ML dseb62 w4 - Nguyen Tuan Duy

October 2022

1 problem 1

using the formula:

$$W = (X^T X)^{-1} X^T T$$

but to fit second-order function, X is now:

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ . & . & \\ . & . & \\ 1 & x_n & x_n^2 \end{bmatrix} W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

2 problem 2a

same as problem 1 but:

$$\text{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & . & . & x_1^m \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & x_n & x_n^2 & x_n^3 & . & . & x_n^m \end{bmatrix} \text{W} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ . \\ .w_m \end{bmatrix}$$

3 Problem 2b

Using k-fold cross-validation taking available data and partitioning it into k groups. Then k-1 of the groups are used to train a set of modes that are then evaluated on the remaining groups. this procedure is then repeated for all k choices for the validation groups. In this problem I use mean squared error (MSE) as an addition way to evaluate the model in each fold

$$MSE = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{n}$$

4 Problem 2c

Ridge regression
Loss function:

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

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

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

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 $db = \frac{-2}{m}(w_0 + w_1 x_i - y_i)$