

SMAI-2020-Homework

September 2020

1 Objective Question [1 mark]

Consider the following Ridge regression objective:

$$J(\mathbf{w}) = \arg \min_{\mathbf{w}} (\mathbf{Y} - \mathbf{XW})^2 + \lambda \|\mathbf{W}\|_2^2$$

What will the solution be for $\lambda = 0$ and $\lambda = \infty$. Select the correct pair from the options below.

1. 0, MLE
2. MLE, 0
3. 0, MAP
4. MAP, 0

Here, MLE is the maximum likelihood estimate and MAP is the maximum a posteriori estimate.

2 Subjective Question [2 marks]

You are interested in the problem of estimating the price of houses. A property broker told you that according to him the price of each house can depend on 9 different factors like locality, number of rooms in the houses, area of the property, etc. He provides you the detailed report of 7 such properties (i.e. for each of the 7 houses, you have been provided with the values of the 9 features). Having studied SMAI, you decide to use Linear regression for the estimating the housing prices and attempt to use the analytical or the closed form solution for the task. What is the chief issue in applying a closed form solution for this task? Will employing Ridge regression help? Explain your answer in brief.

3 Programming Question [3 marks]

Visualize the optimal solution obtained by Ridge regression and Lasso for different values of the hyper-parameter λ .

1. Complete the notebook 'Regression exercise.ipynb'. (2 marks)
2. Observe the output of the Ridge regression and Lasso and comment whether the following statement is **true or false**. “ The Lasso L1 term provides a much more aggressive regularization because the intersection between the constraint function and the cost function happens at the “vertices” of the diamond where either/or each variable is equal to zero. The effect of such regularization is to “cancel” out each variable, and by varying λ we can test the impact of each variable on the model. In effect this is automatic variable selection. On the other hand, Ridge regression provides a less aggressive form of regularization where the coefficients tend to zero in the limit only.” (1 mark)