ML dseb62 w4 - Nguyen Tuan Duy

October 2022

1 Problem 2c

Ridge regression
Loss function:

$$L_{ridge} = \frac{1}{2N}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)^2 + \lambda w_1^2$$
$$= \frac{1}{2N}(||Y - XW||_2^2 + \lambda||W||^2)$$

minimizing L_{ridge} is the same as minimizing $L = ||Y - XW||_2^2 + \lambda||W||^2$

$$\frac{\delta L}{\delta W} = 2X^T(Y - XW) + 2\lambda W = 0$$
$$W = (X^T X + I\lambda)^{-1} X^T Y$$

Lasso regression:

$$L_{lasso} = \frac{1}{2N}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)^2 + \lambda|w_1|$$

because $\lambda|w_1|$ cannot be derived we have to optimize the function using gradient descent
* algorithm:
we update $w = w - \alpha * dw$ and $b = b - \alpha * db$ where α is the learning rate and d
if $w_i > 0$:
$dw = \frac{-2}{m}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)^2 + \lambda$
if $w_i \leq 0$:
$dw = \frac{-2}{m}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)^2 - \lambda$
and $db = \frac{-2}{m}(w_0 + w_1 x_i - y_i)$