ML HW2

107061218 謝霖泳

1. The difference between Maximum Likelihood and Bayesian Linear Regression

- definition of **w**

  The main difference between maximum likelihood and Bayesian linear regression is the `w` they use. Refer to `equation 3.15` on the textbook, `w` for maximum likelihood is given as

  $$w_{ML} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$

  And the `w` for Bayesian linear regression is given in `equation 3.53`

  $$m_N = \beta S_N \Phi^T \mathbf{t}$$

  where S_N is given in `equation 3.54`

  $$S_N = (\alpha I + \beta \Phi_T \Phi)^{-1}$$

  Therefore, majority of codes in `BLR()` and `MLR()` are exactly the same, except for the calculation of **w**. Accroding to the above-mentioned definitions, we can see corresponding calculation of **w** as listed.

  - `MLR()`

    ```
    w = np.matmul(np.matmul(np.linalg.inv(np.matmul(np.transpose(phi), phi)), \
        np.transpose(phi)), chance_of_admit)
    ```

  - `BLR()`

    ```
    sn = np.linalg.inv(alpha * I + beta * np.matmul(np.transpose(phi), phi))
    w = beta * np.matmul(sn, np.matmul(np.transpose(phi), chance_of_admit))
    ```

- regularization term

  The second difference of these two methods is the use of `regularization term` , i.e. α and β which are used in `BLR()` but not used in `MLR()` .

2. The impact of different choices of O_1 and O_2

I tune the value of (O_1, O_2) from deault value (5, 5) into (2, 2) and (8, 8) and (20, 20). The results are shown below.

O1	O2	MSE of BLR	MSE of MLR
2	2	0.00757	0.00700
5	5	0.00743	0.00877
8	8	0.00762	0.09691
20	20	0.00832	$5.03062 * 10^{37}$

I find that no matter how O_1 and O_2 change, MSE of `BLR` changes very little. However, MSE of `MLR()` may surge drastically.

The key leading to this difference is the use of `regularization term`, i.e. α and β. `MLR()` doesn't involve these terms, so it tends to overfit. Therefore, when O_1 and O_2 change, the error may increase a lot. However, `BLR()` method involves regularization term, so it lowers the effect of O_1 and O_2.

As for the best choice for O_1 and O_2, I think there isn't a specific answer. I use for-loop to iterate over some possible (O_1, O_2) combinations and found that optimal (O_1, O_2) for each method doesn't happen concurrently. Take the above table for example. When we lower (O_1, O_2) from default value (5, 5) to (2, 2), we can find MSE of `BLR()` slightly increases, but MSE of `MLR()` drops slightly. However, when we increase (O_1, O_2) to (8, 8), both MSE increase.

3. My result

O1	O2	MSE of BLR	MSE of MLR
5	5	0.00743	0.00877

The above result is estimated under (α, β) = (0.5, 0.5).