6.8.15

(4)

(a) General form of Ridge regression is

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_j)^2 - \lambda \sum_{i=1}^{p} \beta_j^2$$

$\beta_0 = 0$; $n = p = 2$

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

$x_1 = x_{12} = x_{11}$ & similarly x_2

$$(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda (\hat{\beta}_1^2 + \hat{\beta}_2^2)$$

(b) Differentiate the expression wrt to β_1 & β_2 & equating them to 0.

$$\Rightarrow \frac{\partial}{\partial \hat{\beta}_1} (2\hat{\beta}_1 x_{1}^2 - 2x_1 y_1 + 2\beta_2 x_1 x_1) + (2\hat{\beta}_1 x_2^2 - 2x_2 y_2 + 2\hat{\beta}_2 x_2^2) + 2\lambda\hat{\beta}_1 = 0$$

$$\Rightarrow \hat{\beta}_1 (x_1^2 + x_2^2) + \hat{\beta}_2 (x_1^2 + x_2^2) + \lambda \hat{\beta}_1 = x_1 y_1 + x_2 y_2$$

$\Rightarrow \lambda\hat{\beta}_1 = x_1y_1 + x_2y_2 + 2\hat{\beta}_1 x_1 x_2 + 2\hat{\beta}_2 x_1 x_2$

Repeating the same for $\hat{\beta}_2$

$$\lambda\hat{\beta}_2 = x_1y_1 + x_2y_2 + 2\hat{\beta}_2 x_1 x_2 + 2\hat{\beta}_2 x_1 x_2$$

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

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

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

(d) Similarly like part (b), replacing just the constraint term. we get

$$\frac{\partial}{\partial(\hat{\beta}_1)}(\lambda|\hat{\beta}_1|) = \frac{\lambda|\hat{\beta}_1|}{\beta_1}$$

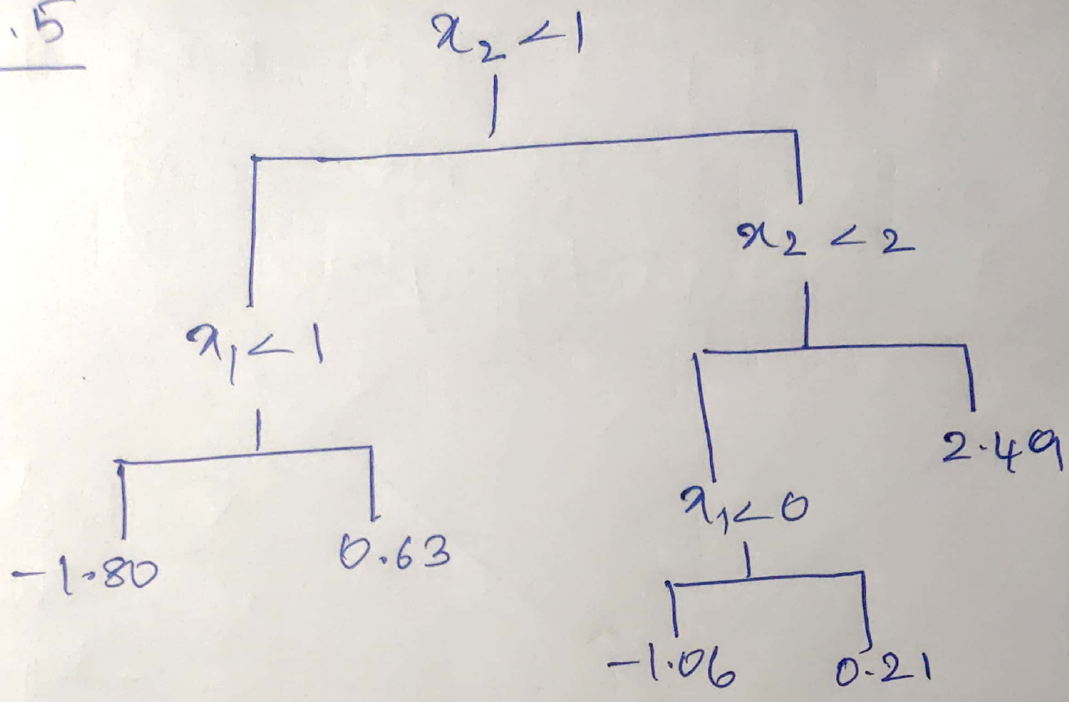
$$\frac{\partial}{\partial(\hat{\beta}_2)}(\lambda|\hat{\beta}_2|) = \frac{\lambda|\hat{\beta}_2|}{\beta_2}$$

$$\frac{\lambda|\hat{\beta}_1|}{\beta_1} = \frac{\lambda|\hat{\beta}_2|}{\beta_2}$$

Both β_1 & β_2 are both +ve or -ve

8.4.5

5

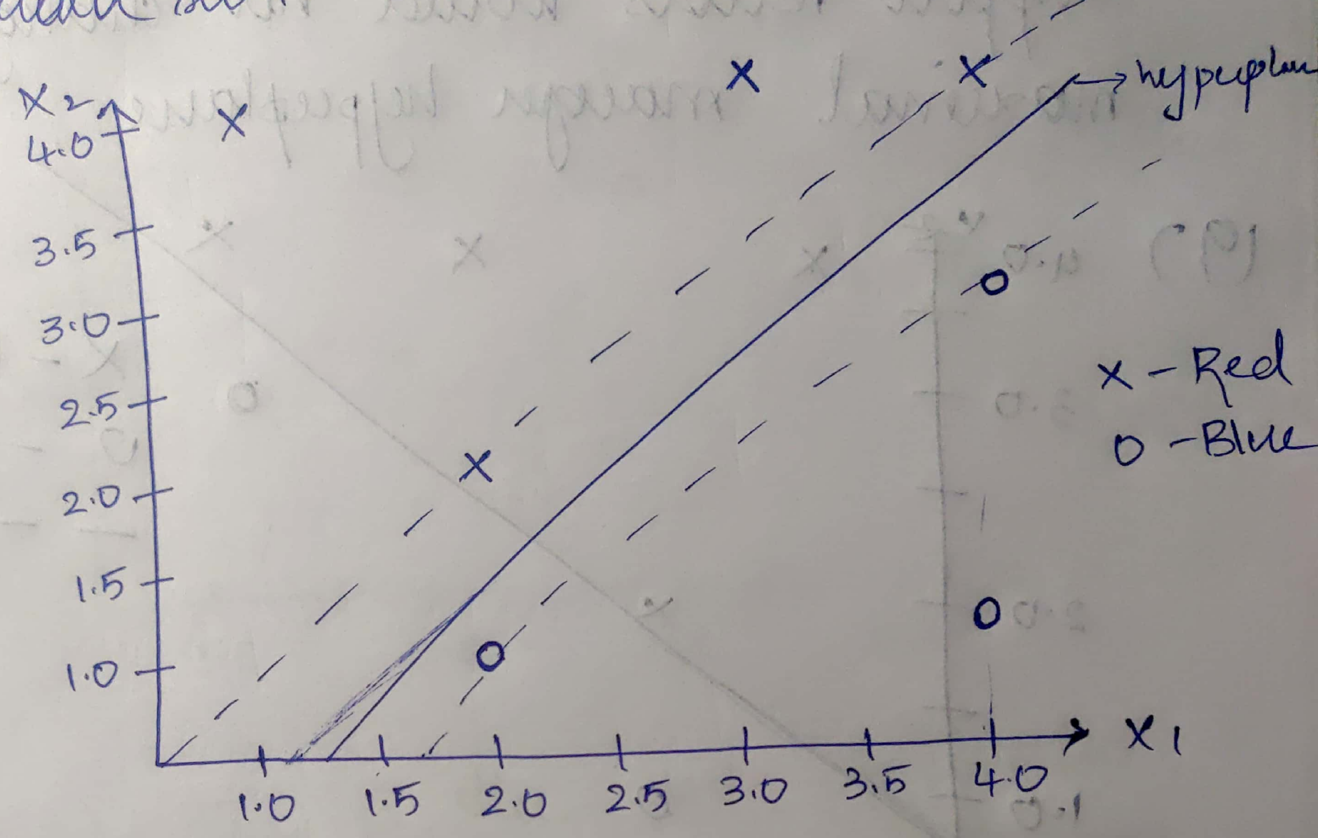


- In majority vote approach, the observation would belong to the Red class because the assignment is based on the class with the max occurrences, ~~bas~~ according to a cutoff.

- In avg approach, avg of probabilities is considered, therefore X does not belong to Red class according to this method.

9.7.3 Maximal margin classifier based on toy
(6) data set.

(a)
&
(b)
&
(d)



Equation of hyperplane ⇒ ~~$x_1 + X_2 - \frac{1}{2} = 0$~~

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

(c) Classification rules

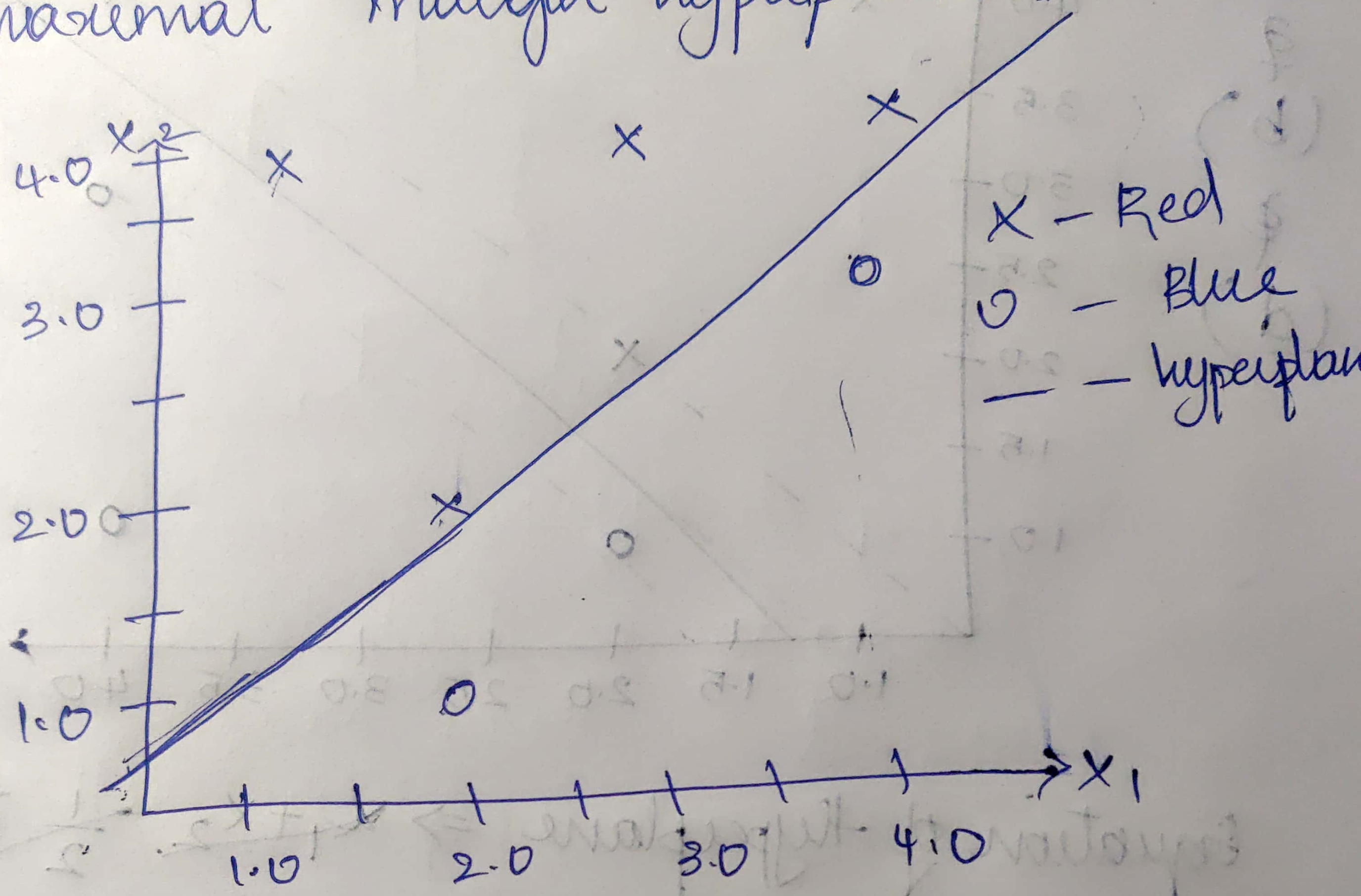
$X_1 - X_2 - \frac{1}{2} > 0 \longrightarrow$ Blue

$X_1 - X_2 - \frac{1}{2} \leq 0 \longrightarrow$ Red.

(e) Support vector for maximal margin classifier are pts (2,1) (4,3), (2,2), (4,4)

(f) Changing any point other than the support vectors would not change the maximal margin hyperplane.

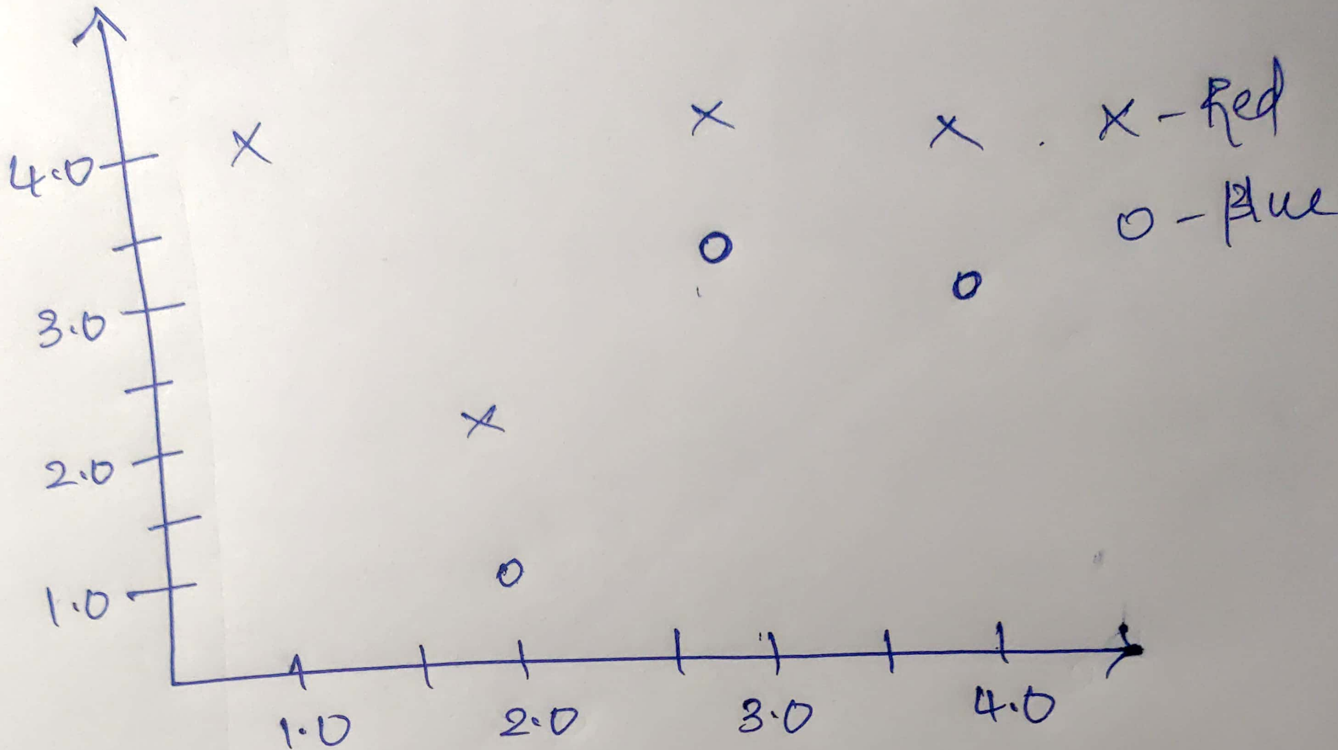
(g)



Any line b/w the (2,1), (2,2) & (4,3) (4 would be the hyperplane which is not optimal as its not passing through the midpoints of the pairs of points mentioned above.

Eg:- $X_1 - X_2 - 0.2 = 0$

h)



(3,3) – Blue is the new observation so that the 2 classes are no longer seperable by a hyperplane.