ridge regression

(3) ordinary least squares — minimize sum of squared residuals

$$\min_{\theta} \sum_{i=1}^{N} (\theta^T x^i - y^i)^2$$

$x^i \to$ feature vector

$y^i \to$ vector

θ parameters/weights

adding data points to the dataset as:

$x^{(N+1)} = (\sqrt{\lambda}, 0, 0, \ldots, 0)$ $x^{N+2} = (0, \sqrt{\lambda}, 0, \ldots 0)$

$y^{(N+1)} =$ $x^{N+d} = (0, 0, 0, \sqrt{\lambda})$

so now you have $\min_{\theta} \sum_{i=N+1}^{N+d} (\theta^T x^i - y^i)^2$

so,

$$= \boxed{\min_{\theta} \sum_{i=1}^{N} (\theta^T x^i - y^i)^2 + \min_{\theta} \sum_{i=N+1}^{N+d} (\theta^T x^i - 0)^2}$$

$\underline{i = N+1}$

$x^{(N+1)} = (\sqrt{\lambda}, 0, 0, \cdots, 0)$, $y^{(N+1)} = 0 = (\theta^T x^{N+1} - 0)^2 =$

$(\theta_1 \cdot \sqrt{\lambda})^2 = \lambda \theta_1^2$

$\underline{i = N+2}$

$x^{(N+2)} = (0, \sqrt{\lambda}, 0, \ldots, 0)$, $y^{(N+2)} = 0$

$(\theta^T x^{(N+2)} - 0)^2 = (\theta_2 \cdot \sqrt{\lambda})^2 = \lambda \theta_2^2$

$\underline{i = N+d}$

$$\sum_{i=N+1}^{N+d} (\theta^T x^i)^2 = \lambda \sum_{j=1}^{d} \theta_j^2$$

this is the same as the ridge regression ridge regression

① ↓ rmse means better fit. So, the left would have the highest and the middle has the smallest.

↳ left: data points are flat so no clear relation between x and y = 4 with little variance

↳ middle: linear relationship where a regression model would fit a line through it.

↳ right: linear trend with scatter which means RMSE would increase compared to the middle but still be lower than plot 1 (left) since it has a sort-of linear relation

② R^2 close to 1 is a good fit, R^2 close to 0 is a bad one. This is measuring proportional variance in dependent variable.

↳ leftmost plot is flat so no relationship can be seen so R^2 would be close to 0

↳ middle plot is a linear relationship so R^2 would be close to 1 and reflect variance in y

↳ right plot shows a linear

therefore the middle plot would have the highest R^2 and the left would have the smallest since there's little relation between x and y.