TANMAY GUPTA
Eg 219

Math-680
Homework-4

1.

LASSO Problem

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij} \beta_j)^2, \text{ subject to } \sum_{j=1}^{p} |\beta_j| \le t$$

where $t \geq 0$ is constant

(a) if $t = 0$

$\hat{\beta}_{j\,Lasso} \Rightarrow \beta_{Lasso} = \beta_{OLS} + \Lambda(\sum|\beta_j|)$

so when $t = 0$

that means $\Lambda \sum|\beta_i| = 0$

$$\boxed{\text{So } \hat{\beta}_{j\,LASSO} = \beta_{OLS}}$$

(b) $t_0 = \sum_{j=1}^{p} |\hat{\beta}_j^{\,OLS}|$

To prove $\Rightarrow$ if $t \geq t_0$ then $\hat{\beta}_j$ Lasso $= \hat{\beta}_j$ ols

There are 2 ways to prove this. First $\rightarrow$

(1) if $t \geq t_0$ Λ corresponds to 0 or $\lambda = 0$

$$\sum_{i=0} (y_i - \sum x_i \beta)^2 + 0 . \sum|\beta|$$

which is equivalent to $\beta_{j\,OLS}$.

So $\underline{\underline{\beta_j^{\text{lasso}} = \beta_j^{\text{ols}}}}$

(ii) Second $\to$

$\min RSS = \hat{\beta}_{\text{lasso}}$

subject to $\Sigma|\beta_j| \le t$

(from part a)

$\Rightarrow RSS(\hat{\beta}^{LS}) \le RSS(\hat{\beta}^{\text{lasso}})$

As $\hat{\beta}^{\text{lasso}}$ is optimal solution

So $\hat{\beta}_j^{\text{lasso}} = \hat{\beta}_j^{\text{ols}}$

2 Assumption $\Rightarrow X^T X = I$

To prove

$$\min \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \sum_{j=1}^{p} J(\beta_j)$$

$$= \min_{\beta_j} (\beta_j^{\text{ols}} - \beta_j)^2 + J(\beta_j), \text{ for each } j = 1, p)$$

Where J is the penalty function.

Proof →

As $X^T X = I$ So orthogonal matrix

$$(y_i - x^T\beta)^2 \Rightarrow (y_i - x^T\beta_{LS} + x^T\beta_{LS} - x^T\beta)^2$$

$$\Rightarrow (y_i - x^T\beta_{LS})^2 + (x^T\beta_{LS} - x^T\beta)^2$$

As identity matrix, so the cross product term = 0.

$$2\underbrace{(y_i - x^T\beta_{LS})^T}_{\text{residual term}}(x^T\beta_{LS} - x^T\beta) = 0$$

So $\min \sum (y_i - x^T\beta_{LS})^2 + \sum J(\beta_j)$

$$\Rightarrow \min (x^T\hat{\beta}_a - x^T\beta_{LS})^2 + \sum_j(\beta_j)$$

$$\Rightarrow \min \cancel{(x^T\beta - x^T\beta} \; X^TX(\hat{\beta} - \hat{\beta}_{LS})^T(\hat{\beta}_{LS} - \beta) + J(\beta_j)$$

$$\Rightarrow \min (\hat{\beta}_{jLS} - \beta_j)^T(\hat{\beta}_{jLS} - \beta_j) + J(\beta_j)$$

$$= \min (\hat{\beta}_{jLS} - \beta_j)^2 + J(\beta_j)$$

($\hat{\beta}_{LS}$ & $\hat{\beta}_{OLS}$ is same)

Hence proved

3. $Y = \sum_{j=1}^{4} x_j \beta_j + \varepsilon$

X is orthogonal matrix

$\hat{\beta}_{ols} = (1.1, -0.8, 0.3, -0.1)$

(a) β_{ridge} for $\lambda = 1$ & $\lambda = 0.4$.

$\hat{\beta}_{ridge} = \dfrac{\beta_{ols}}{1+\lambda}$

for $\lambda = 1$, $\hat{\beta}_{ridge} \Rightarrow$

$\Rightarrow \left(\dfrac{1.1}{1+1}, \dfrac{-0.8}{1+1}, \dfrac{0.3}{1+1}, \dfrac{-0.1}{1+1} \right)$

$\Rightarrow \left(\dfrac{1.1}{2}, \dfrac{-0.8}{2}, \dfrac{0.3}{2}, \dfrac{-0.1}{2} \right)$

$\Rightarrow (0.55, -0.4, 0.15, -0.05)$.

for $\lambda = 0.4$, $\hat{\beta}_{ridge} \Rightarrow$

$$\Rightarrow \left(\frac{1.1}{1+0.4}, \frac{-0.8}{1+0.4}, \frac{0.3}{1+0.4}, \frac{-0.1}{1+0.4} \right)$$

$$\Rightarrow \left(0.78, -0.57, 0.21, -0.07 \right).$$

As λ goes near 0 the $\hat{\beta}$ value increases.

(d) LASSO

$$\beta_{j\,Lasso} = \operatorname{sign}(\hat{\beta}_j^{ls})\left(|\hat{\beta}_j^{ls}| - \frac{\lambda}{2}\right)_+$$

$$= \begin{cases} \hat{\beta}_j^{ls} - \frac{\lambda}{2} & \text{if } \hat{\beta}_j^{ls} > \frac{\lambda}{2} \\ 0 & \text{if } |\hat{\beta}_j^{ls}| \le \frac{\lambda}{2} \\ \hat{\beta}_j^{ls} + \frac{\lambda}{2} & \text{if } \hat{\beta}_j^{ls} < -\frac{\lambda}{2} \end{cases}$$

if $\lambda = 1$ then $\frac{\lambda}{2} = 0.5$.

$$\beta_j\ lasso = \left(1.1 > 0.5, -0.8 < 0.5, 0.3 < 0.5, -0.1 < 0.5\right)$$

$$\Rightarrow \left(1.1 - 0.5, -0.8 + 0.5, \overset{0}{\cancel{0.3 + 0.5}}, -\overset{0}{\cancel{0.1 + 0.5}}\right).$$

$$\Rightarrow \left(0.6, -0.3, -\cancel{0.2}, 0\right).$$

if $\lambda = 0.4$ $\quad \lambda/2 = 0.2$

$\beta_{LASSO} = \left(1.1 > 0.2,\ -0.8 < -0.2,\ 0.3 > 0.2,\ \underset{\le 0.2}{|-0.1|}\right)$

$\Rightarrow (1.1 - 0.2,\ -0.8 + 0.2,\ 0.3 - 0.2,\ 0)$

$\Rightarrow (0.9,\ -0.6,\ 0.1,\ 0)$

(c)

for ridge regression as λ decreases from 1 to 0.4 the β's go away from zero

In LASSO if the β_{OLS} is near 0 it truncates the small coefficients and bring it to zero. As λ dec from 1 to 0.4 the β ~~in~~ go away from zero (only the ones that are sig enough in β_{OLS})

4.

(a)

9. (a) Linear reg. model in the R file.

$R^2 = 0.69$ or 69%

p-values of reg. coeff ->

coeff.	p-value
lcavol	1.47
lweight	0.007
age	0.16
lbph	0.04
svi	0.01
lcp	0.067
gleason	0.88
pgg45	0.08

The significant parameters at $\alpha = 0.05$ are lweight, lbph & svi because their p-value are below 0.05.

Train error => 0.439

Test error => 0.521

(b)

Model in the R file giving coefficients please refer to R file. Training error =>

~~Optimal BIC = -27.00272~~

Important variables BIC =>

(Intercept)	lcavol	weight
-1.0434396	0.6274074	0.7383757

Test error given model = 0.792

Training error => (of 8 models)

=> 0.664 0.553 0.521 0.489 0.478

0.455 0.4393 0.439

[For regression coefficient please refer to R file as it was hard to write]

Optimal BIC is on index 2 which is => -27.00272

Important variables selected by BIC ->

(Intercept)	lcavol	lweight
-1.0494	0.6276	0.7383

Refitted the model with important var & the test error reduced to 0.492

(C)

Important var selected by AIC.

(Intercept)	lcavol	lweight	age
0.2590	0.5739	0.6192	-0.019

lbph	svi	lcp	pgg45
0.1444	0.7417	-0.2854	0.0089

Scanned by CamScanner

Refitting the model with the import. var. selected by AIC & the test error for that is 0.516

5.

(a) Model in the R file.

Best λ = 0.7777778

Estimates of coeff. ->

lcavol	lweight	age	lbph	svi
0.5191	0.5764	-0.0113	0.1267	0.6095

lcp	gleason	pgg45
-0.0950	0.0000	0.0060

The test error using this CV method is

0. 477

(5)

Best λ = 0.333

estimated reg. coeff. =>

lcavol	lweight	age	lbph	svi
0.4356	0.3146	0.000	0.0000	0.1480

lcp	gleason	pgg45
0.0000	0.0000	0.0000

Test mse using one-SD CV

0.510