TÉCNICO LISBOA

Aprendizagem 2021/22

Homework II

Deadline 27/10/2021 23:59 via Fenix as PDF

- Homework limited to 4 pages (2.5−3pp for part I, 1−1.5pp for part II) according to the provided template
- Include your programming code as an Appendix (maximum 1 page)
- Submission Gxxx.PDF in Fenix where xxx is your group number. Please note that it is possible to submit several times on Fenix to prevent last-minute problems. Yet, only the last submission is considered valid
- Exchange of ideas is encouraged. Yet, if copy is detected after automatic/manual clearance, homework is nullified and IST guidelines apply for content sharers and consumers, irrespectively of the underlying intent
- Please consult the FAQ before posting questions to your faculty hosts

I. Pen-and-paper [13.5v]

Given the following integer data:

		y_1	y_2	y_3	$output$
Train	$\mathbf{x}_1$	1	1	0	1
	$\mathbf{x}_2$	1	1	5	3
	$\mathbf{x}_3$	0	2	4	2
	$\mathbf{x}_4$	1	2	3	0
	$\mathbf{x}_5$	2	0	7	6
	$\mathbf{x}_6$	1	1	1	4
	$\mathbf{x}_7$	2	0	2	5
	$\mathbf{x}_8$	0	2	9	7
test	$\mathbf{x}_9$	2	0	0	2
	$\mathbf{x}_{10}$	1	2	1	4

1) [5v] Consider the following basis function $\phi_j(\mathbf{x}) = \|\mathbf{x}\|_2^{\,j}$
   Learn the following polynomial regression model:
$$f(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{3} w_j \cdot \phi_j(\mathbf{x})$$
2) [2v] Identify the RMSE on the testing data.
3) [5v] Consider an equal depth binarization of y_3 and class targets to be defined as:
$$t_i = \begin{cases} \text{P}, & output_i \geq 4 \\ \text{N}, & \text{else} \end{cases}$$
   Learn a decision tree using ID3.
4) [1.5v] Identify the classification accuracy on the testing data.

II. Programming and critical analysis [6.5v]

Recall the `breast.w.arff` from previous homework.

5) [3.5v] In a single plot, compare the training and testing accuracy of a decision tree with a varying:
   i. number of selected features in {1,3,5,9} using mutual information (tree with no fixed depth)
   ii. maximum tree depth in {1,3,5,9} (with all features and default parameters)
6) [1.5v] Identify two reasons for the observed correlation.
7) [1.5v] Select a specific depth. Justify

END