ECE 595: Homework 2

Yi Qiao, Class ID 187

(Spring 2019)

Exercise 1

(a)

Refer to code in the back.

(b)

The data was successfully read, shown by the screen below.

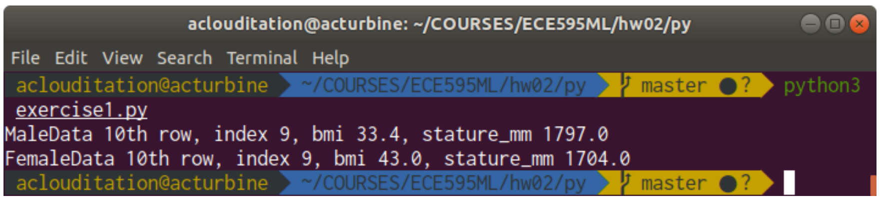


Figure 1: screenshot for reading data

Exercise 2

(a)

$$
\begin{aligned}
\begin{bmatrix} \boldsymbol{\omega}^* \\ \omega_0^* \end{bmatrix} &= \underset{\boldsymbol{\omega},\omega_0}{argmin} \sum_{j=1}^{N} (\boldsymbol{\omega}^T \boldsymbol{x}_j + \omega_0 - y_j)^2 \\
set\ \boldsymbol{\theta} &= \begin{bmatrix} \boldsymbol{\omega}^* \\ \omega_0^* \end{bmatrix} \\
\boldsymbol{\theta}^* &= \underset{\boldsymbol{\theta}}{argmin} \sum_{j=1}^{N} \left(\begin{bmatrix} \boldsymbol{x}_j^T & 1 \end{bmatrix} \boldsymbol{\theta} - y_j \right)^2
\end{aligned} \tag{1}
$$

thus,

$$
A = \begin{bmatrix} -\boldsymbol{x}_1- & 1 \\ -\boldsymbol{x}_2- & 1 \\ \cdots & \cdots \\ -\boldsymbol{x}_N- & 1 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{bmatrix}
$$

(b)

by least square

$$
\begin{aligned}
\boldsymbol{A}^T \boldsymbol{A} \boldsymbol{\theta}^* &= \boldsymbol{A}^T \boldsymbol{b} \\
\boldsymbol{\theta}^* &= (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{b}
\end{aligned} \tag{2}
$$

if A^TA is invertible, A needs to be full column rank (or $null(A) = 0$). TODO: find out how to avoid this issue

(c)

Refer to the code in the back.
The optimal weight θ^* is:

$$\boldsymbol{\theta}^* = \begin{bmatrix} -1.23396767e-2 \\ 6.67486843e-3 \\ -1.07017505e+1 \end{bmatrix}$$

(d)

Refer to the code in the back again.
The weight vector computed using CVXPY is the same as the one computed in the previous step.

$$\boldsymbol{\theta}^* = \begin{bmatrix} -1.23396767e-2 \\ 6.67486843e-3 \\ -1.07017505e+1 \end{bmatrix}$$

Exercise 3

(a)

(i)

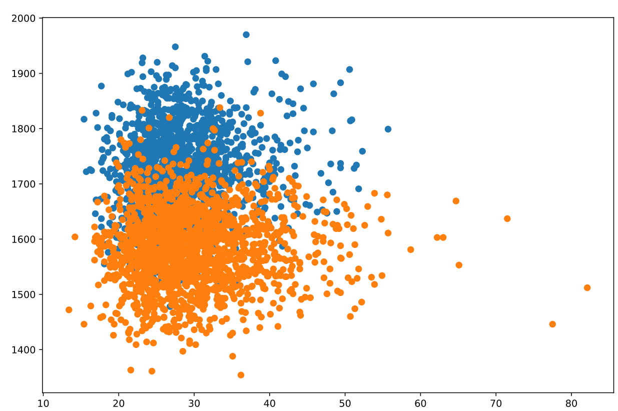


Figure 2: Training data

(ii)

$$\begin{aligned} \boldsymbol{\omega}^{*T}\boldsymbol{x} + \omega_0^* &= 0 \\ \omega_1^* x_1 + \omega_2^* x_2 + \omega_0^* &= 0 \\ x_2 &= -\frac{\omega_1^* x_1 + \omega_0^*}{\omega_2^*} \end{aligned} \tag{3}$$

(iii)

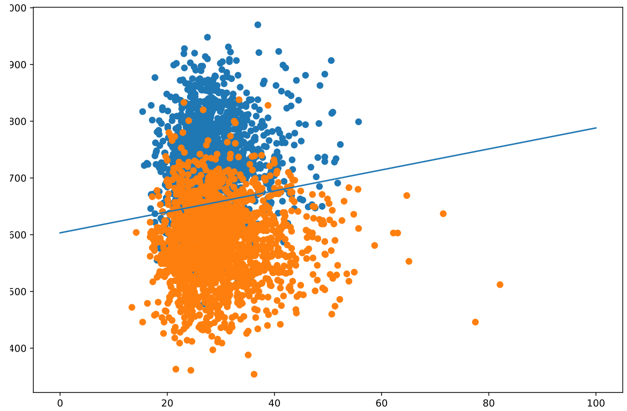


Figure 3: Training data with decision boundary

(b)

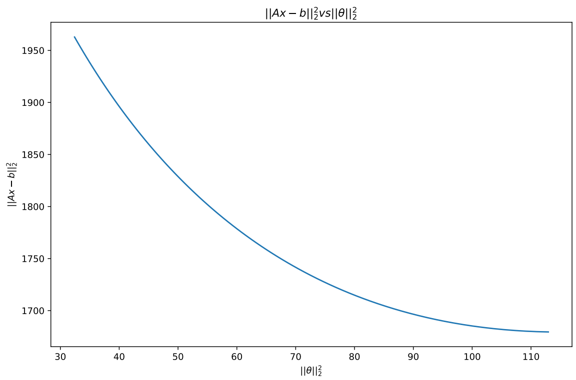
Refer to the code in the back, the success rate is:

$$success\ rate = 83.93\%$$

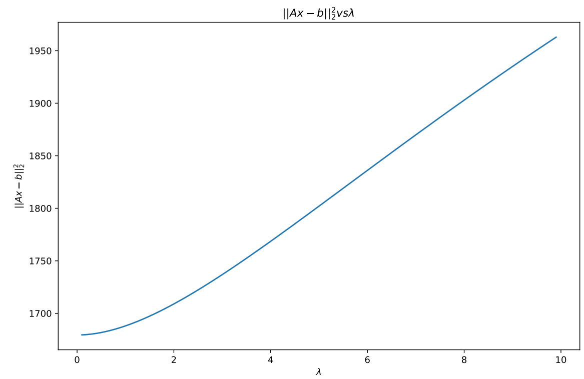
Exercise 4

(a)

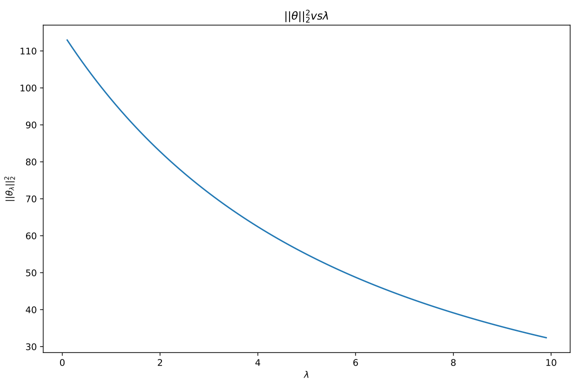
(i)



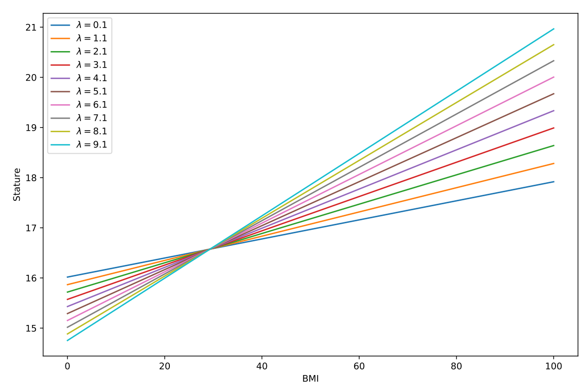
(a) 1



(b) 2



(c) 3



(d) 4

(ii)

TODO: fix this

(b)

(i)

for equation (8), $\boldsymbol{\theta}_\lambda = \underset{\boldsymbol{\theta}}{argmin}||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{\theta}||_2^2$:

$$\mathcal{L}(\boldsymbol{\theta}) = ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{\theta}||_2^2 \tag{4}$$

Since no constraint, the vertex can be found at:

$$\begin{aligned} \nabla_\theta||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 + \lambda\nabla_\theta||\boldsymbol{\theta}||_2^2 = 0 \\ 2\boldsymbol{A}^T(\boldsymbol{A\theta} - \boldsymbol{b}) + 2\lambda\boldsymbol{\theta} = 0 \\ \boldsymbol{A}^T\boldsymbol{A\theta} + \lambda\boldsymbol{\theta} = \boldsymbol{A}^T\boldsymbol{b} \\ \boldsymbol{\theta}_\lambda = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} \end{aligned} \tag{5}$$

for equation (9), $\boldsymbol{\theta}_\alpha = \underset{\boldsymbol{\theta}}{argmin}||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 \ subject\ to\ ||\boldsymbol{\theta}||_2^2 \leq \alpha,\ \alpha - ||\boldsymbol{\theta}||_2^2 \geq 0$:

$$\mathcal{L}(\boldsymbol{\theta}, \gamma) = ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 - \gamma(\alpha - ||\boldsymbol{\theta}||_2^2) \tag{6}$$

KKT conditions:

$$\begin{aligned}
\nabla_\theta||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 + \gamma\nabla_\theta(||\boldsymbol{\theta}||_2^2 - \alpha) = 0 & \qquad \textit{stationarity} \\
\alpha - ||\boldsymbol{\theta}||_2^2 \geq 0 & \qquad \textit{primal feasibility} \\
\gamma \geq 0 & \qquad \textit{dual feasibility} \\
\gamma(\alpha - ||\boldsymbol{\theta}||_2^2) = 0 & \qquad \textit{complementary slackness}
\end{aligned} \tag{7}$$

for equation (10), $\boldsymbol{\theta}_\epsilon = \underset{\boldsymbol{\theta}}{argmin}||\boldsymbol{\theta}||_2^2 \ subject\ to\ ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 \leq \epsilon,\ \epsilon - ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 \geq 0$:

$$\mathcal{L}(\boldsymbol{\theta}, \gamma) = ||\boldsymbol{\theta}||_2^2 - \gamma(\epsilon - ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2) \tag{8}$$

KKT conditions:

$$\begin{aligned}
\nabla_\theta||\boldsymbol{\theta}||_2^2 + \gamma\nabla_\theta(||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 - \epsilon) = 0 & \qquad \textit{stationarity} \\
\epsilon - ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 \geq 0 & \qquad \textit{primal feasibility} \\
\gamma \geq 0 & \qquad \textit{dual feasibility} \\
\gamma(\epsilon - ||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2) = 0 & \qquad \textit{complementary slackness}
\end{aligned} \tag{9}$$

(ii)

for equation (9), stationarity suggests that:

$$\nabla_\theta||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 + \gamma\nabla_\theta(||\boldsymbol{\theta}||_2^2 - \alpha) = 0 \tag{10}$$

by doing the algebra, we get:

$$\begin{aligned}
2\boldsymbol{A}^T(\boldsymbol{A\theta} - \boldsymbol{b}) + 2\gamma\boldsymbol{\theta} = 0 \\
\boldsymbol{\theta}_\alpha = (\boldsymbol{A}^T\boldsymbol{A} + \gamma\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b}
\end{aligned} \tag{11}$$

if we set $\gamma_\alpha = \lambda$, referring to equation (5) above:

$$\begin{aligned}
\boldsymbol{\theta}_\alpha &= (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} \\
&= \boldsymbol{\theta}_\lambda
\end{aligned} \tag{12}$$

where primal feasibility:

$$\alpha - ||\boldsymbol{\theta}_\alpha||_2^2 \geq 0$$

is satisfied if we choose $\alpha = ||\boldsymbol{\theta}_\lambda||_2^2$ since $\boldsymbol{\theta}_\alpha = \boldsymbol{\theta}_\lambda$,

$$\alpha - ||\boldsymbol{\theta}_\alpha||_2^2 = ||\boldsymbol{\theta}_\alpha||_2^2 - ||\boldsymbol{\theta}_\alpha||_2^2 = 0 \geq 0$$

Dual feasibility is automatically satisfied since we fix $\gamma_\alpha = \lambda > 0$
Complementary slackness is satisfied since $\alpha - ||\boldsymbol{\theta}_\alpha||_2^2 = 0$

Thus, the two problems (8) and (9) are equivalent.

(iii)

for equation (9), stationarity suggests that:

$$\nabla_{\theta}||\boldsymbol{\theta}||_2^2 + \gamma\nabla_{\theta}(||\boldsymbol{A\theta} - \boldsymbol{b}||_2^2 - \epsilon) = 0 \tag{13}$$

by doing the algebra, we get:

$$\begin{aligned} 2\boldsymbol{\theta} + 2\gamma\boldsymbol{A}^T(\boldsymbol{A\theta} - \boldsymbol{b}) = 0 \\ \boldsymbol{\theta}_{\epsilon} = (\gamma\boldsymbol{A}^T\boldsymbol{A} + \boldsymbol{I})^{-1}\boldsymbol{A}^T\gamma\boldsymbol{b} \\ \boldsymbol{\theta}_{\epsilon} = (\gamma^{-1}\gamma\boldsymbol{A}^T\boldsymbol{A} + \gamma^{-1}\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} \\ \boldsymbol{\theta}_{\epsilon} = (\boldsymbol{A}^T\boldsymbol{A} + \gamma^{-1}\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} \end{aligned} \tag{14}$$

if we set $\gamma_{\epsilon} = \lambda^{-1}$, referring to equation (5) above:

$$\begin{aligned} \boldsymbol{\theta}_{\epsilon} &= (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} \\ &= \boldsymbol{\theta}_{\lambda} \end{aligned} \tag{15}$$

where primal feasibility:

$$\epsilon - ||\boldsymbol{A\theta}_{\epsilon} - \boldsymbol{b}||_2^2 \geq 0$$

is satisfied if we choose $\epsilon = ||\boldsymbol{A\theta}_{\lambda} - \boldsymbol{b}||_2^2$ since $\boldsymbol{\theta}_{\epsilon} = \boldsymbol{\theta}_{\lambda}$,

$$\epsilon - ||\boldsymbol{A\theta}_{\epsilon} - \boldsymbol{b}||_2^2 = 0 \geq 0$$

Dual feasibility is automatically satisfied since we fix $\gamma_{\epsilon}^{-1} = \lambda > 0$, thus, $\gamma_{\epsilon} > 0$
Complementary slackness is satisfied since $\epsilon - ||\boldsymbol{A\theta}_{\epsilon} - \boldsymbol{b}||_2^2 = 0$

Thus, the two problems (8) and (10) are also equivalent.

(c)

(i)

(ii)

(d)