ISLR - 6.6.3

Q $\sum_{i=1} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2$ subject to $\sum_{j=1}^{p} |\beta_j| \leq s$

A) increase s from 0, RSS will - (iv)

→ as S ↑, RSS ↓ consistently till it becomes a minimum value

B) increase s for test RSS - (ii)

→ Test RSS will decrease until a point & then start increasing because it will tend to overfit the training data as S ↑

c) Repeat (a) for variance - (iii)

→ increasing s means giving more flexibility means reducing the bias and increasing variance and error

d) Repeat (a) for squared bias - (iv)

→ according to part c, bias decreases.

e) Repeat (a) for irreducible error (v)

→ Irreducible error does not dependent of s

ISLR 6.6.5

$n = 2,\ p = 2,\ x_{11} = x_{12},\ x_{21} = x_{22}$

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

$\hat{\beta}_0 = 0$

a) Ridge regression

$$\sum (y_i - \hat{\beta}_0 - \sum_{j=1}^{p=2} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{\hat{x}} \beta_j^2$$

$$= (y_1 - \hat{\beta}_0 - (\hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12}))^2 + (y_2 - \hat{\beta}_0 - (\hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22}))^2 + \lambda \hat{\beta}_1^2 + \lambda \hat{\beta}_2^2 \quad \text{--- (1)}$$

b) the ridge coeff estimates statisfy $\hat{\beta}_1 = \hat{\beta}_2$

substituting $y_1 = -y_2,\ x_{11} = -x_2,\ x_{12} = -x_{22}$

$$\rightarrow [y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12}]^2 + [-y_1 - \hat{\beta}(-x_{11}) - \hat{\beta}_2(-x_{12})]^2 + \lambda \hat{\beta}_1^2 + \lambda \hat{\beta}_2^2$$

$$= [y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12}]^2 + [y_1 - \hat{\beta}_2 x_{12}]^2 + \lambda \hat{\beta}_1 + \lambda \hat{\beta}_2$$

$$= 2[y_1^2 + \hat{\beta}_1^2 x_{11}^2 + \hat{\beta}_2^2 + x_{12}^2 - 2y_1 x_{11} \hat{\beta}_1 - 2y_1 x_{12} \hat{\beta}_2 + 2\hat{\beta}_1 \hat{\beta}_2 x_{11} x_{12}] + \lambda \hat{\beta}_1^2 + \lambda \beta_2^2$$

- by differenciating it by β_1 & β_2 &

$$\frac{\partial}{\partial \beta_1} = 2\left[2\beta_1 x_{11}^2 - 2y_1 x_{11} + 2\hat{\beta}_2 x_{11} x_{12}\right] + 2\lambda \beta_1 = 0.$$

$$\beta_1 = \frac{2y_1 x_{11} - 2\hat{\beta}_2 x_{11} x_{12}}{2x_{11}^2 + \lambda}$$

$$\hat{\beta}_1 = \frac{2y_1 x_{11} - 2\hat{\beta}_2 x_{11}^2}{2x_{11}^2 + \lambda}$$

$$\hat{\beta}_2 = \frac{2y_1 x_{12} - 2\beta_1 x_{11} x_{12}}{2x_{12}^2 + \lambda}$$

$$\hat{\beta}_2 = \frac{2y_1 x_{11} - 2\hat{\beta}_1 x_{11}^2}{2x_{11}^2 + \lambda}$$

after solving these eqn^n we get $\hat{\beta}_1 = \hat{\beta}_2$

$$\hat{\beta_1} = \frac{2y_1 x_{11}}{\lambda + 2x_{11}^2} - \frac{2x_{11}^2 \hat{\beta_2}}{\lambda + 2x_{11}^2}$$

$$\hat{\beta_1} = A - B\hat{\beta_2} \quad (1)$$

$$\hat{\beta_2} = A - B\hat{\beta_1} \quad (2)$$

$$\hat{\beta_1} - \hat{\beta_2} = -B\hat{\beta_2} + B\hat{\beta_1}$$

$$\hat{\beta_1} - \hat{\beta_2} = B(\hat{\beta_1} - \hat{\beta_2})$$

$$\boxed{B = 1}$$

$$\hat{\beta_1} = A - \hat{\beta_2} \quad \& \quad \hat{\beta_2} = A - \hat{\beta_1}$$

$$\Rightarrow \boxed{\hat{\beta_1} = \hat{\beta}_2}$$

H. P.

c) <u>Lasso</u> expression

$$\sum_{i=1}^{n}\left[y_i - \hat{\beta}_0 - \left(\sum_{j=1}^{p} \beta_j x_{ij}\right)\right]^2 + \lambda \sum_{j=1}^{p} |\hat{\beta}_j)$$

$$= (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 +$$
$$(y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 +$$
$$\lambda |\beta_1| + \lambda |\beta_2|$$

2

d) i. $\hat{\beta}_1$ & $\hat{\beta}_2$ = not unique,

Lasso eq RSS = $\sum (y_i - \beta_0 - \sum(\beta_j x_{ij}))$

subject to $\sum |\beta_j| \le s$

to minimize RSS, set it 0

$= \sum [y_i - \beta_0 - \sum \beta_j x_{ij}]^2 \ge 0$

= Square in $\beta_1, \beta_2 \rightarrow$ 2 solutions

~~$= [y_1 - \beta_0 - (\hat{\beta}_1 x_{11}) + (y_2 - \beta_0 - (\beta$~~

= for this question, we have

$2(y_1 - (\hat{\beta}_1 + \hat{\beta}_2) x_{11})^2 \ge 0$

with $|\hat{\beta}_1| + |\hat{\beta}_2| \le s$. — ①

$\frac{y_1}{x_{11}} = \hat{\beta}_1 + \hat{\beta}_2$ — ② since RSS = 0 minimum.

lasso has sharp edges, using two contsaints we find that Eq ② touches the lasso diamond at many places instead of 1

ISLR 8.4.5

Q 10 bootstrapped samples R, G, classificatn tree
10 estimates $P(R|X)$
0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, 0.75

a) use majority vote apporoach
$P(R|X) < 0.5 \rightarrow$ green
$P(R|X) \geq 0.5 \rightarrow$ Red.

Acc to this example: G, G, G, G, $\underbrace{R, R, R, R, R, R}$
majority vote for green.

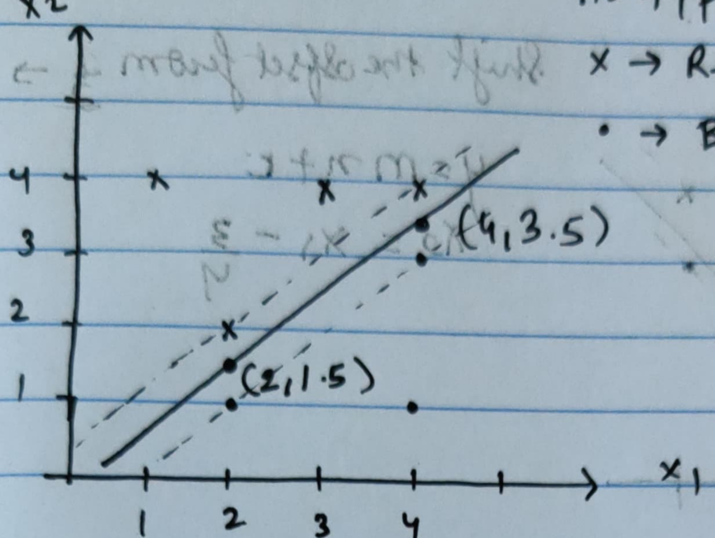
b) Avg propability

$= \frac{\text{Sum of all prob}}{10} = 0.45$

= Green.

ISLR 9.7.3

a, b, d



$n = 7,\ P = 2$

X → Red

• → Blue

$y = mx + c$

$3.5 = m * 4 + c$

$1.5 = m * 2 + c$

$2.0 = 2m$

$m = 1$

$3.5 = 1 \times 4 + c$

$c = -0.5 \Rightarrow y = x - \frac{1}{2}$

$$\boxed{x_1 - x_2 = \frac{1}{2}}$$

c) classification

$B_0 = -\frac{1}{2}$

$B_1 = 1$

$B_2 = -1$

$x_1 - x_2 - \frac{1}{2} > 0$ Red.

$x_1 - x_2 - \frac{1}{2} < 0$ Blue

e) Support vector

(2, 2) → Red

(4, 4) → Red

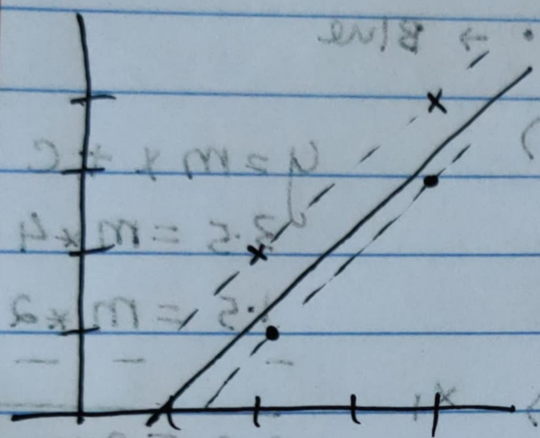
(2, 1), (4, 3) → Blue.

f) The 7th vector is not a support vector

∴ movement in (4, 1) should not affect the hyperplane

g) not optimal

Shift the offset from $\frac{1}{2} \to \frac{3}{4}$



$y = mx + c$

$x_2 = x_1 - \frac{3}{4}$

h) additional point $(2,3)$ as blue will make the classes inseparable