ISLR 6.8.3:

minimizing function:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right) \quad \text{subject to} \quad \sum_{j=1}^{p}|\beta_j| \leq s$$

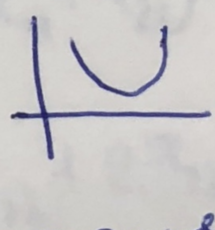
a) when 's' is "0" it means all the coefficients estimates are zero, which model depends only on intercept.

$\boxed{\text{RSS} \rightarrow \text{train RSS}}$

Here train RSS is largest.
As we increase "s", "train RSS" starts decreasing. i.e (iv) is correct (steadily decreases)

b) Test RSS starts decreasing initially. But then variance out weighs the Bias and starts increasing. i.e it has definite "U" shape

(ii) is correct



c) As 's' increases it nears to linear Regression model.
(iii) Steadily increases.

(d) As model nears to Linear Regression model ~~it~~ ~~d~~ the bias decreases.

(iv) Steadily decreases.

(e) No change (iv) Remains constant.

As this is some thing which is not dependent on any of the variables

6.8.5

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

$+ \lambda(\beta_1^2 + \beta_2^2)$

b) taking derivatives w.r.t $\hat{\beta}_1$ & $\hat{\beta}_2$ after expanding the equation.

we get

$$\frac{d}{d\hat{\beta}_1} f(x) = -2(x_{11} + x_{21}) + 2\lambda\hat{\beta}_1 = 0$$

$$\frac{d}{d\hat{\beta}_2} f(x) = -2(x_{12} + 2x_{22}) + 2\lambda\hat{\beta}_2 = 0$$

$\therefore \; -2(x_{11}+x_{21}) + 2\lambda\hat{\beta}_1 =$

$-2(x_{12}+2x_{22}) + 2\lambda\hat{\beta}_2$

given $x_{11}+x_{21} = 0$

$x_{12}+x_{22} = 0$

$2\lambda\hat{\beta}_1 = 2\lambda\hat{\beta}_2$

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

c) Lasso optimization problem.

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

$+ \lambda|\hat{\beta}_1| + \lambda|\hat{\beta}_2|$

d) taking derivatives $\frac{d}{\text{w.r.t}}$ $\hat{\beta}_1$ & $\hat{\beta}_2$ and setting them to '0'

$\frac{d}{d\beta_1} f(x) = -2(x_{11}+x_{21}) + \lambda \frac{\hat{\beta}_1}{|\hat{\beta}_1|} = 0$

$$\frac{d}{d\hat{\beta}_2} f(x) = -2(x_{12} + 2x_{22}) + \lambda \frac{\hat{\beta}_2}{|\hat{\beta}_2|} = 0$$

$$-2(x_{11} + x_{21}) + \lambda \frac{\hat{\beta}_1}{|\hat{\beta}_1|} = -2(x_{12} + 2x_{22}) + \lambda \frac{\hat{\beta}_2}{|\hat{\beta}_2|}$$

$$\frac{\lambda |\hat{\beta}_1|}{\hat{\beta}_1} = \lambda \frac{|\hat{\beta}_2|}{\hat{\beta}_2}$$

$\hat{\beta}_1$ & $\hat{\beta}_2$ are either both +ve or -ve

ignoring case '0'.

8.4.5

As the Probabilities are for "Red" and ~~the~~ assuming threshold probability as "0.5".

$$\hat{P}_{Red} = \frac{6}{10} = 0.6$$

(as 6 of them has Probs > 0.5)

$\hat{P}_{green} = 1 - 0.6 = 0.4.$

Classification result is "red".

Average Probabilities:

$$\hat{P}_{red} = \frac{0.1 + 0.15 + 0.2 + 0.2 + 0.55 + 0.6 + 0.6 + 0.65 + 0.7 + 0.75}{10}$$

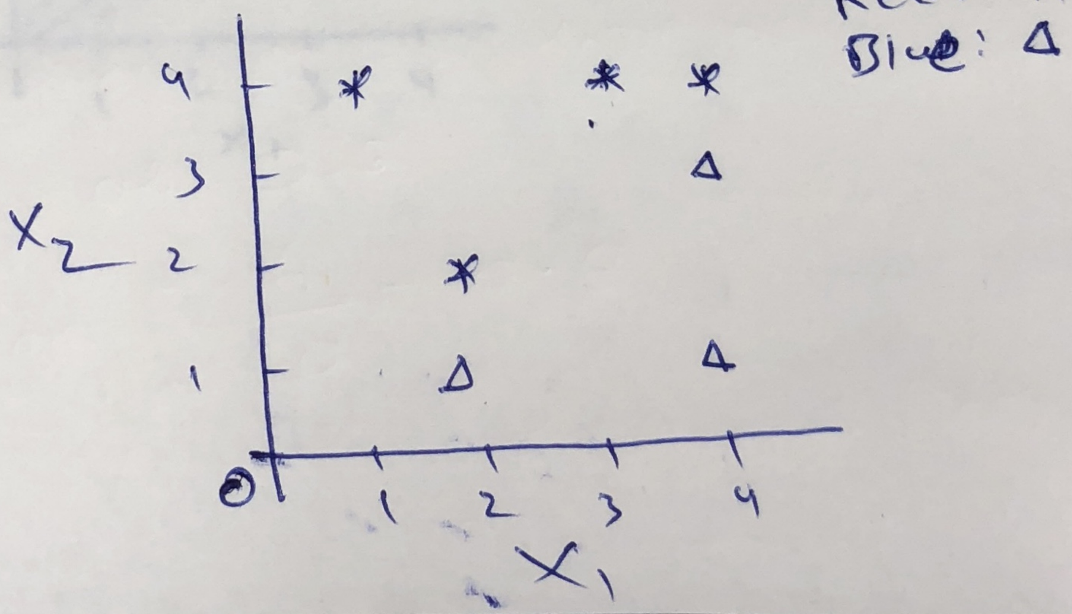
$$= 0.45.$$

$$\hat{P}_{green} = 1 - \hat{P}_{red} = 0.55$$

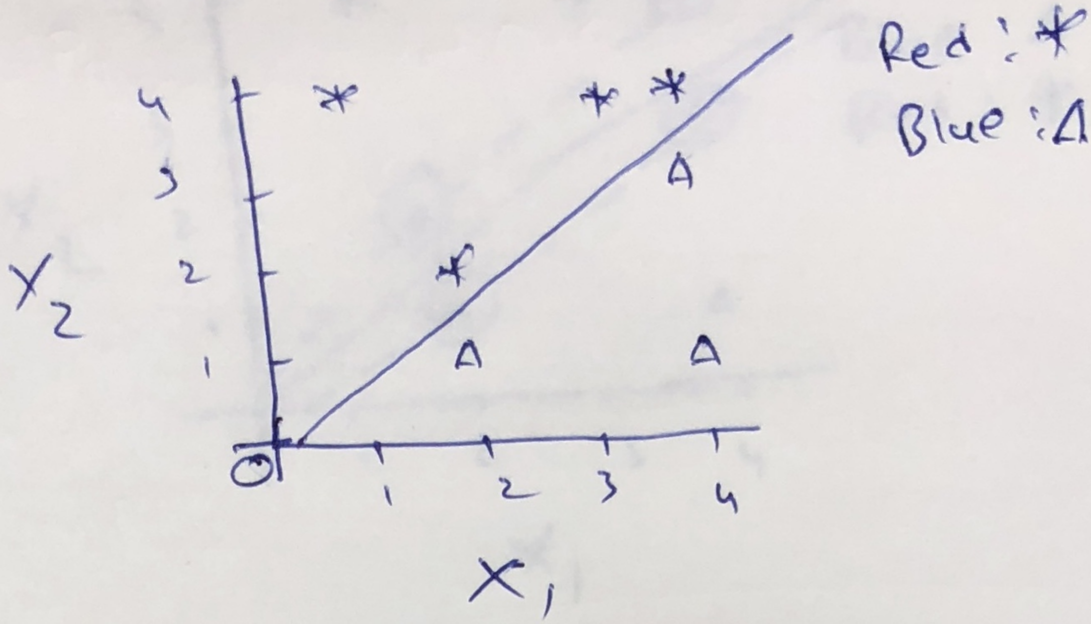
The classification result is "green". as $\hat{P}_{green}$ is large.

9.7.3

a)



b)

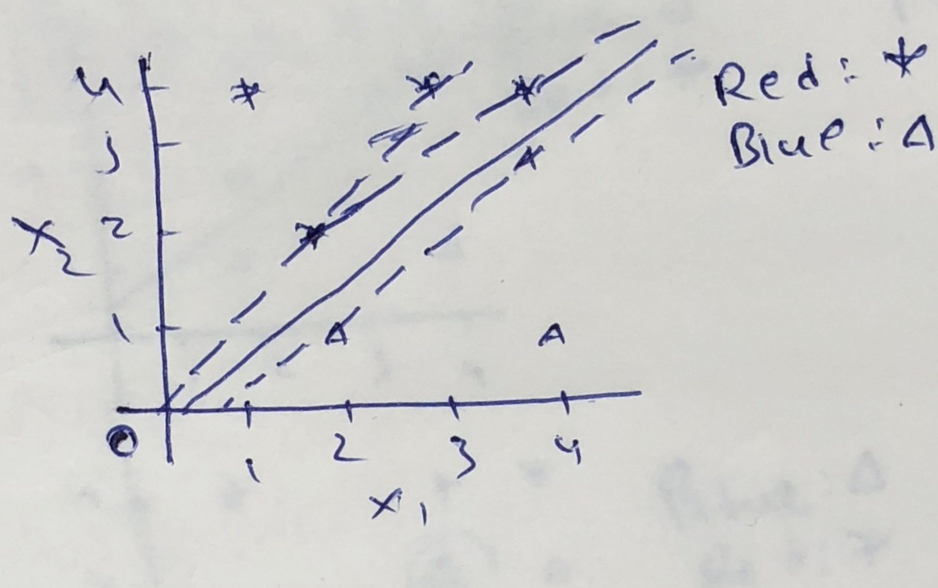


$$-0.5 + x_1 - x_2 = 0$$

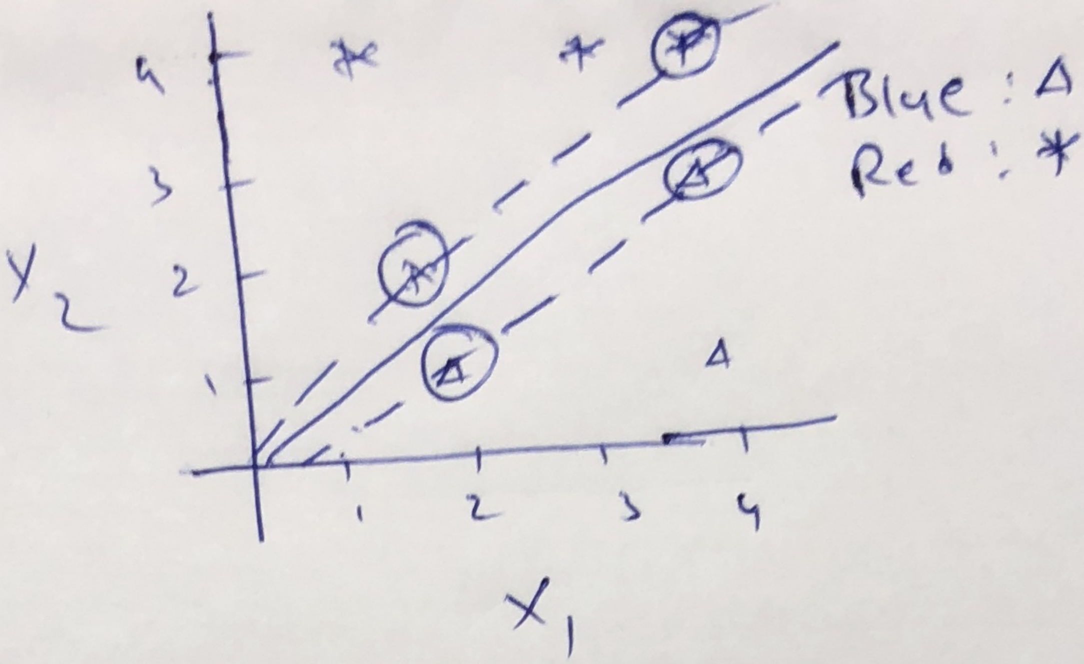
c) Red: $-0.5 + x_1 - x_2 > 0$

blue: $-0.5 + x_1 - x_2 \leq 0$

d)

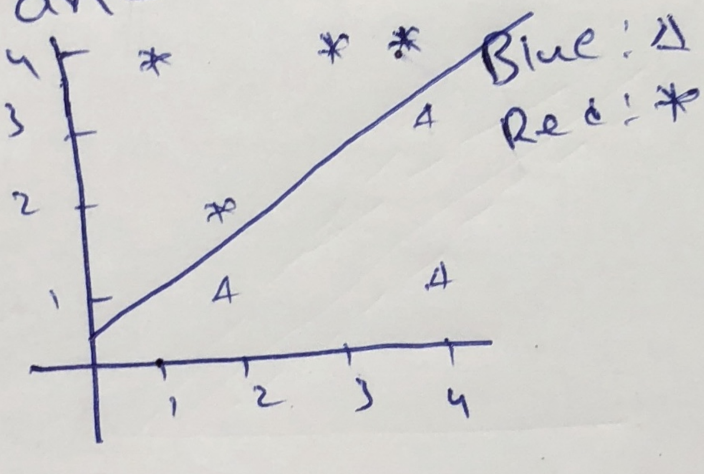


e)



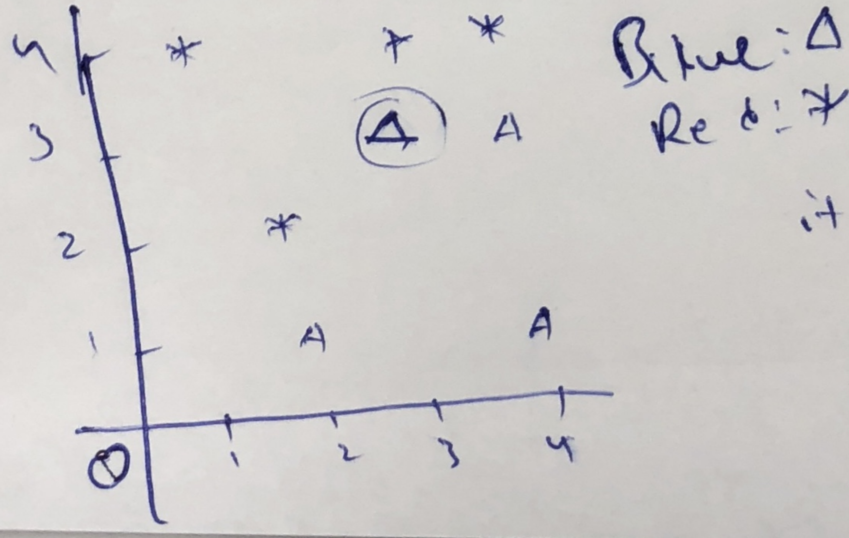
f) It would not affect because it is far from the region marked dashed and solid lines

g)



$1+4x_1-5x_2=0$

h)



it cannot classify well.