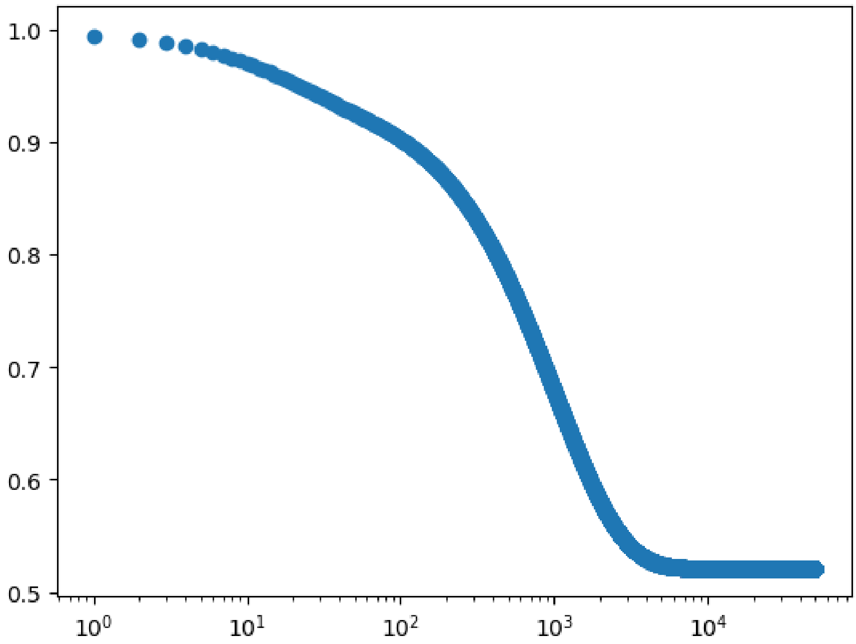


Figure 2. plot of the training loss

g.

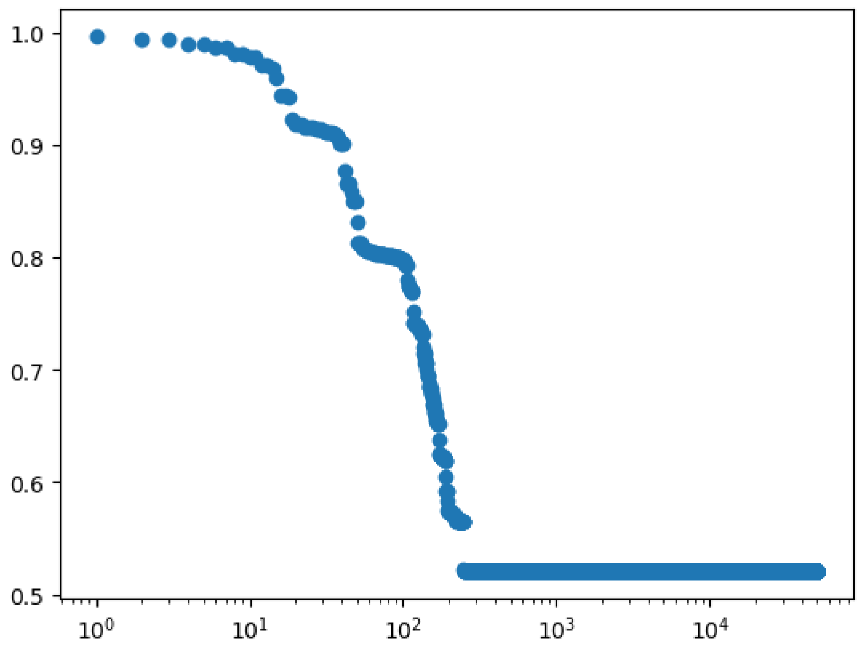
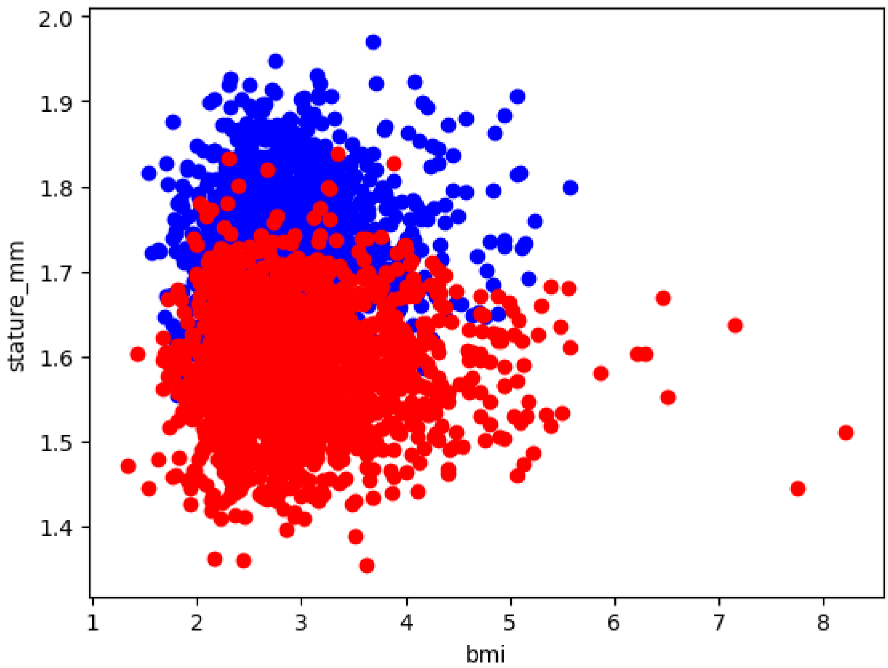


Figure 3. plot of the training loss

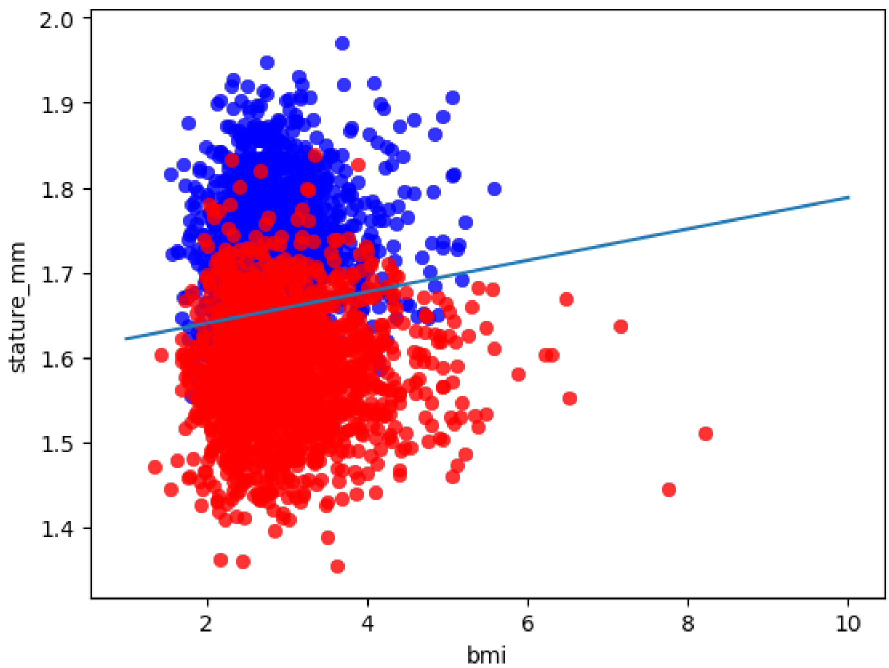
Exercise 3: Visualization and Testing

a.

i.



ii.

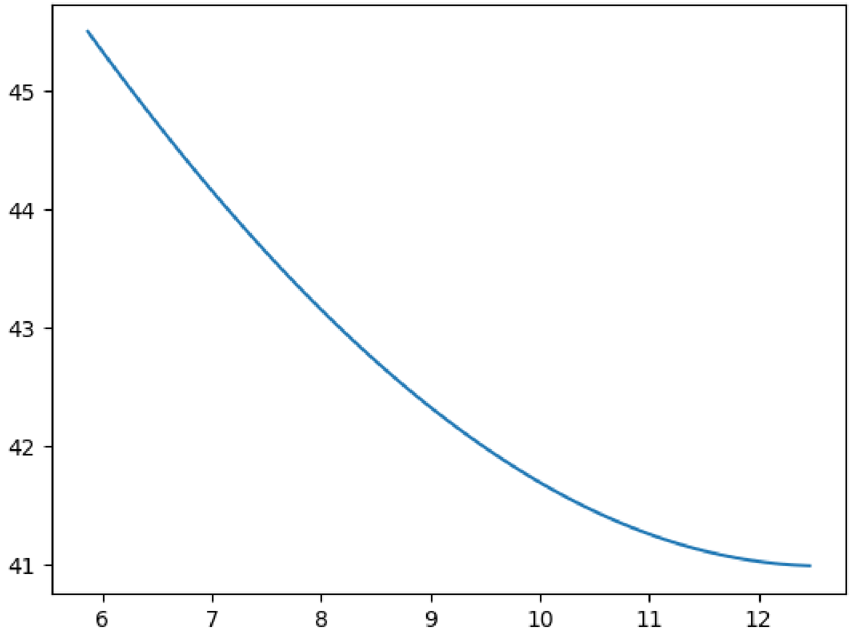


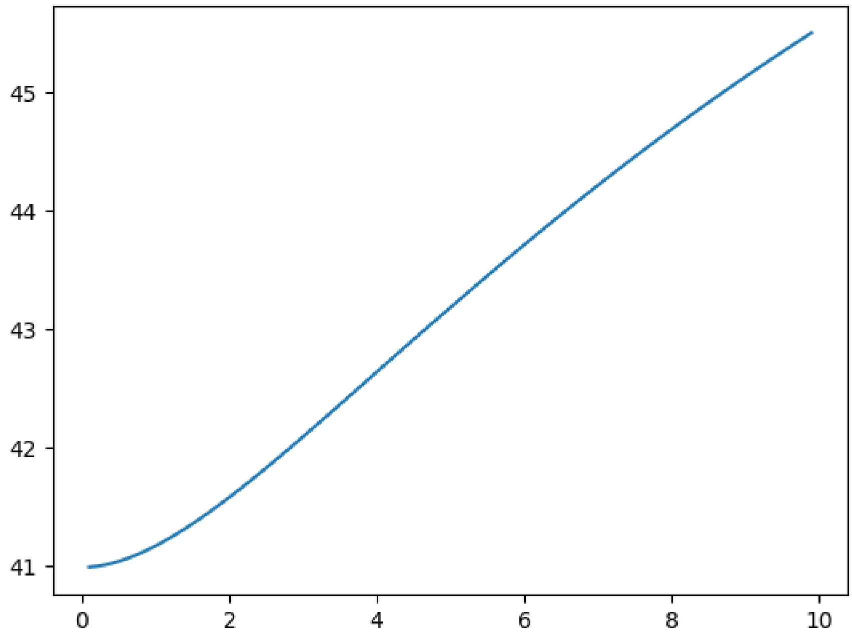
b

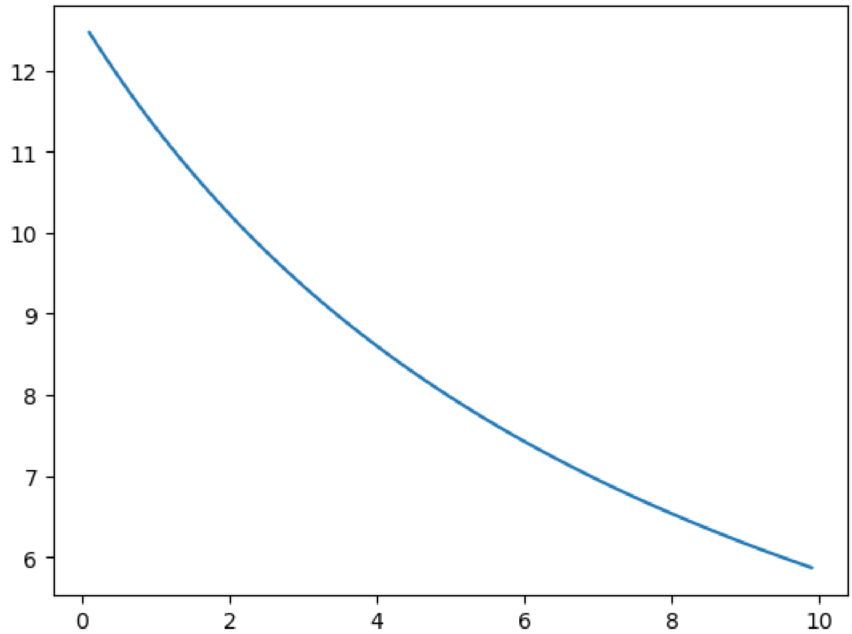
False Alarm = 15.024038461538846%
Miss = 18.653846153846154%
Precision = 0.9998804240631164
Recall = 0.9999036898201988

Exercise 4: Regularization

a.







b.

b) i) $\mathcal{L}(\underline{x}) = \|X\theta - y\|_2^2 + \lambda\|\theta\|_2^2$

ε $\quad \mathcal{L}(\underline{x}, \gamma_\alpha) = \|X\theta - y\|_2^2 - \sum_{i=1}^{m} \gamma_{\alpha i}(\alpha - \|\theta\|_2^2)$

$g_\alpha = \alpha - \|\theta\|_2^2$

$g_\varepsilon = \varepsilon - \|X\theta - y\|_2^2$

$\mathcal{L}(X, \gamma_\varepsilon) = \|\theta\|_2^2 - \sum_{i=1}^{m} \gamma_\varepsilon(\varepsilon - \|X\theta - y\|_2^2)$

ii) 1 $\nabla_x \mathcal{L}(X^*) = 0$, $\nabla_x \mathcal{L}(X^*, \gamma_\alpha^*) = 0$, $\nabla_x \mathcal{L}(X^*, \gamma_\varepsilon^*) = 0$

2 $\alpha - \|\theta^*\|_2^2 \geq 0$, $\varepsilon - \|X\theta^* - y\|_2^2 \geq 0$

3 $\gamma_\alpha^* \geq 0$, $\gamma_\varepsilon^* \geq 0$

4 $\gamma_\alpha^*(\alpha - \|\theta^*\|_2^2) = 0$, $\gamma_\varepsilon^*(\varepsilon - \|X\theta^* - y\|_2^2) =$

iii)

$$\hat{\theta}_\lambda = \underset{\theta \in \mathbb{R}^d}{\arg\min} \|X\theta - y\|_2^2 + \lambda \|\theta\|_2^2$$

$$\mathcal{L}(x) = \|X\theta - y\|_2^2 + \lambda \|\theta\|_2^2$$

$$\nabla_x \mathcal{L}(x) = 2\|X\theta - y\| = 0$$

$$\|X\theta - y| = 0 \Rightarrow X\theta = y \qquad \hat{\theta}_\lambda = X^{-1} y$$

$$\gamma_\alpha (\alpha - \|\hat{\theta}_\lambda\|_2^2) = 0 \Rightarrow r_\alpha (\alpha - \|X^{-1} y\|_2^2) = 0$$

$$\cancel{\gamma_{\alpha i} (\alpha - (x_i^{-1} y_i)^2) = 0}$$

$$\cancel{\text{if } \gamma_{\alpha i} = 0 \quad \alpha - (x_i^{-1} y_i)^2 > 0 \quad \alpha > (x_i^{-1} y_i)^2}$$

if $r_\alpha = 0$, $\alpha \geq \|X^{-1} y\|_2^2$

if $\alpha - \|X^{-1} y\|_2^2 = 0$, $r_\alpha > 0$

$$\alpha = \|X^{-1} y\|_2^2$$

iv) $\gamma_\varepsilon (\varepsilon - \|\cancel{XX^{-1} y - y}\|_2^2) = 0 \Rightarrow r_\varepsilon \cdot \varepsilon = 0$

$r_\varepsilon = 0$ or $\varepsilon = 0$

v) $\nabla_\theta \left(\|X\theta - y\|_2^2 - \cancel{\sum_{i=1}^{m}} \gamma_\alpha (\alpha - \|\theta\|_2^2) \right) = 0$

$$2(X\theta - y) + \gamma_\alpha \cdot 2\theta = 0 \qquad \cancel{2} X\theta - \cancel{2} y + r_\alpha \cdot \cancel{2}\theta = 0$$

$$(\cancel{2} X + r_\alpha)\theta = y \qquad \theta = (X + r_\alpha)^{-1} y$$

If $r_\alpha = 0$, . Yes

$$r_\alpha (\alpha - \|(X + r_\alpha)^{-1} y\|_2^2) = 0$$

If $r_\alpha > 0$, ~~Yes~~ No.

Since $r_\alpha > 0$

$$(X + r_\alpha)^{-1} y \neq X^{-1} y \neq \alpha$$

~~Yes~~