

HW-3

I

1.

$$OLS: w_{OLS} = (X^T X)^{-1} X^T y$$

$$y = \begin{bmatrix} 3.5 \\ 1.0 \\ 3.8 \\ 10.1 \\ 8.5 \end{bmatrix} \quad X = \begin{bmatrix} 1 & \phi_1 \\ 1 & \phi_2 \\ 1 & \phi_3 \\ 1 & \phi_4 \\ 1 & \phi_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 6 \\ 1 & 18 \\ 1 & 8 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 5 & 37 \\ 37 & 441 \end{bmatrix} \quad X^T y = \begin{bmatrix} 26.9 \\ 287.6 \end{bmatrix}$$

$$\det(X^T X) = 836 \quad (X^T X)^{-1} = \frac{1}{836} \begin{bmatrix} 441 & -37 \\ -37 & 5 \end{bmatrix}$$

$$w_{OLS} = \frac{1}{836} \begin{bmatrix} 441 & -37 \\ -37 & 5 \end{bmatrix} \begin{bmatrix} 26.9 \\ 287.6 \end{bmatrix} \\ = \begin{bmatrix} 1.46136 \\ 0.52954 \end{bmatrix}$$

$$\hat{y}_{OLS} = 1.46136 + 0.52954(y_1, y_2)$$

$$2. \quad \Lambda = \begin{bmatrix} 0 & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\det[(X^T X + \Lambda)^{-1}] = 841$$

$$\text{Ridge: } w_{\text{ridge}} = (X^T X + \Lambda)^{-1} X^T y$$

$$w_{\text{ridge}} = (X^T X + \Lambda)^{-1} X^T y = \frac{1}{841} \begin{bmatrix} 441 & -37 \\ -37 & 5 \end{bmatrix} \begin{bmatrix} 26.7 \\ 287.6 \end{bmatrix}$$

$$= \begin{bmatrix} 1,48466 \\ 0,526391 \end{bmatrix}$$

$$\hat{y}_{\text{ridge}} = 1,48466 + 0,526391(y_1)$$

Comparison: With ridge the coefficients shrink due to regularization (large coefficients get pulled towards zero).

3.

Train

$$\bullet \text{ OLS: } \hat{y}_{OLS} = 1,46136 + 0,52954(y_1, y_2)$$

$$\bullet \text{ Ridge: } \hat{y}_{Ridge} = 1,48466 + 0,526391(y_1, y_2)$$

$$\text{Error} = y_{\text{mean}} - \hat{y}_{\text{product}}$$

$x$	$y_1, y_2$	$\hat{y}_{OLS}$	$\hat{y}_{Ridge}$	$ \text{Error} _{OLS}$	$ \text{Error} _{Ridge}$
$x_1$	4	3,57952	3,590248	0,07952	0,09025
$x_2$	7	1,99090	2,011057	0,99090	1,011057
$x_3$	6	4,63860	4,643042	0,83860	0,84304
$x_4$	18	10,99308	10,95981	0,89308	0,85981
$x_5$	8	5,69768	5,69584	2,80232	2,80416

$$MAE_{\text{train, OLS}} = \frac{0,07952 + 0,99090 + 0,83860 + 0,89308 + 2,80232}{5} \approx 1,12088$$

$$MAE_{\text{train, ridge}} = \frac{0,09025 + 1,011057 + 0,84304 + 0,85981 + 2,80416}{5} \approx 1,12066$$

Test

$x$	$y_1, y_2$	$\hat{y}_{OLS}$	$\hat{y}_{Ridge}$	$ \text{Error} _{OLS}$	$ \text{Error} _{Ridge}$
$x_6$	0	1,46136	1,48466	0,46136	0,48466
$x_7$	12	7,81584	7,80142	1,61584	1,60142
$x_8$	5	4,10906	4,17665	0,50906	0,51665

$$MAE_{\text{test, OLS}} = \frac{0,46136 + 1,61584 + 0,50906}{3} \approx 0,86207$$

$$MAE_{\text{test, ridge}} = \frac{0,48466 + 1,60142 + 0,51665}{3} \approx 0,86758$$

Explanation: The ridge model shows a slightly higher training MAE and a similar test MAE compared to OLS. This behavior is what we expect from regularization (ridge adds a small bias aiming to reduce overfitting and improve stability on new data [although, that is not prominent with these results]).

