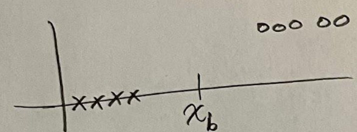
Exercise 1

1.i) Think of the 1D case.

$h_\Theta(x) = \frac{1}{1+\exp\{-w_1 x + w_0\}}$. If our data is linearly seperable, it means there is a value x_0 where class A is $< x_0$ and class B is $> x_0$



To minimize the loss, are sigmoid would need to look like a step function. This happens when $w \to \infty$.

In higher dimensions, we still linearly combine each dimension into one scaler value. If our weights approach infinity, we will get a higher dimension step function

1.ii) $|w_0| < C_2$ is sort of like our decision boundry. We don't want to limit this or else our decision boundry ~~can~~ ~~will~~ can be in the wrong location

$||w_1|| < C_1$ would make our ~~decis~~ decision boundry less sharp. This shouldn't hurt our overall predictions, but we may not be able to be as confident in our predictions

Other way $\hat{\Theta} = \underset{\Theta}{\arg\min} -\sum_{n=1}^{N}\{y_n \log h_\Theta(x_n) + (1-y_n)(1-\log h_\Theta(x_n))\} + \lambda||\Theta||_2$

iii) No, this is because logistic regression we are not looking for an exact decision boundry. We can think of the logistic regression as almost the probability a data point belongs to a class. If the dataset is linearly sep. this boundry is clear and the sigmoid is sharp where the boundry is at. ~~However~~ Other methods aim to find the decision boundry and this issue does not arise

Exercise 2

2a)

2a) $$J(\vec{\theta}) = -\sum_{n=1}^{N}\left\{ y_n \log h_{\vec{\theta}}(\vec{x}_n) + (1-y_n)\log(1-h_{\vec{\theta}}(\vec{x}_n)) \right\}$$

$$= -\sum_{n=1}^{N}\left\{ y_n \log \frac{h_{\vec{\theta}}(\vec{x}_n)}{1-h_{\vec{\theta}}(\vec{x}_n)} + \log(1-h_{\vec{\theta}}(\vec{x}_n)) \right\}$$

$$\log \frac{h_{\vec{\theta}}(\vec{x}_n)}{1-h_{\vec{\theta}}(\vec{x}_n)} = \log\left\{\frac{1/(1+\exp(-\vec{\theta}^T\vec{x}))}{1-1/(1+\exp(-\theta^T\vec{x}))}\right\} = \log \frac{1}{1+\exp(-\vec{\theta}^T\vec{x})-1} = -\vec{\theta}^T\vec{x}$$

$$J(\vec{\theta}) = -\sum_{n=1}^{N}\left\{-y_n\vec{\theta}^T\vec{x}_n + \log(1-h_{\vec{\theta}}(\vec{x}_n))\right\}$$

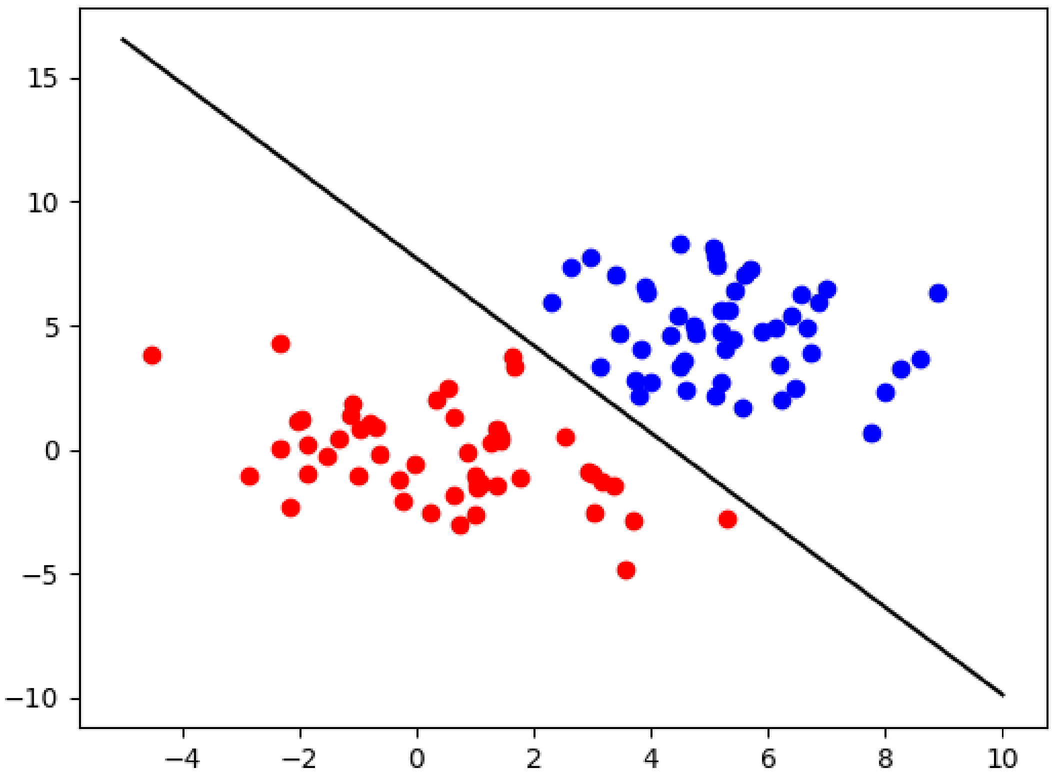
$$\log\left(1-\frac{1}{1+e^{-\theta^T x}}\right) = \log\left(\frac{(\cancel{1}+e^{-\vec{\theta}^T x})-\cancel{1}}{1+e^{-\theta^T x}}\right) = -\log\left(\frac{1+e^{-\vec{\theta}^T\vec{x}}}{e^{-\vec{\theta}^T\vec{x}}}\,\frac{e^{\vec{\theta}^T x}}{e^{\vec{\theta}^T x}}\right) = -\log(1+e^{\vec{\theta}^T x})$$

$$J(\vec{\theta}) = -\sum_{n=1}^{N}\left\{y_n\vec{\theta}^T x - \log(1+e^{\vec{\theta}^T x})\right\} = -\left(\sum_{n=1}^{N} y_n x_n\right)^T\vec{\theta} + \sum_{n=1}^{N}(1+e^{\vec{\theta}^T x})$$

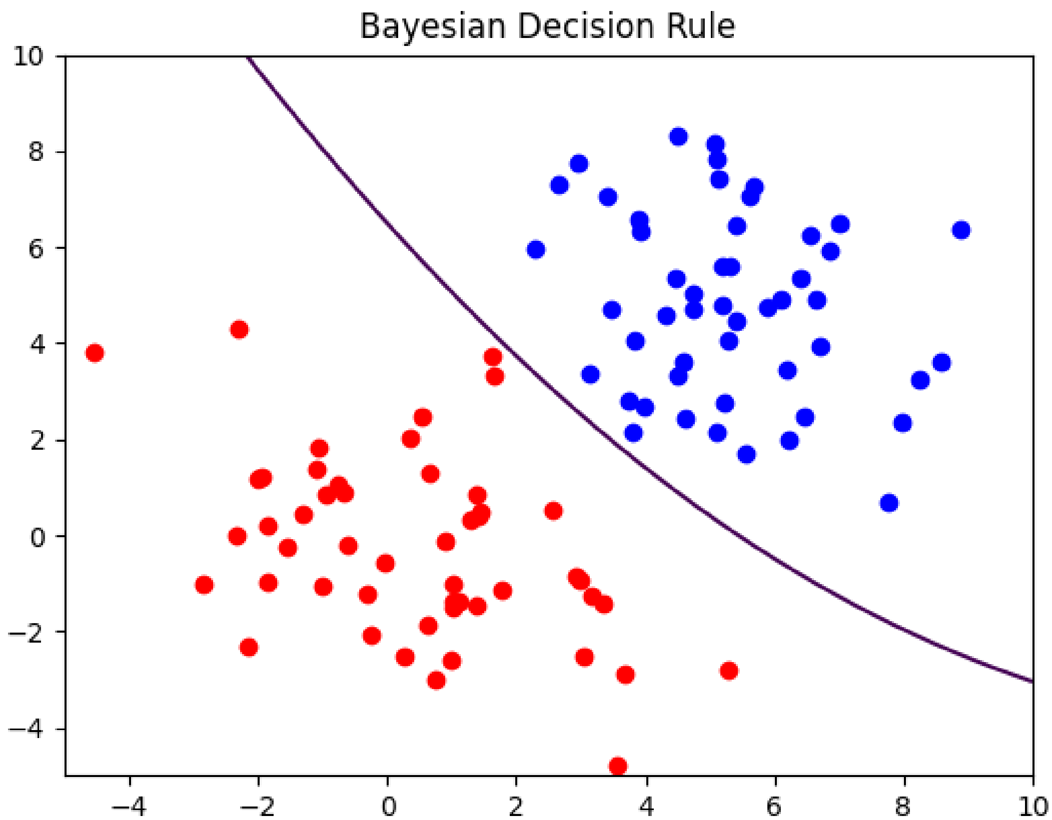
$$w_0x_1 + w_1x_2 + w_2 = 0$$

$$x_2 = \frac{-w_2 - w_0x_1}{w_1}$$

2b-c)



2d)



Exercise 3

a)

1.00	0.37	0.02	0.00	0.00	0.00
0.37	1.00	0.37	0.02	0.00	0.00
0.02	0.37	1.00	0.37	0.02	0.00
0.00	0.02	0.37	1.00	0.37	0.02
0.00	0.00	0.02	0.37	1.00	0.37
0.00	0.00	0.00	0.02	0.37	1.00

b)

c)
First 2 alphas:
[-0.95245074 -1.21046707]

d)