Yanan Zhou 4366462176 HW6 — ISLR Questions

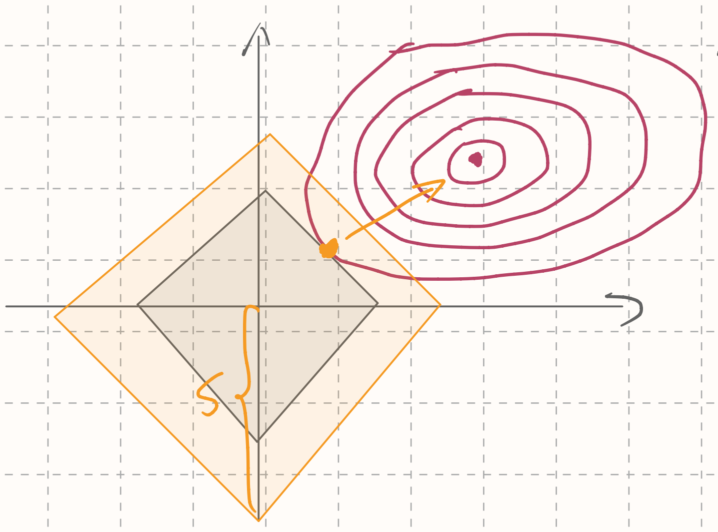
ISLR 6.6.3

(a). iv.

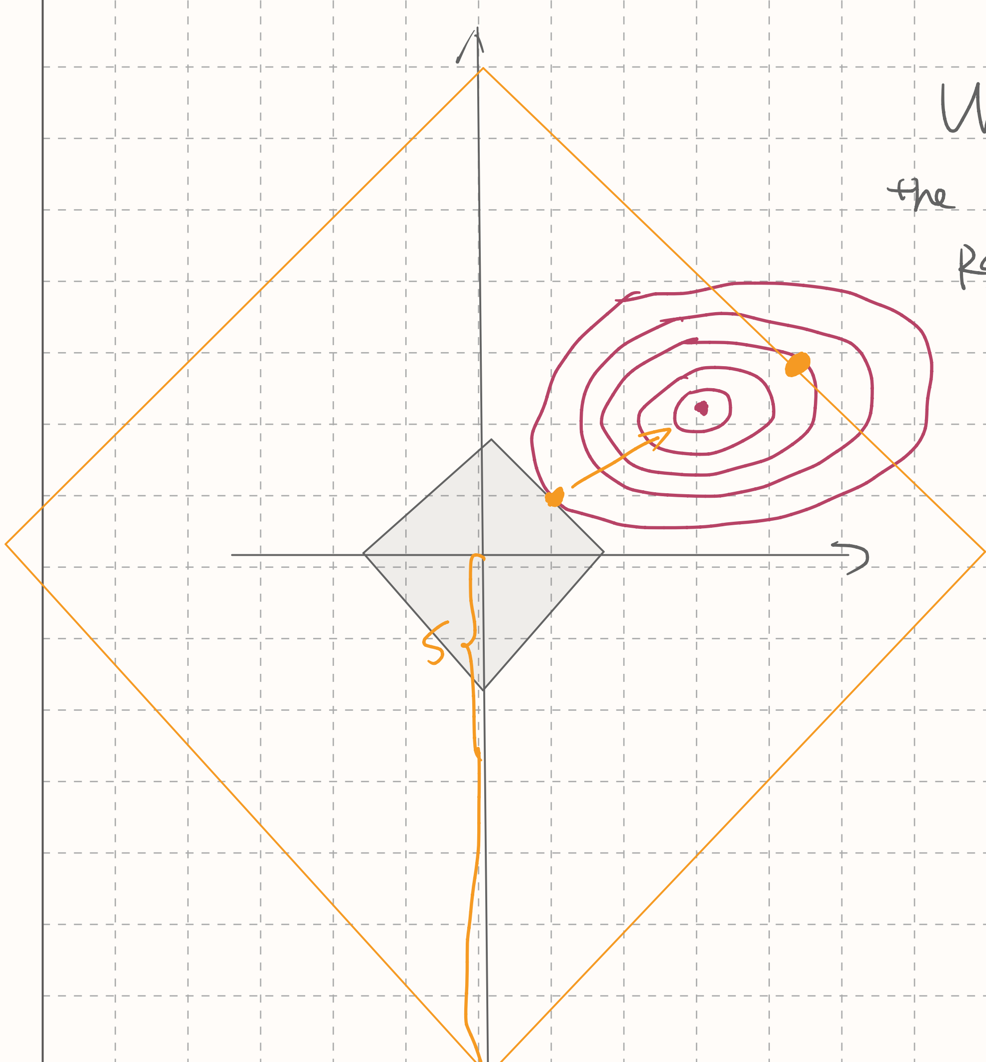
Actually I think training RSS will decrease first and when s reach some certain value, it stops decreasing.

Because as s increase, the constraints on β_j decrease, and the feasible region increases. So the optimal solution of the loss function can get closer and closer to the minimum point with no constraints. And after the feasible region includes the minimum point, it will hardly change.

Take 2 dimension as an example:



As s increase, the region increase and the intersection point becomes closer and closer to the minimum point.



Untit the feasible region includes the minimal point, the minimum traing RSS will always be the value of that point.

(b) ii

First decrease as the model get closer and closer to the real model.
Then increase because when the estimated model becomes more and more complex,
it begins to explain the irreducible errors, the variance becomes larger and larger,
so the test error increases.

(c). iii

As s increases, the constraints on the loss function decrease, the model fits better and better to the train set, bias decrease, variance increase.

(d). iv.

Same reason with (c).

(e). v.

irreducible error is the error cannot be explained by the model even the model is exactly the true model. So no matter how accurate our estimated model is, the irreducible error will stay the same as long as we are alway estimate the same model.

ISLR 6.6.5

(a).

$$\min \quad (y_1 - \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2$$

$$\text{s.t.} \quad (\hat{\beta}_1)^2 + (\hat{\beta}_2)^2 \le s$$

$\Downarrow$ Lagrange

$$\min \quad (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda\left[(\hat{\beta}_1)^2 + (\hat{\beta}_2)^2 - s\right]$$

$\Downarrow$

$$\min \quad (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda\left[(\hat{\beta}_1)^2 + (\hat{\beta}_2)^2\right]$$

(b)

Loss function of Ridge

$$f_R(\hat\beta_1,\hat\beta_2)=(y_1-\hat\beta_1x_{11}-\hat\beta_2x_{12})^2+(y_2-\hat\beta_1x_{21}-\hat\beta_2x_{22})^2+\lambda\left[(\hat\beta_1)^2+(\hat\beta_2)^2\right]$$

this is a convex problem, to calculate the optimal solution, we can find the stationary point by setting $\partial f=0$.

$$\frac{\partial f_R}{\partial\hat\beta_1}=-2x_{11}(y_1-\hat\beta_1x_{11}-\hat\beta_2x_{12})-2x_{21}(y_2-\hat\beta_1x_{21}-\hat\beta_2x_{22})+2\lambda\hat\beta_1$$

$\because$ $x_{11}=x_{12}$, $x_{21}=x_{22}$, $y_1+y_2=0$

$x_{11}+x_{21}=0$, $x_{12}+x_{22}=0$

$\therefore$ $x_{11}=x_{12}=-x_{21}=-x_{22}$, $y_1=-y_2$

$$\therefore\ \frac{\partial f_R}{\partial\hat\beta_1}=-4y_1x_{11}+4x_{11}^2\hat\beta_2+4x_{11}^2\hat\beta_1+2\lambda\hat\beta_1=0$$

$$\Downarrow$$

$$\boxed{\hat\beta_1=\frac{2y_1x_{11}-2x_{11}^2\hat\beta_2}{\lambda+2x_{11}^2}}\quad ---\ ①$$

$$\frac{\partial f_R}{\partial\hat\beta_2}=-2x_{12}(y_1-\hat\beta_1x_{11}-\hat\beta_2x_{12})-2x_{22}(y_2-\hat\beta_1x_{21}-\hat\beta_2x_{22})+2\lambda\hat\beta_2$$

plug in $x_{11}=x_{12}=-x_{21}=-x_{22}$, $y_1=-y_2$

$$\frac{\partial f_R}{\partial\hat\beta_2}=-4y_1x_{11}+4x_{11}^2\hat\beta_2+4x_{11}^2\hat\beta_1+2\lambda\hat\beta_2=0$$

$$\Downarrow$$

$$\boxed{\hat\beta_2=\frac{2y_1x_{11}-2x_{11}^2\hat\beta_1}{\lambda+2x_{11}^2}}\quad —\ ②$$

plug ① into ②

$$\hat\beta_2=\frac{2y_1x_{11}-2x_{11}^2\cdot\frac{2y_1x_{11}-2x_{11}^2\hat\beta_2}{\lambda+2x_{11}^2}}{\lambda+2x_{11}^2}$$

We can set $\frac{2y_1x_{11}}{\lambda+2x_{11}^2}=a$, $\frac{-2x_{11}^2}{\lambda+2x_{11}^2}=b$

$$\hat{\beta}_2 = a - b \cdot (a - b\hat{\beta}_2)$$

$$\hat{\beta}_2 = \frac{a(1+b)}{(1-b^2)}$$

similar, plug ② to ①.

$$\hat{\beta}_1 = a - b(a - b\hat{\beta}_1)$$

$$\hat{\beta}_1 = \frac{a(1+b)}{(1-b^2)}$$

$\therefore \hat{\beta}_1 = \hat{\beta}_2$

(c)

$$\min \quad (y_1 - \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2$$

$$\text{s.t.} \quad |\hat{\beta}_1| + |\hat{\beta}_2| \le s$$

$\Downarrow$ Lagrange

$$\min \quad (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda\left[|\hat{\beta}_1| + |\hat{\beta}_2| - s\right]$$

$\Downarrow$

$$\min \quad (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda\left(|\hat{\beta}_1| + |\hat{\beta}_2|\right)$$

(d).

plug in $y_1 = -y_2$, $x_{11} = x_{12} = -x_{21} = -x_{22}$

the optimization problem:

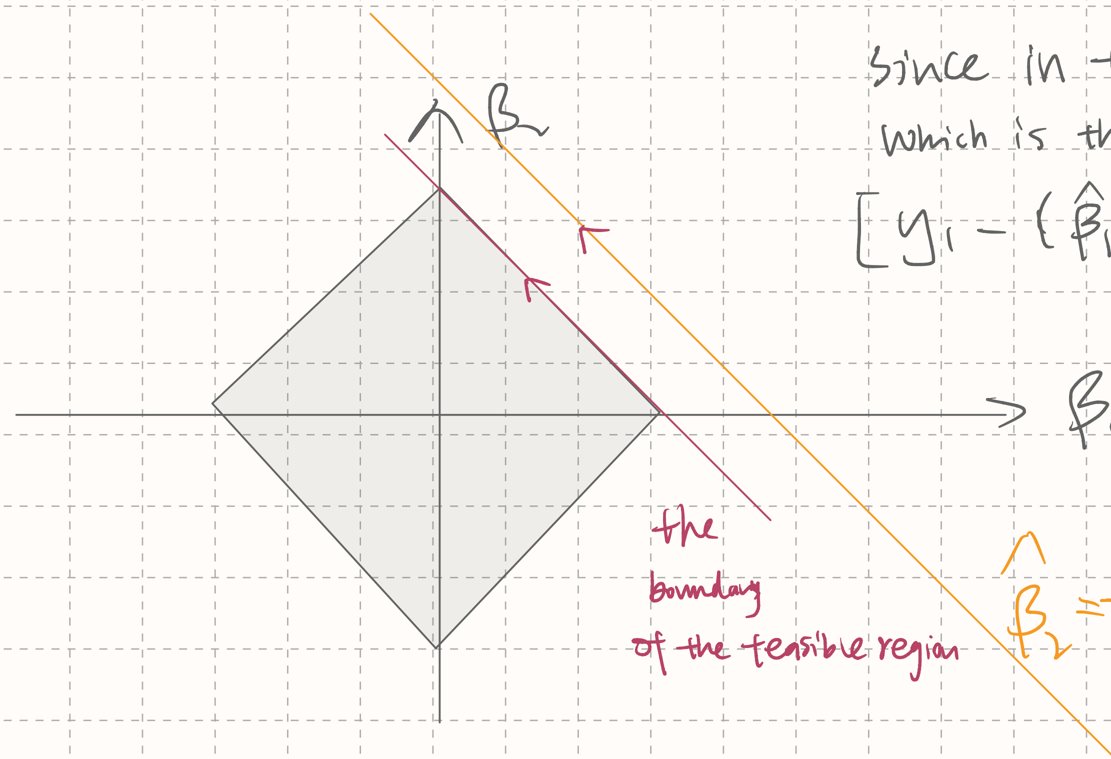
$$\min \quad [y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_{11}]^2 + [-y_1 + (\hat{\beta}_1 + \hat{\beta}_2)x_{11}]^2$$

$$\text{s.t.} \quad |\hat{\beta}_1| + |\hat{\beta}_2| \le s$$

$\Downarrow$

Yanan Zhou 4366462176 HWG — ISLR Questions

$$\min \sum \left[y_1 - (\hat\beta_1 + \hat\beta_2) x_{11} \right]^2$$
$$\text{s.t. } |\hat\beta_1| + |\hat\beta_2| \le s$$



Since in this problem, RSS can be 0, which is the minimum

$$\left[y_1 - (\hat\beta_1 + \hat\beta_2) x_{11} \right]^2 = 0$$

$$\hat\beta_2 = -\hat\beta_1 + \frac{y_1}{x_{11}}$$

The boundary of the feasible region is paralleled with the minimum solution line.

So when the feasible region intersects with the contour line of the objective function, there are infinite solutions for the optimal objective value.

So in Lasso $\hat\beta_1$ $\hat\beta_2$ are not unique.

ISLR 8.4.5

Approach 1: Red

Among the Prob of 10 bootstrapped samples, there are 6 shows $P(\text{Red}|X) > 0.5$, so according to the majority vote, the final classification is "Red".

Approach 2: Green

$$\text{Average prob} = \frac{0.1 + 0.15 + 0.2 + 0.2 + 0.55 + 0.6 + 0.6 + 0.65 + 0.7 + 0.75}{10}$$

Average prob = 0.45

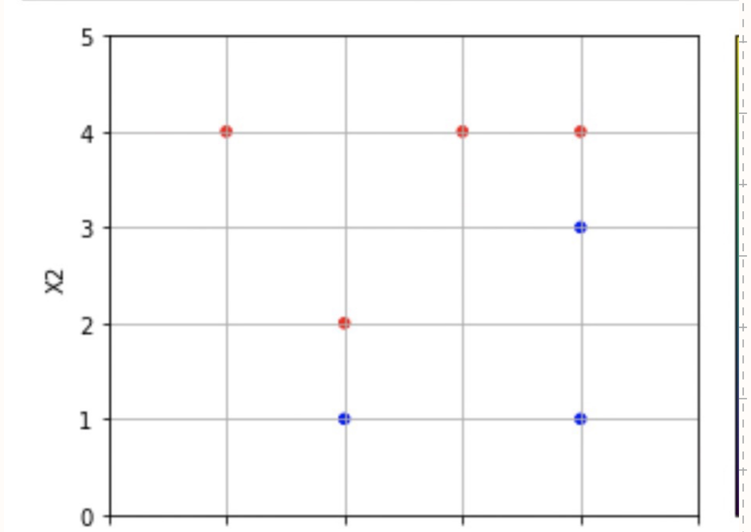
According to the average probability, $P(\text{Red} \mid X) < 0.5$

So the final classification is 'green'

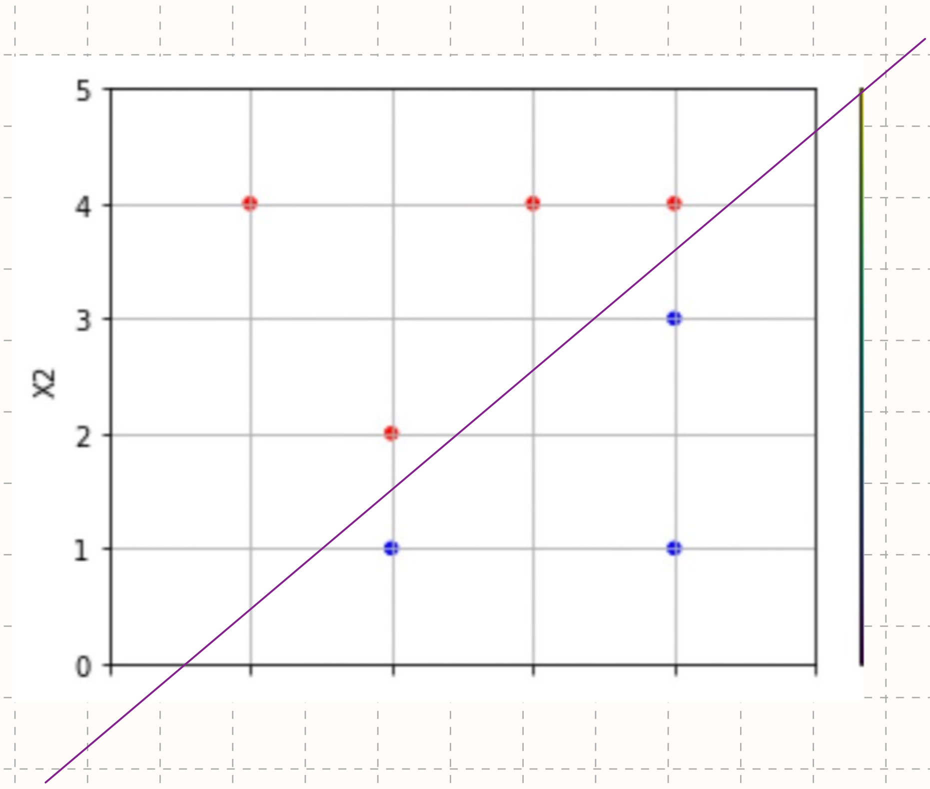
ISLR 9.7.3

(a).

```
data.plot.scatter('X1', 'X2', c = 'Y')
plt.xlim((0, 5))
plt.ylim((0, 5))
plt.xlabel('X1')
plt.ylabel('X2')
plt.grid()
```



(b)

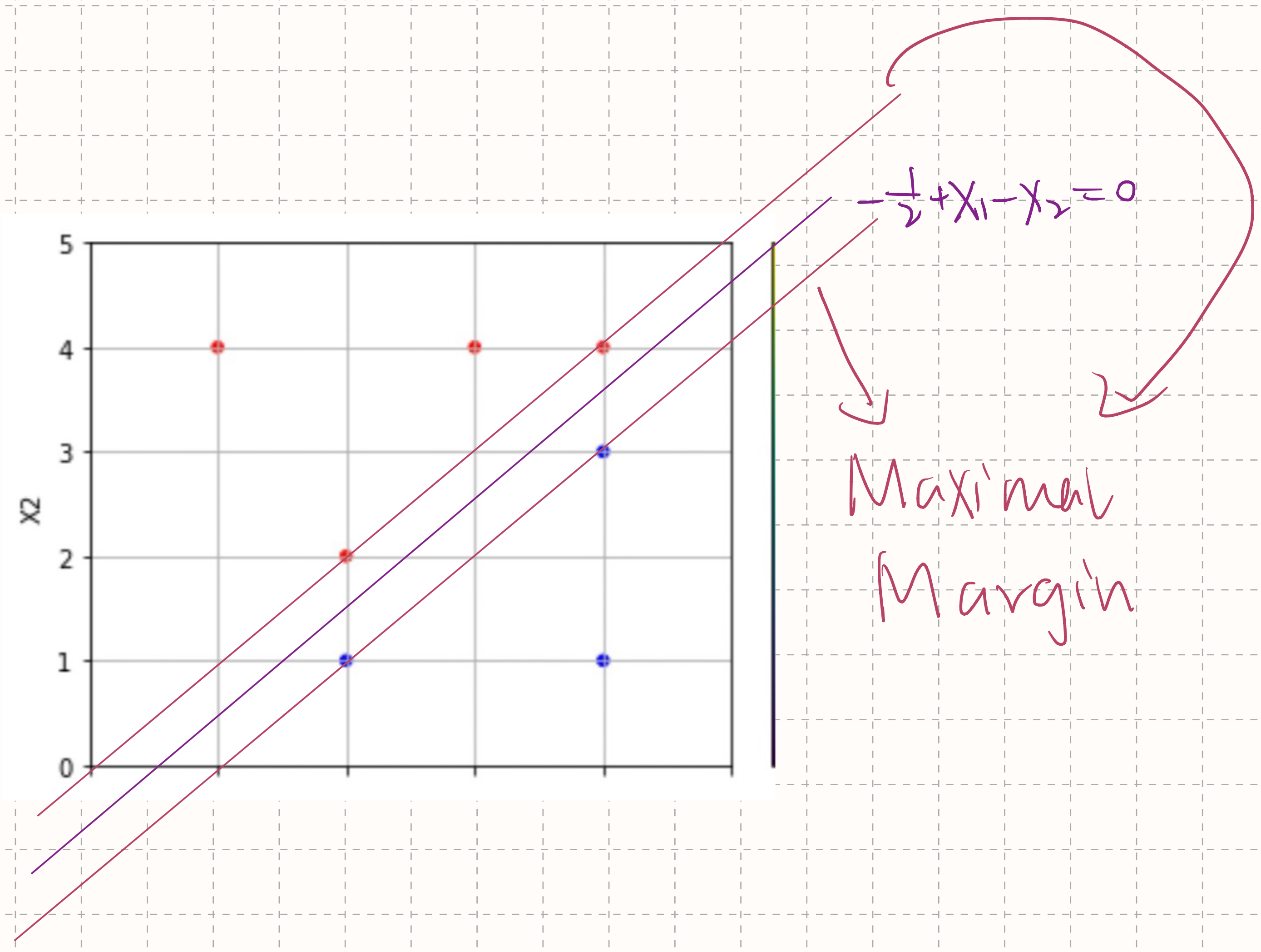


$-\frac{1}{2}+X_1-X_2=0$

Yanan Zhou 4366462176 HW6 — ISLR Questions

(c) $\begin{cases} \text{If } -\frac{1}{2}+X_1-X_2>0, & \text{then "Blue"} \\ \text{If } -\frac{1}{2}+X_1-X_2\le 0, & \text{then "Red"} \end{cases}$

(d).



(e). The support vectors are the points on or violate the margin

There are 4:

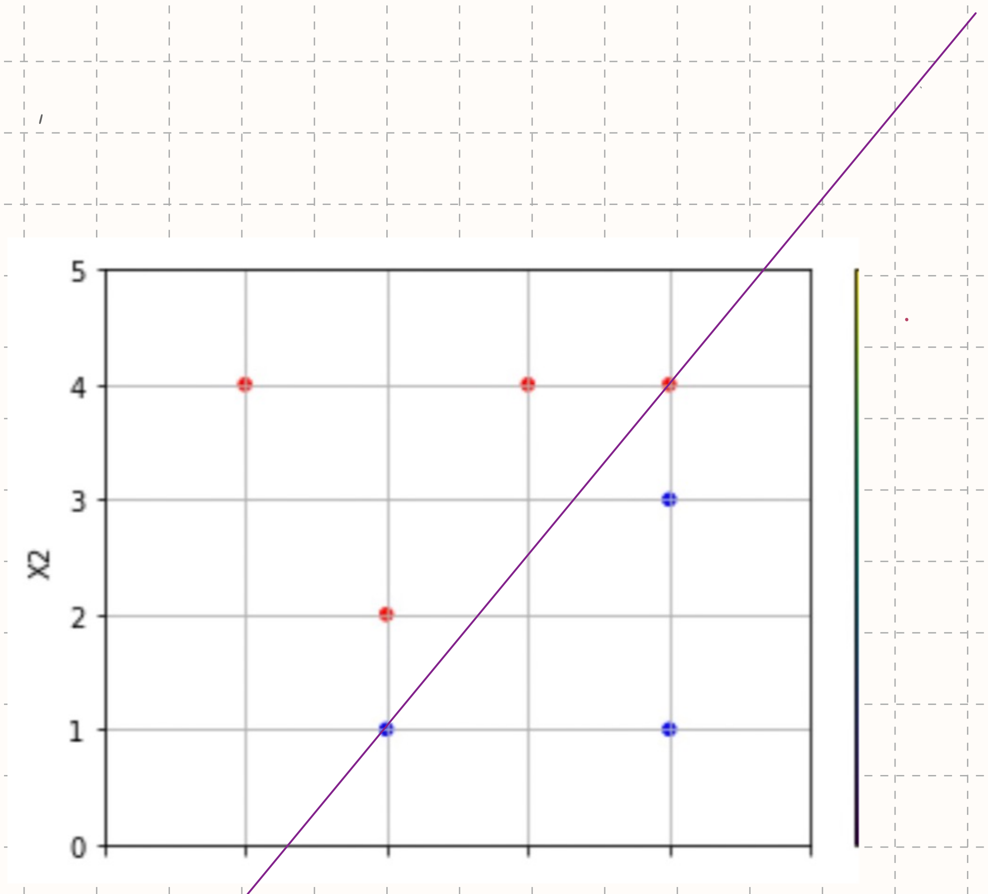
(X_1, X_2)	Y
(2, 2)	Red
(4, 4)	Red
(2, 1)	Blue
(4, 3)	Blue

(f.) Because the margin and the hyperplane are only determined by support vectors.

And the 7th point is not a support vector,

so the slight movement will not influence the maximal margin hyperplane.

(g).



the equation is $\frac{3}{2}X_1 - X_2 - 2 = 0$

(h).