علیرضا درویشی - 96109674 - تمرین 8 - تحلیل رگرسیون

تمرینات نظری :

1-

a) مدل best subset کمترین error روی training را دارد چون روی مجموعه‌ی بزرگتری از مدل‌ها بهینه گرفته می‌شود.

b) قابل پیش‌بینی نیست. اگر تعداد observation ها زیاد باشد، احتمالاً test error برای bestsub کمتر از روش‌های دیگر است.

c)

i. درست. مدل k+1، همان k متغیر پیش‌بینی شده‌ی مدل k را دارد و یک متغیر به آن اضافه شده است.

ii. درست. مدل k، یکی از متغیرهای پیش‌بینی شده در مدل k+1 را حذف کرده است.

iii. غلط. دلیلی ندارد که گزاره‌ی گفته شده صحیح باشد. هرچند می‌شود به صورت اتفاقی صحیح باشد.

iv. غلط. دلیلی ندارد که گزاره‌ی گفته شده صحیح باشد. هرچند امکان دارد به صورت اتفاقی صحیح باشد.

v. غلط. مانند بخش قبلی

~~این روش‌ها بایاس بیشتری نسبت به least square ها دارند پیش‌بینی flexible است.~~
~~اما ... همه درست است و ... در این ... است.~~

2.

a) iii. روش lasso، بایاس بیشتری نسبت به least square دارد.

b) iii. مشابه lasso، بایاس بیشتری دارد.

c) ii. روش‌های غیرخطی، Variance بیشتری دارند و زمانی بهبود می‌دهند که افزایش واریانس کمتر از کاهش بایاس باشد.

3.

a) iv. چون Variance مدل زیاد می‌شود.

b) ii. اول تست ارور کم می‌شود. بعد از گذشتن از نقطه‌ی بهینه، زیاد می‌شود.

c) واریانس در کل بازه زیاد می‌شود. (iii)

d) بایاس کم می‌شود. (iv)

e) v.

4.

a) i. چون Variance مدل کم می‌شود

b) iii

c) iv

d) ii

e) ii

-5

(a

$$Y = \begin{pmatrix} y \\ -y \end{pmatrix}, \quad X = \begin{pmatrix} x & x \\ -x & -x \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\min \ (Y - X\beta)^T (Y - X\beta) + \lambda \beta^T \beta = f(\beta)$$

(b

$$\nabla_\beta f(\beta) = -2X^T Y + 2X^T X \beta + 2\lambda\beta = 0 \Rightarrow (X^T X + \lambda I)\beta = X^T Y$$

$$\Rightarrow \beta = (X^T X + \lambda I)^{-1} X^T Y \longrightarrow X^T X = \begin{pmatrix} x & -x \\ x & -x \end{pmatrix}\begin{pmatrix} x & x \\ -x & -x \end{pmatrix} = \begin{pmatrix} 2x^2 & 2x^2 \\ 2x^2 & 2x^2 \end{pmatrix}$$

$$(X^T X + \lambda I)^{-1} = \begin{pmatrix} 2x^2+\lambda & 2x^2 \\ 2x^2 & 2x^2+\lambda \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2x^2+\lambda}{(2x^2+\lambda)^2 - 4x^4} & \frac{-2x^2}{(2x^2+\lambda)^2 - 4x^4} \\ \frac{-x^2}{(2x^2+\lambda)^2 - 4x^4} & \frac{2x^2+\lambda}{(2x^2+\lambda)^2 - 4x^4} \end{pmatrix}$$

$$\Rightarrow \beta = \frac{1}{\lambda^2 + 4\lambda x^2} \begin{pmatrix} 2x^2+\lambda & -2x^2 \\ -2x^2 & 2x^2+\lambda \end{pmatrix} \underbrace{\begin{pmatrix} x & -x \\ x & -x \end{pmatrix}\begin{pmatrix} y \\ -y \end{pmatrix}}_{\begin{pmatrix} 2xy \\ 2xy \end{pmatrix}} = \boxed{\frac{2xy\lambda}{\lambda^2 + 4\lambda x^2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

(c

$$\min \ (Y - X\beta)^T (Y - X\beta) + \lambda \sum_i |\beta_i| = f(\beta)$$

(d

$$\nabla_\beta f(\beta) = -2X^T Y + 2X^T X\beta + \lambda\,\mathrm{sign}(\beta) = 0$$

$$\Rightarrow (X^T X)\beta = X^T Y - \frac{\lambda}{2}\mathrm{sign}(\beta), \quad X^T X = 2x^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \ \mathrm{rank}(X^T X) = 1$$

$$\Rightarrow \mathrm{sign}(\beta) = \pm \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \boxed{\beta_1, \beta_2 \geq 0, \quad 2x^2(\beta_1 + \beta_2) = 2xy - \frac{\lambda}{2}\mathrm{sign}(\beta_1)}$$