HW6

6.6.3 LASSO Regression.

a) (iv) steadily decrease.

As we $\uparrow$ s from 0, we are reducing the constraints, allowing our model to be more flexible and similar to least squares. Thus RSS will decrease & reach least squares.

b) (ii) With constraints reducing, test RSS will decrease (as fit improves) and then will start increasing due to overfitting.

c) (iii) steadily increases.

As model becomes more flexible & similar to least squares, variance will keep increasing

d) (iv) Bias steadily decreases
as the model becomes more flexible (more features)

e) (v) Unchanges
Irreducible error doesnt depend on model selection.

6.6.5

same data pt, 2 features

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

$y_1+y_2=0$, $x_{12}+x_{21}=0$. $x_{12}+x_{22}=0$

$\hat{\beta_0}=0$

(a) Minimize

$$f=\sum_{i=1}^{n}\left(y_i-\beta_0-\sum\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

Let $x_1=x_{12}=x_{11}$. $x_2=x_{21}=x_{22}$

$\beta_0=0$.

$$\therefore f=(y_1-\beta_1x_1-\beta_2x_1)^2+$$
$$(y_2-\beta_1x_2-\beta_2x_2)^2+\lambda\beta_1^2+\lambda\beta_2^2$$

$$=y_1^2-2y_1\beta_1x_1-2y_1\beta_2x_1+\beta_1^2x_1^2$$
$$+2\beta_1\beta_2x_1^2+\beta_2^2x_1^2+y_2^2$$
$$-2y_2\beta_1x_2-2y_2\beta_2x_2+\beta_1^2x_2^2$$
$$+2\beta_1\beta_2x_2^2+\beta_2^2x_2^2+\lambda\beta_1^2+\lambda\beta_2^2$$

(b) $$\frac{\partial f}{\partial\beta_1}=-2y_1x_1+2\beta_1x_1^2+2\beta_2x_{12}^2$$
$$-2y_2x_2+2\beta_1x_2^2+2\beta_2x_2+$$
$$2\lambda\beta_1=0$$

$$\frac{\partial f}{\partial\beta_2}=-2y_1x_1+2\beta_1x_1^2+2\beta_2x_1^2-2y_2x_2$$
$$+2\beta_1x_2^2+2\beta_2x_2^2+2\lambda\beta_2$$
$$=0.$$

① $\Rightarrow \lambda\beta_1 = y_1x_1 + y_2x_2 - \beta_1x_1^2 - \beta_2x_2^2$
$-\beta_2x_2^2$

② $\rightarrow \lambda\beta_2 = y_1x_1 + y_2x_2 - \beta_1x_1^2 - \beta_2x_1^2$
$-\beta_1x_2^2 - \beta_2x_2^2$

Thus $\lambda\beta_1 = \lambda\beta_2 \Rightarrow \beta_1 = \beta_2$

(c) Lasso.

Minimize

$$\sum(y_i - \beta_0 - \sum\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

$$(y_1 - \beta_1x_1 - \beta_2x_1)^2 + (y_2 - \beta_1x_2 - \beta_2x_2)^2$$
$$+\lambda|\beta_1| + \lambda|\beta_2|$$

(d) Constraint form:

$\min\left((y_1 - \hat{\beta}_1x_1 - \hat{\beta}_2x_1)^2 + (y_2 - \hat{\beta}_1x_2 - \beta_2x_2)^2\right)$
subject to $|\hat{\beta}_1| + |\hat{\beta}_2| \le s$.

Since $x_1 + x_2 = 0$ & $y_1 + y_2 = 0$
$x_2 = -x_1$ / $y_2 = -y_1$.

Minimize

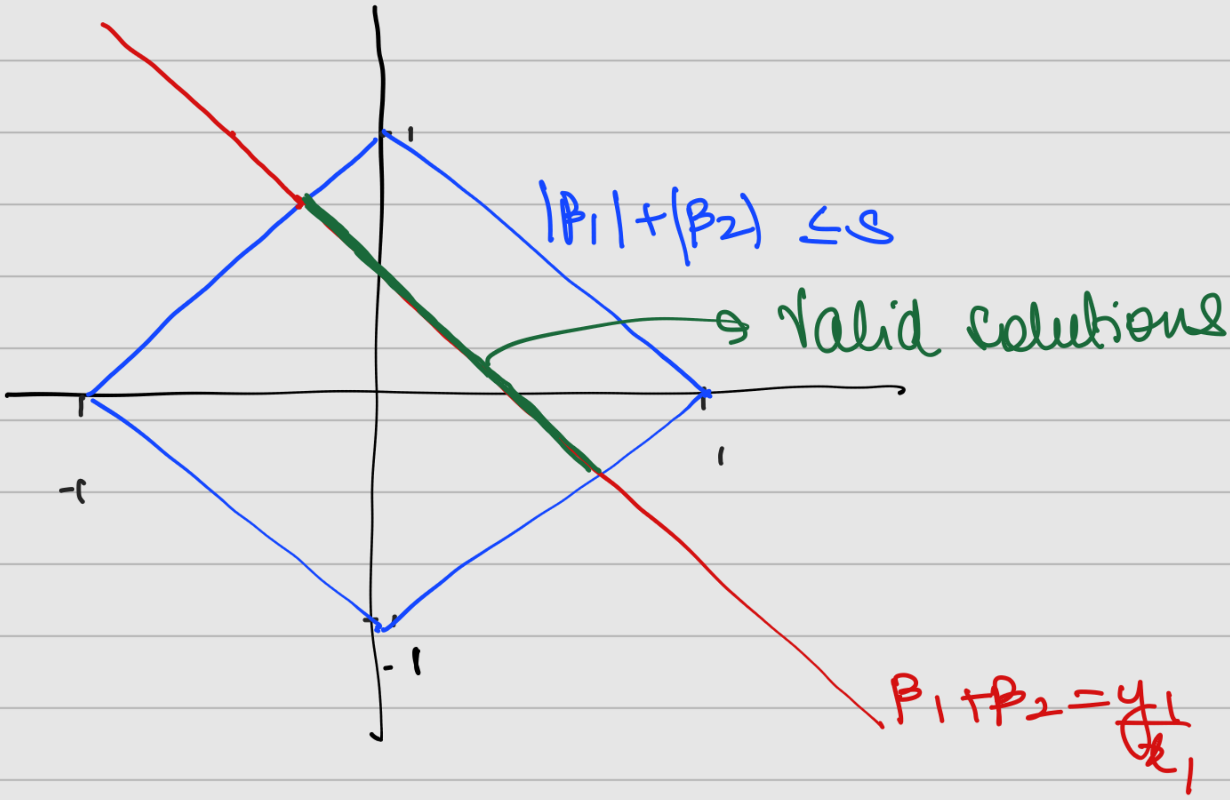
$\therefore (y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_1)^2 + (-y_1 - (\beta_1 + \beta_2)(-x_1))^2$

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

This is minimized when $\frac{y_1}{x_1} = \hat{\beta}_1 + \hat{\beta}_2$

This is a line with slope -1 in (β_1, β_2).

The constraint $|\hat{\beta_1}| + |\hat{\beta_2}| \leq s$ is a diamond centered at origin.
The line $\hat{\beta_1} + \hat{\beta_2} = \frac{y_1}{x_1}$ intersects feasible region along a segment.
$\therefore$ all points in segment are valid solutions, i.e. solution is not unique



8.4.5

(1) Majority voting for classification

Count of $P(\text{Class is Red} \mid x) < 0.5 = 4$.

Count of $P(\text{Class is Red} \mid x) > 0.5 = 6$.

$\therefore$ 'x' classified as Red

(2) Majority voting:

$P(\text{Class is Red} \mid x) =$

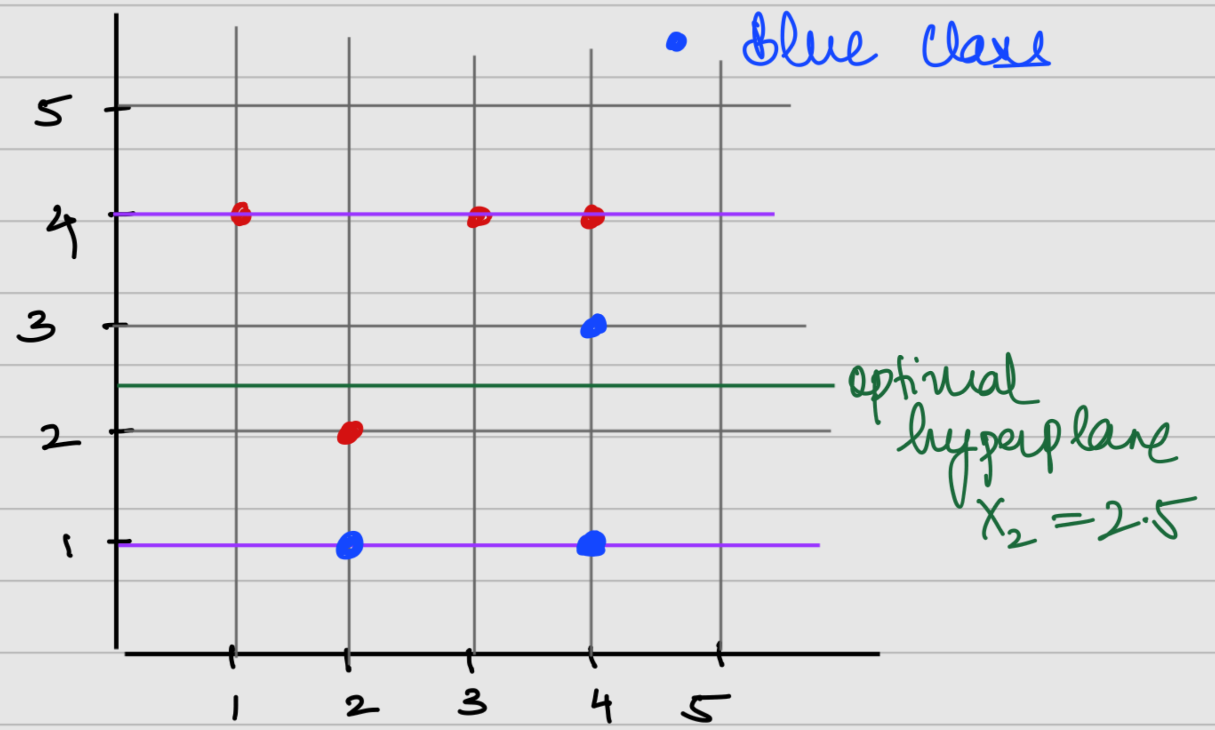
$$\frac{0.1 + 0.15 + 0.2 + 0.2 + 0.55 + 0.6 + 0.6 + 0.65 + 0.7 + 0.75}{10}$$

$$= 0.45$$

Predicⁿ: Green under avg prob. method.

9.7.3

(a)



(b) Optimal hyperplane → line that separates red and blue classes with max margin.

$$X_2 = 2.5 \rightarrow \text{hyperplane}.$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0.$$

$$0 + 0 \cdot X_1 + 1 \cdot X_2 - 2.5 = 0.$$

$$\beta_0 = -2.5, \ \beta_1 = 0, \ \beta_2 = 1$$

(c) Classification rule.

Classify to Red if $X_2 > 2.5$.
otherwise classify to blue

(d) Margin: distance from hyperplane to closest point in either class
Red $X_2 = 4$.
Blue: $X_2 = 1$

$$\text{margin} = \frac{4-1}{2} = \underline{\underline{1.5}}$$

(e) Support vectors: points that lie exactly on margin.
Red : $X_2 = 4 \rightarrow$ Obs. 1, 3, 4
Blue $X_2 = 1 \rightarrow$ Obs 5, 7

(f) Observation: 7: (4,1) is a blue point on the margin. Slight movement would not affect max. margin hypothesis unless it crosses the margin boundary.

(g) Non optimal separating hyperplane is any point that separates 2 classes but does not maximize margin.
eg. $X_2 = 3$.
$-3 + X_2 = 0$; $\beta_1 = 0$, $\beta_2 = 1$.
still separates red & blue but not with largest margin.