Credit Task 2

Submission requirements: Submit a report (pdf format) and the code file (.ipynb), which is testable by the marker.

1.   What are the differences between hyperparameters and parameters of a machine learning (ML) model?   Explain your answer using at least two machine learning models that you have learned in this unit.

2.   Prove that Elastic net can be used as either LASSO or Ridge regulariser.

Background

The recently started human and other genome projects are likely to change the situation of molecular biology. Comprehensive analyses of whole genomic sequences will enable us to understand the general mechanisms of how protein and nucleic acid functions are encoded in the sequence data.

Dataset filename: yeast2vs4.csv

Dataset description: There are 8 features and one target in the dataset. All the features are in a numerical format, and the target is in text format. For further information about the attributes, please read "Data Set Information.pdf".

3.   Analyse the importance of the features for predicting presence or absence of protein using two different approaches. Explain the similarity/difference between outcomes.

4.   Create three supervised machine learning (ML) models except any ensemble approach for predicting presence or absence of protein.

  a.   Report performance score using a suitable metric. Is it possible that the presented result is an overfitted one? Justify.

  b.   Justify different design decisions for each ML model used to answer this question.

  c.   Have you optimized any hyper-parameters for each ML model? What are they? Why have you done that? Explain.

  d.   Finally, make a recommendation based on the reported results and justify it.

Assessment feedback

The results with comments will be released within 5 business days from the due date.

Referencing

You must correctly use the Harvard method in this assessment. See the Deakin referencing guide.

Academic integrity, plagiarism and collusion

Plagiarism and collusion constitute extremely serious breaches of academic integrity. They are forms of cheating, and severe penalties are associated with them, including cancellation of marks for a specific assignment, for a specific unit or even exclusion from the course. If you are ever in doubt about how to properly use and cite a source of information refer to the referencing site above.

Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgement as to its authorship, the work of any other person or resubmits their own work from a previous assessment task.

Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose, with the intent of obtaining an advantage in submitting an assignment or other work.

Work submitted may be reproduced and/or communicated by the university for the purpose of assuring academic integrity of submissions: https://www.deakin.edu.au/students/study-support/referencing/academic-integrity.