6.8.3 (a) iv. As $s \uparrow$, restrictions on $\beta_j \downarrow$. Model will be more flexible as s increases and therefore fit training data better. Training RSS $\downarrow$.

(b) ii. As model flexibility $\uparrow$, test RSS will decrease initially due to the model's improvement, then increase due to overfitting.

(c) iii. Model flexibility $\uparrow$ variance $\uparrow$

(d) iv. Model flexibility $\uparrow$ bias $\downarrow$

(e) v. It's "irreducible" by definition.

6.8.5 (a) $f = (y_1 - \hat\beta_1 x_{11} - \hat\beta_2 x_{12})^2 + (y_2 - \hat\beta_1 x_{21} - \hat\beta_2 x_{22})^2 + \lambda(\hat\beta_1^2 + \hat\beta_2^2)$

(b) $x_{11} = x_{12}$, $x_{21} = x_{22}$ $\Rightarrow$ let $x_1 = x_{11} = x_{12}$, $x_2 = x_{21} = x_{22}$

$\Rightarrow f = (y_1 - \hat\beta_1 x_1 - \hat\beta_2 x_1)^2 + (y_2 - \hat\beta_1 x_2 - \hat\beta_2 x_2)^2 + \lambda(\hat\beta_1^2 + \hat\beta_2^2)$

$$\nabla f(\hat\beta_1, \hat\beta_2) = \begin{bmatrix} y_1 x_1 + y_2 x_2 - \hat\beta_1(x_1^2 + x_2^2 + \lambda) - \hat\beta_2(x_1^2 + x_2^2) \\ y_1 x_1 + y_2 x_2 - \hat\beta_1(x_1^2 + x_2^2) - \hat\beta_2(x_1^2 + x_2^2 + \lambda) \end{bmatrix}$$

set $\nabla f(\hat\beta_1, \hat\beta_2) = 0 \Rightarrow \hat\beta_1 = \hat\beta_2$

(c) $f = (y_1 - \hat\beta_1 x_{11} - \hat\beta_2 x_{12})^2 + (y_2 - \hat\beta_1 x_{21} - \hat\beta_2 x_{22})^2 + \lambda(|\hat\beta_1| + |\hat\beta_2|)$

(d) let $x_1 = x_{11} = x_{12}$, $x_2 = x_{21} = x_{22}$, given $x_{11} + x_{21} = 0$, $x_{12} + x_{22} = 0$

$\Rightarrow x_1 + x_2 = 0$.

$\Rightarrow f = (y_1 - \hat\beta_1 x_1 - \hat\beta_2 x_1)^2 + (y_2 - \hat\beta_1 x_2 - \hat\beta_2 x_2)^2 + \lambda(|\hat\beta_1| + |\hat\beta_2|)$

$\Rightarrow \min (y_1 - \hat\beta_1 x_1 - \hat\beta_2 x_1)^2 + (y_2 - \hat\beta_1 x_2 - \hat\beta_2 x_2)^2$

s.t. $\lambda(|\hat\beta_1| + |\hat\beta_2|) < s$

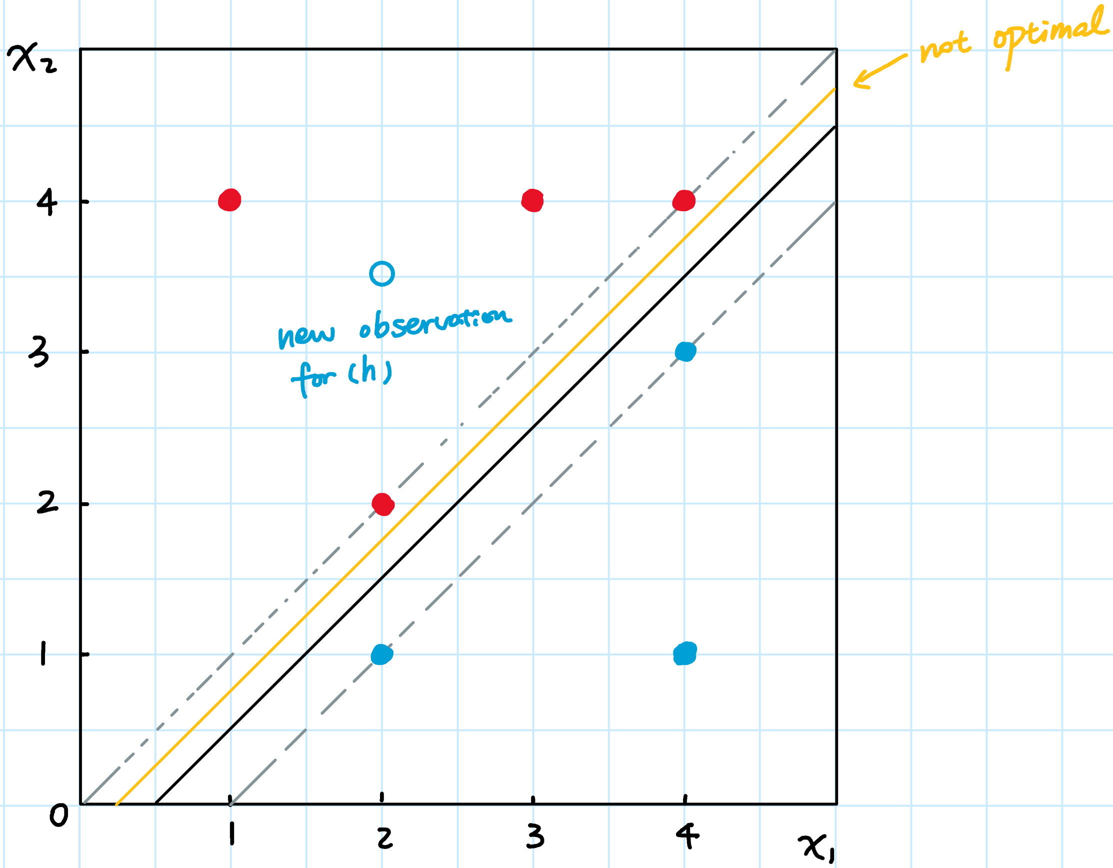
$\Rightarrow \min 2(y_1 - (\hat\beta_1 + \hat\beta_2)x_1)^2 \geq 0$

$$\Rightarrow \begin{cases} \hat\beta_1 + \hat\beta_2 = s, & \hat\beta_1 \geq 0, \ \hat\beta_2 \geq 0 \\ \hat\beta_1 + \hat\beta_2 = -s, & \hat\beta_1 \leq 0, \ \hat\beta_2 \leq 0 \end{cases}$$

8.4.5 Majority vote : Red (6 for Red vs. 4 for Green)

Average probability : Green (avg. prob = 0.45 < 0.5)

9.7.3



(c) $-x_1 + x_2 + 0.5 > 0 \Rightarrow \beta_0 = 0.5 \quad \beta_1 = -1 \quad \beta_2 = 1$

(e) support vectors : (2, 2) (4, 4) (2, 1) (4, 3)

(f) (4, 1) is not a support vector, moving it won't affect the maximal margin hyperplane

(g) $x_1 - x_2 + 0.75 = 0$