# Homework 1

Runmin Lu

September 4, 2021

## Exercise 1.3

### (a)

Given that $\mathbf{x}(t)$ is misclassified by $\mathbf{w}(t)$, we know that $y(t) \neq \text{sign}(\mathbf{w}^T(t)\mathbf{x}(t))$. Either $y(t) > 0$ and $\mathbf{w}^T(t)\mathbf{x}(t) < 0$ or $y(t) < 0$ and $\mathbf{w}^T(t)\mathbf{x}(t) > 0$. Their product must be negative.

### (b)

$$\begin{aligned}
y(t)\mathbf{w}^T(t+1)\mathbf{x}(t) &= y(t)(\mathbf{w}(t) + y(t)\mathbf{x}(t))^T\mathbf{x}(t) \\
&= y(t)\mathbf{w}^T(t)\mathbf{x}(t) + y(t)y(t)\mathbf{x}^T(t)\mathbf{x}(t) \\
&= y(t)\mathbf{w}^T(t)\mathbf{x}(t) + y(t)^2||\mathbf{x}(t)||^2 \\
&> y(t)\mathbf{w}^T(t)\mathbf{x}(t) \\
&\text{because } y(t)^2||\mathbf{x}(t)||^2 > 0 \text{ since } y(t) \neq 0, \mathbf{x}(t) \neq \mathbf{0}
\end{aligned}$$

### (c)

In order to classify a datum correctly, we need $y(t)$ and $\mathbf{w}^T(t)\mathbf{x}(t)$ to have the same sign. Equivalently, we want their product to be positive. After moving $\mathbf{w}(t)$ to $\mathbf{w}(t+1)$, we make the product greater, which is more positive, as shown in part (b).

## Exercise 1.5

(a) learning

(b) design

(c) learning

(d) design

(e) design

# Exercise 1.6

## (a)

supervised

input: each user's browsed books

output: whether he/she read or ignored the books

## (b)

reinforcement

input: games played against human

some output: win/loss

grade: win = +1, loss = -1

## (c)

supervised

input: existing movies

output: their classifications

## (d)

unsupervised

input: existing music

## (e)

supervised

input: existing customers

output: their maximum debt

# Exercise 1.7

## (a)

The hypothesis that always returns $\bullet$ gets picked.

3 points: 1

2 points: 3

1 point: 3

none: 1

### (b)

The hypothesis that always returns $\circ$ gets picked.

3 points: 1

2 points: 3

1 point: 3

none: 1

### (c)

$g(101) = \circ$

$g(110) = \circ$

$g(111) = \bullet$

3 points: 1

2 points: 3

1 point: 3

none: 1

### (d)

$g(101) = \bullet$

$g(110) = \bullet$

$g(111) = \circ$

3 points: 1

2 points: 3

1 point: 3

none: 1

## Problem 1.1

$$\begin{aligned} P(\text{first black}) &= \frac{3}{4} \\ P(\text{two blacks}) &= \frac{1}{2} \\ P(\text{two blacks}|\text{first black}) &= \frac{1}{2}/\frac{3}{4} \\ &= \boxed{\frac{2}{3}} \end{aligned}$$

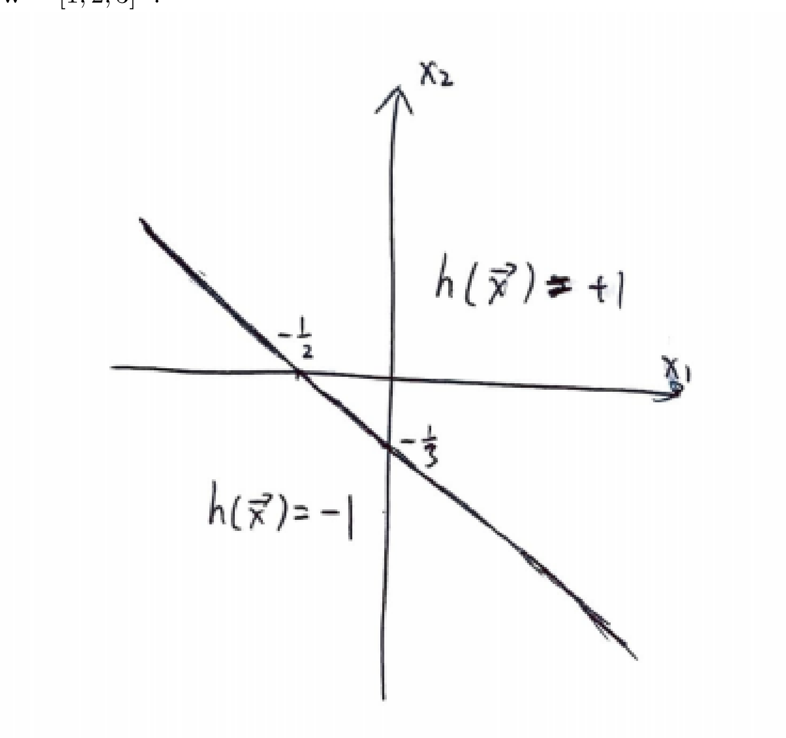
## Problem 1.2

### (a)

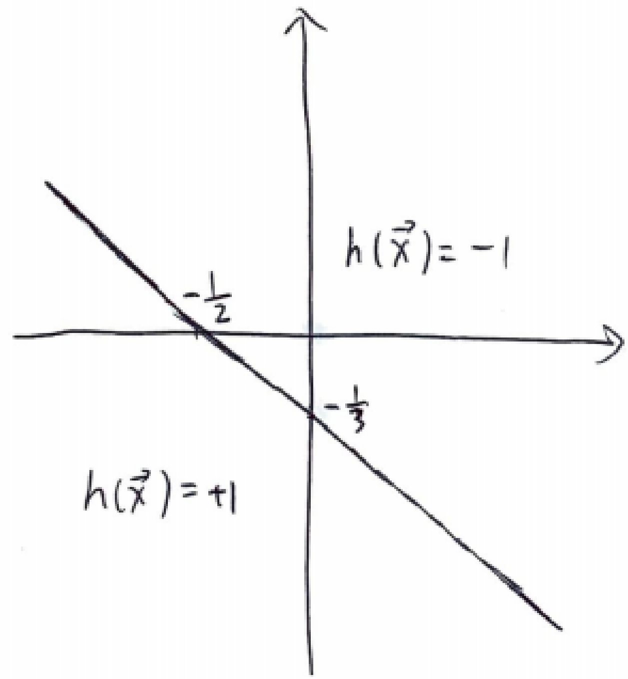
$$
\begin{aligned}
\text{What separates the region is represented by } \mathbf{w}^T\mathbf{x} = 0 \\
\text{which is a line: } w_0 + w_1 x_1 + w_2 x_2 = 0 \\
w_2 x_2 = -w_1 x_1 - w_0 \\
w_2 = -\frac{w_1}{w_2} x_1 - \frac{w_0}{w_2} \\
a = -\frac{w_1}{w_2} \\
b = -\frac{w_0}{w_2}
\end{aligned}
$$

### (b)

$\mathbf{w} = [1, 2, 3]^T$:

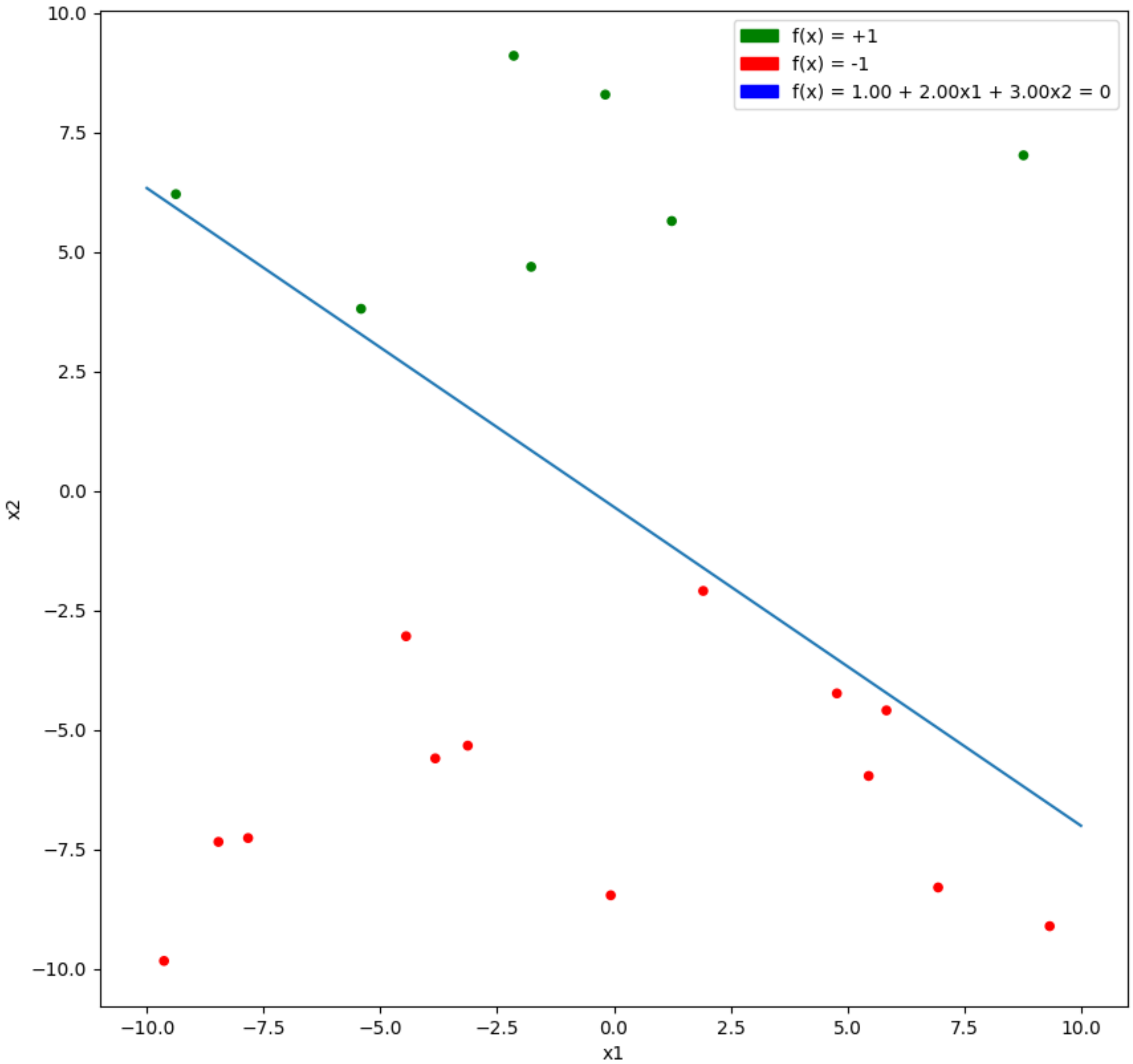



$\mathbf{w} = -[1, 2, 3]^T$:

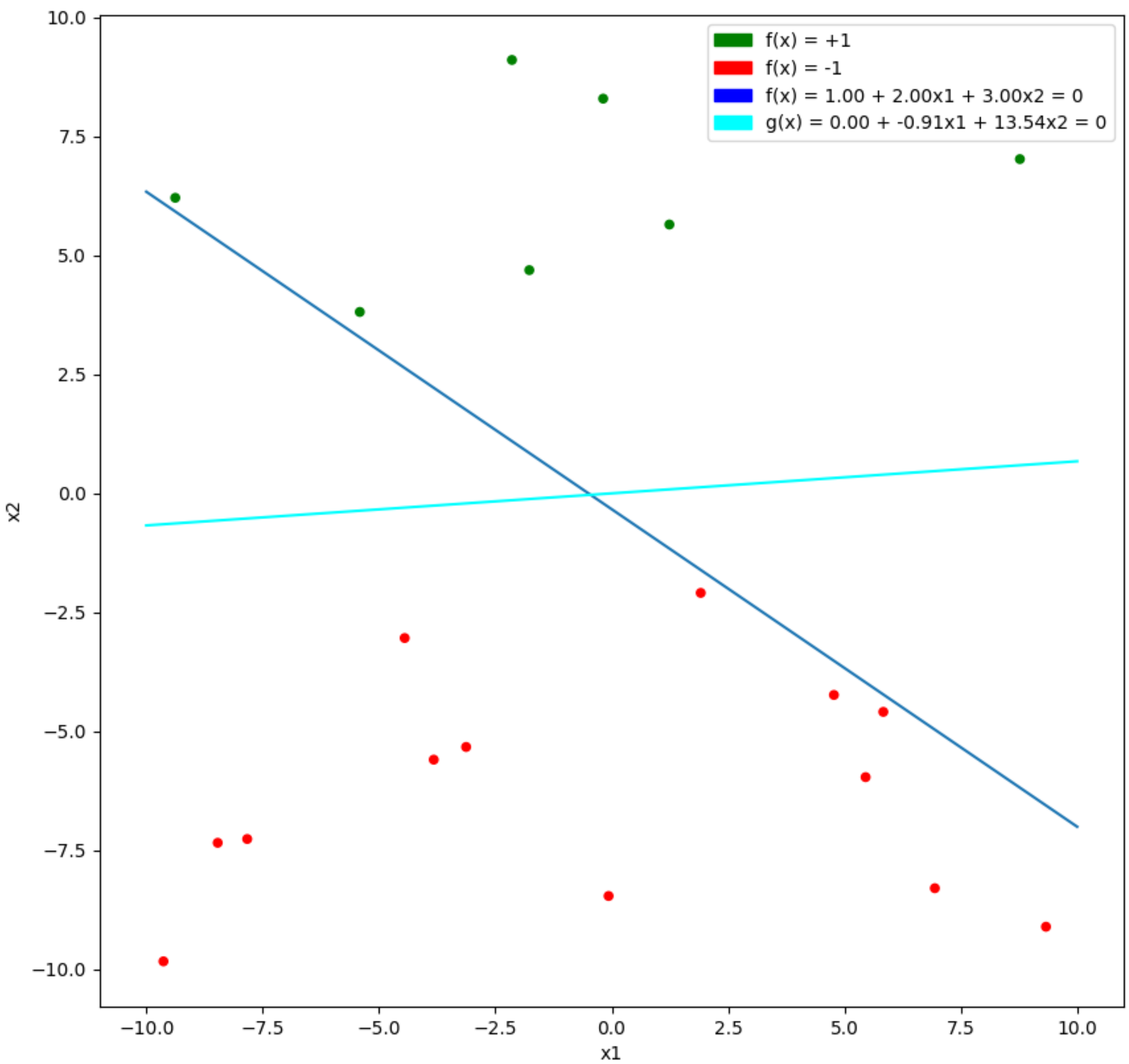

$h(\vec{x}) = -1$
$-\frac{1}{2}$
$-\frac{1}{3}$
$h(\vec{x}) = +1$

## Problem 1.4

### (a)

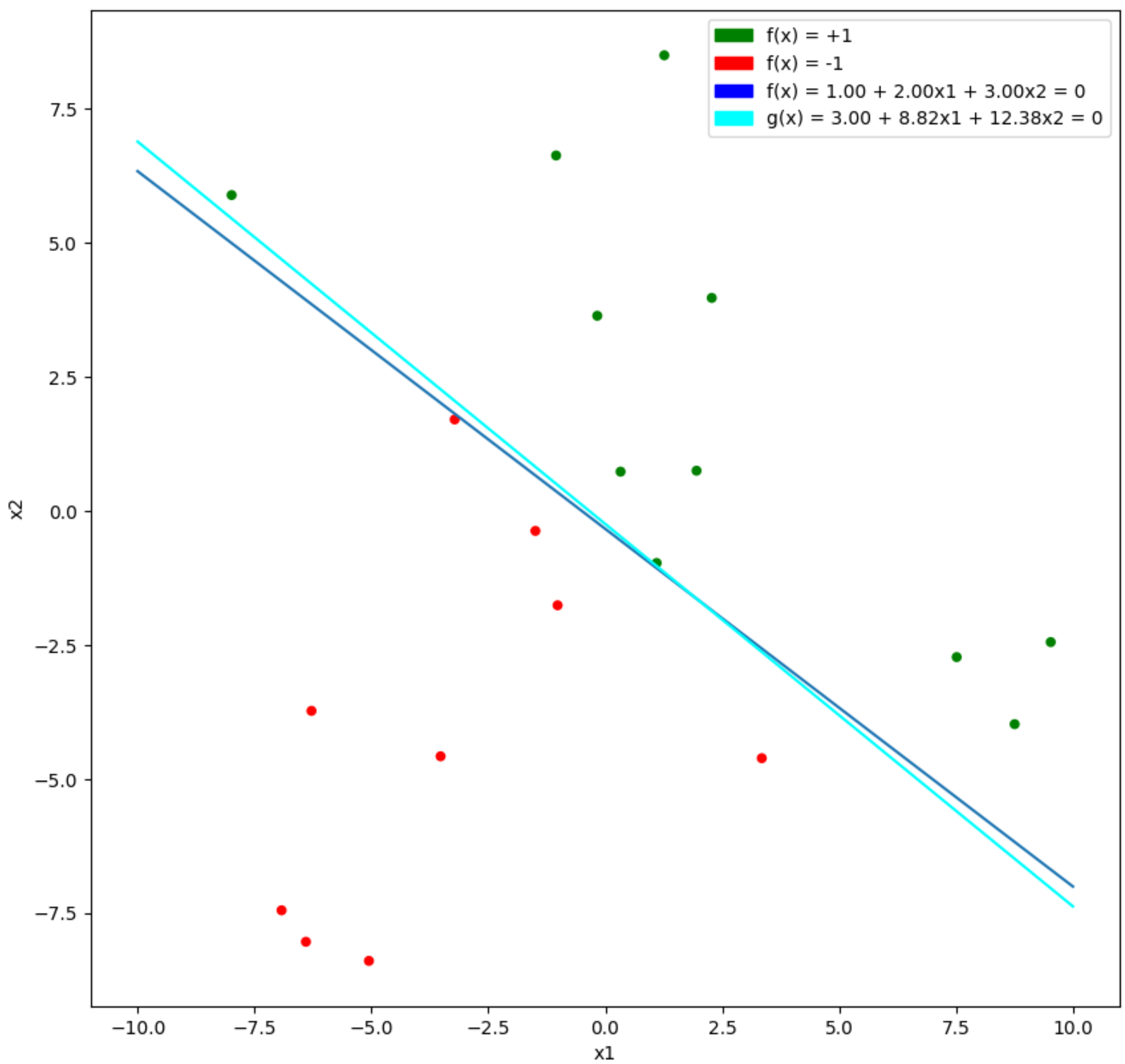


(b)



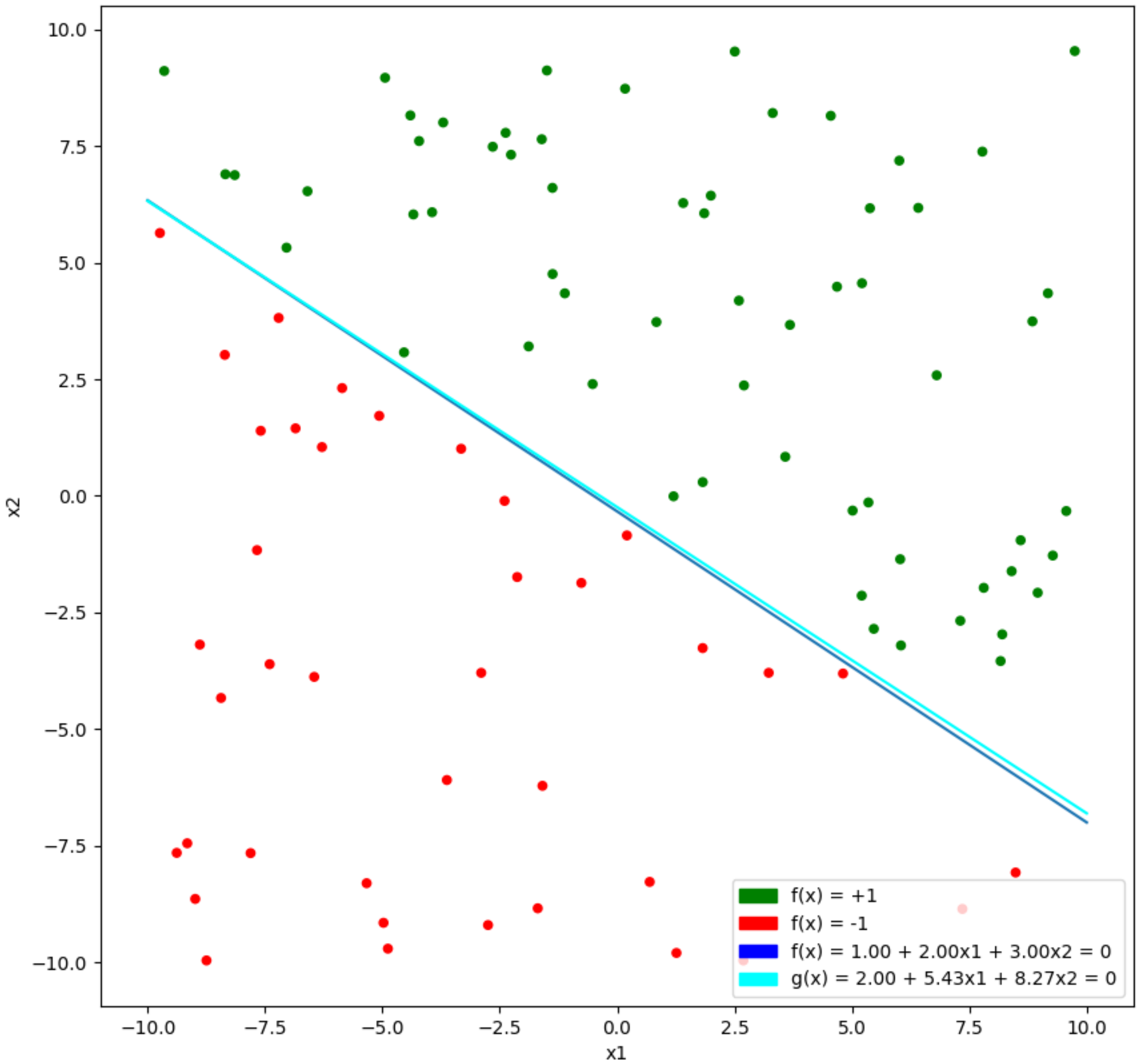

$g$ is not close to $f$ at all because the size of the data is not large enough.

**(c)**



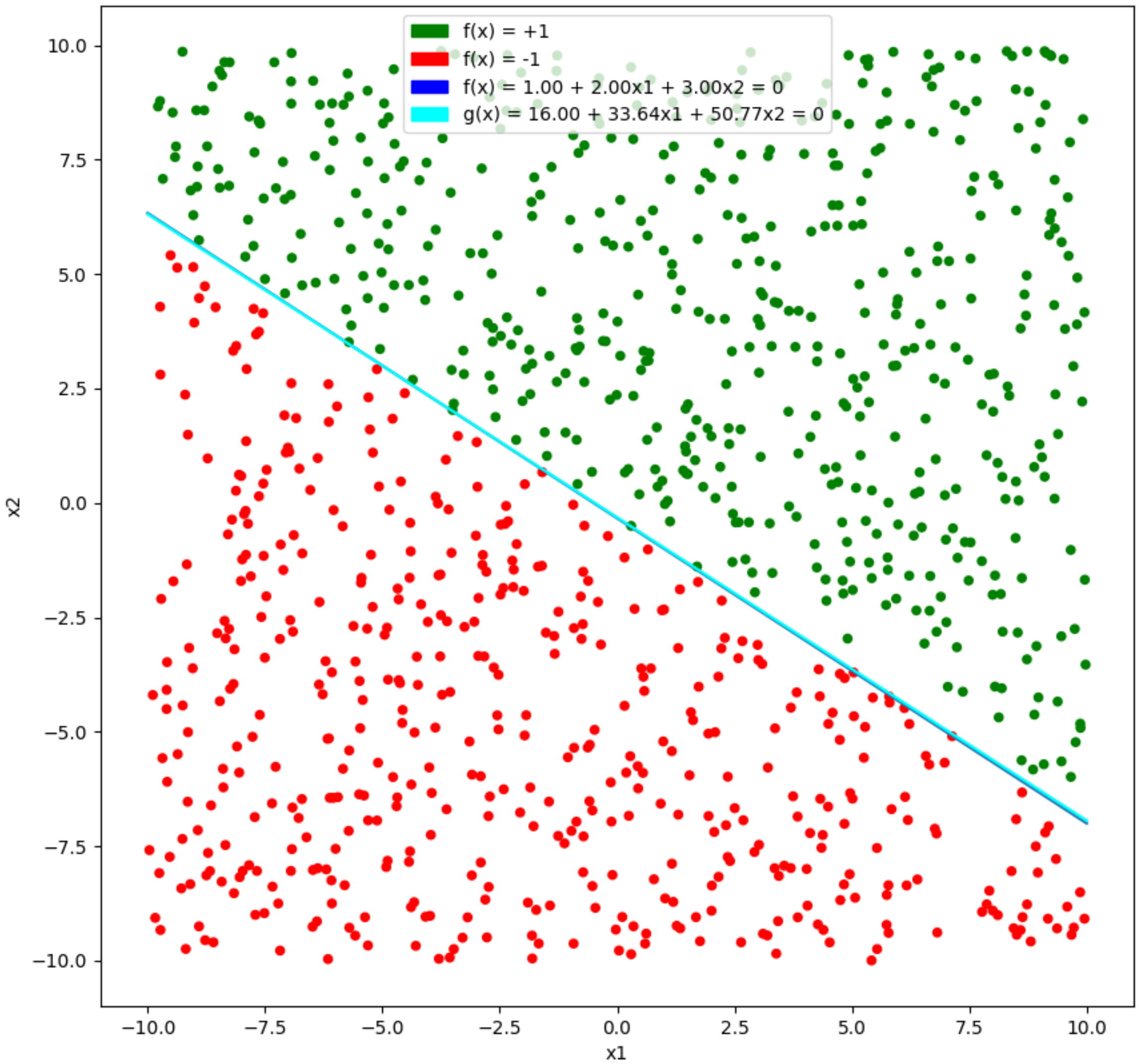

This time $g$ is pretty close to $f$ because we got lucky that the data is better structured.

**(d)**




$g$ is even closer to $f$ this time because the data size is much larger, which gives $g$ less freedom to move around without misclassifying any datum.

**(e)**




In this case $g \approx f$ because the data size is even larger.