Homework 8

Juan Diego Sánchez-Torres.[1]

14 April, 2021

[1] Department of Mathematics and Physics, ITESO. dsanchez@iteso.mx

In this activity, an image classification problem is reproduced. The purpose is to become familiar with a form of programming, given in the mentioned example, that is quite convenient in solving data modeling problems. Besides, a problem related to the selection of variables for regression is solved, and the model's performance can be improved.

Introduction

The purpose of this last activity is to solve a problem similar to those encountered in practice, using a reasonably sophisticated code structure appropriate to the solution of the problem. Also, use is made of symbolic transformers for the resolution of a regression problem. It is demonstrated that the use of new inputs can improve the model's predictive capacity.

Activities

Problem 1: Using Transformed Features

Read and reproduce the example about the Boston housing dataset given in Gplearn: Symbolic Transformer. Then,

1. Explain how the *symbolic transformer* method helps to improve the linear regression's performance.
2. Subsequently, solve a problem of your choice from the UC Irvine Machine Learning Repository.

Upload your results to Github in the form of a Jupyter notebook, then make it interactive using Binder, hence submit your results through both links. The use of Google Colab is highly recommended.

To further elaborate on the importance of having adequate model inputs and the application transformers to obtain such inputs, consider the following problem.

Problem 2: An Application to Image Classification

Read and reproduce the Tutorial: image classification with scikit-learn carefully, then complete the mathematical deductions and code descriptions. Upload your results to Github in the form of a Jupyter notebook, then make it interactive using Binder, hence submit your results through both links. The use of Google Colab is highly recommended.

The following documents may help understand the joint use of Google Colab and Github:

- Using Google Colab with GitHub
- Google Colab Tips: Easy export notebook to Github
- Google Drive + Google Colab + GitHub; Don't Just Read, Do It!

Further Lectures and Bibliography

For this activity, consider the following references:

- The Smola's tutorial on SVR[4].
- The book of Abe[1], presents a comprehensive study of support vector machines.
- The book of Suykens[5], introduces the basic concepts of support vector machines.
- The book of Deisenroth[3], presents a detailed calculation of support vector machines (Chapter 12., pp. 370-394).
- The book of Bishop[2], presents a rigorous deduction of support vector machines and kernel based methods (Chapter 6 and 7., pp. 291-358).

References

[1] Shigeo Abe. *Support Vector Machines for Pattern Classification, 2 Ed.* Springer-Verlag London, 2010. ISBN 978-1-84996-097-7. URL https://www.springer.com/gp/book/9781849960977.

[2] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York, 2006. ISBN 97814939938438. URL https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf.

[3] M.P. Deisenroth, A.A. Faisal, and C.S. Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020. ISBN 9781108470049. URL https://mml-book.github.io/book/mml-book.pdf.

[4] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2004. ISSN 1573-1375. doi: 10.1023/B:STCO.0000035301.49549.88. URL https://alex.smola.org/papers/2004/SmoSch04.pdf.

[5] Johan A K Suykens, Tony Van Gestel, Jos De Brabanter, Bart De Moor, and Joos Vandewalle. *Least Squares Support Vector Machines*. World Scientific, 2002. ISBN 9789812381514. URL https://www.worldscientific.com/worldscibooks/10.1142/5089.