TÉCNICO LISBOA

Aprendizagem 2021/22

Homework II

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

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:

$$f(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{3} w_j \cdot \phi_j(\mathbf{x})$$

		$\|\mathbf{x}\|_2$	$\|\mathbf{x}\|_2{}^2$	$\|\mathbf{x}\|_2{}^3$	z
$\mathbf{x}_1$	1	1.414214	2	2.828427	1
$\mathbf{x}_2$	1	5.196152	27	140.2961	3
$\mathbf{x}_3$	1	4.472136	20	89.44272	2
$\mathbf{x}_4$	1	3.741657	14	52.3832	0
$\mathbf{x}_5$	1	7.28011	53	385.8458	6
$\mathbf{x}_6$	1	1.732051	3	5.196152	4
$\mathbf{x}_7$	1	2.828427	8	22.62742	5
$\mathbf{x}_8$	1	9.219544	85	783.6613	7

Closed form solution $\mathbf{w} = (X^T \cdot X)^{-1} \cdot X^T \cdot \mathbf{z}$

where $X^T \cdot X =$

8.00000000e+00	3.58842916e+01	2.12000000e+02	1.48228114e+03
3.58842916e+01	2.12000000e+02	1.48228114e+03	1.14360000e+04
2.12000000e+02	1.48228114e+03	1.14360000e+04	9.35735164e+04
1.48228114e+03	1.14360000e+04	9.35735164e+04	7.93976000e+05

Inverse $(X^T \cdot X)^{-1} =$

8.19551657e+00	-6.23130178e+00	1.30493617e+00	-7.93403817e-02
-6.23130178e+00	5.07809652e+00	-1.10435916e+00	6.86446430e-02
1.30493617e+00	-1.10435916e+00	2.47213973e-01	-1.56648459e-02
-7.93403817e-02	6.86446430e-02	-1.56648459e-02	1.00683060e-03

Product with $(X^T \cdot X)^{-1} \cdot X^T$ yielding

1.76858894	-0.08114793	-0.66940817	-1.00687669	1.37941711	0.90512909	-0.78504492	-0.51065743
-1.06435076	-0.03185877	0.5312166	0.90399365	-1.30698763	-0.39217005	0.85010185	0.51005512
0.19325755	0.0435779	-0.09073507	-0.18677661	0.32320534	0.05237499	-0.19540644	-0.13949766
-0.01074414	-0.00434877	0.00440454	0.01093752	-0.02135529	-0.00220726	0.0122792	0.01103421

Finally, the product with $\mathbf{z}$ yields our parameters $w = \begin{pmatrix} 4.58352 \\ -1.6872 \\ 0.3377 \\ -0.0133 \end{pmatrix}$

Aprendizagem 2021/22

Homework II

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

Grading criteria:

- Transformation Matrix (2 pts):
  - objective: Generate the correct transformation matrix
  - common penalty: Adding bias before transformation (-1 pt)
  - common penalty: Mistake in norm calculation (-1 pt)
  - common penalty: No transformation or transformation not shown (-2 pts)
- Bias (1 pt):
  - objective: Add the bias factor to the transformation matrix
  - common penalty: No bias (-1 pt)
- Formula (1 pt):
  - objective: State the closed form formula used to obtain the parameters
  - common penalty: No or wrong formula (-1 pt)
- Parameters (1 pt):
  - objective: Correct calculation of the parameters
  - common penalty: Mistake in calculation (-0.5 pts)
  - common penalty: Several mistakes in calculation (-1 pt)
  - common penalty: Wrong parameters and calculation steps not shown (-1 pt)
  - common penalty: No parameters (-1 pt)

2) [2v] Identify the RMSE on the testing data.

Observations in the transformed space:

		$\|\mathbf{x}\|_2$	$\|\mathbf{x}\|_2^{\,2}$	$\|\mathbf{x}\|_2^{\,3}$	z
$\mathbf{x}_9$	1	2	4	8	2
$\mathbf{x}_{10}$	1	2.44949	6	14.69694	4

$$\hat{z} = \mathbf{w}^T\mathbf{x}$$

$$\hat{z}_9 = (4.58352 \quad -1.6872 \quad 0.3377 \quad -0.0133)\begin{pmatrix}1\\2\\4\\8\end{pmatrix} = 2.45$$

$$\hat{z}_{10} = (4.58352 \quad -1.6872 \quad 0.3377 \quad -0.0133)\begin{pmatrix}1\\2.44949\\6\\14.69694\end{pmatrix} = 2.28$$

$$RMSE = \sqrt{\frac{1}{2}((2-2.45)^2 + (4-2.28)^2)} = 1.256$$

Grading criteria:

- predictions (1 pt):
  - objective: Correct predictions for the test instances
  - common penalty: Prediction formula not shown (-0.25 pts)
  - common penalty: Small mistake in calculation (-0.25 pts)
  - common penalty: Mistakes in calculation (-0.5 pts)
  - common penalty: Wrong predictions and no calculation steps (-1 pt)
  - common penalty: No or wrong transformation (-1 pt)
  - common penalty: Predictions on training set instead of test set (-1 pt)
  - common penalty: Predictions not shown (-1 pt)
- RMSE (1 pt):
  - objective: Correct computation of the error
  - common penalty: Mistake in calculation (-0.5 pts)
  - common penalty: No RMSE formula but correct value (-0.5 pts)
  - common penalty: Wrong RMSE formula (-1 pt)
  - common penalty: RMSE not shown (-1 pt)

Aprendizagem 2021/22

Homework II

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

3) [5v] Consider an equal depth binarization of y_3 and class targets to be defined as:

$$t_i = \begin{cases} P, & output_i \geq 4 \\ N, & else \end{cases}$$

Learn a decision tree using ID3.

discretization $y_3{}^{disc} = (0,1,1,0,1,0,0,1)$

E(out)	1		
E(out|y1)	0.655639	IG(y1)	0.344361
E(out|y2)	0.688722	IG(y2)	0.311278
E(out|y3)	1	IG(y3)	0

select y1	=2	c=1		
	=0	other variables unable to explain, e.g. c=1		
		y2	y3	out
	=1	1	0	0
		1	1	0
		2	0	0
		1	0	1
select y3	=1	c=0		
	=0	y2	out	
		1	0	
		2	0	
		1	1	
select y2	=2	c=0		
	other	c=1 (with error)		

Additional intermedium calculus:

- E(out|y1=0, y2) = 1 IG(out|y1=0, y2) = 0
- E(out|y1=0, y3) = 1 IG(out|y1=0, y3) = 0
- E(out|y1=1, y2) = 0.689 IG(out|y1=1, y2) = 0.123
- E(out|y1=1, y3) = 0.689 IG(out|y1=1, y3) = 0.123

Grading criteria:

- y3 binarization: 0.5/5v
- Entropy and Information Gain formulas: 0.2/5v
- E and IG for the root: 2.1/5v
- E and IG for the lower levels 1.6/5v
- Tree Presentation: 0.6/5v
- Calculation Mistakes: -0.2v
- Presentation Mistake -0.1v

4) [1.5v] Identify the classification accuracy on the testing data.

After binarization of y3 and output:

	y_1	y_2	y_3	$output$
$\mathbf{x}_9$	2	0	0	0
$\mathbf{x}_{10}$	1	2	0	1

Given the learned tree, $\hat{z}_9 = 1$ and $\hat{z}_{10} = 0$. 1 false negative and 1 false positive. Accuracy is then 0.

Grading criteria

- Correct Accuracy: 0.5/1.5v
- Correct Prediction: 1/1.5v
- Presentation Mistake -0.2v

Aprendizagem 2021/22

Homework II

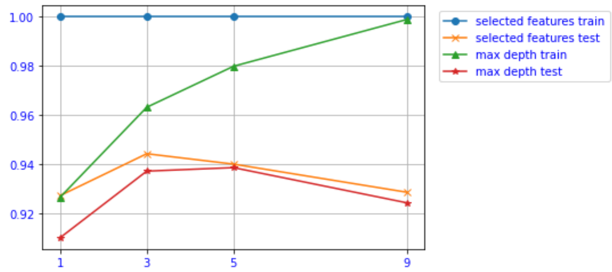
Deadline 27/10/2021 23:59 via Fenix as PDF

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)

Solution note: there is a soft correlation between test accuracies and max depth and selected features. Given the variability of estimates, the observed differences for some seeds are not statistically significant, more commonly with regards to feature selection.



Common discounts:
- Did not plot all the requested metrics (1.75val);
- No plot (0.0val);
- Bar plot with unclear identification of respective metrics (1.75val)
- Missing legend (0.875val).

6) [1.5v] Identify two reasons for the observed correlation.

Solution note: concordant underfitting and overfitting risks per strategy.

Joint analysis: limiting the depth of a decision tree limits the maximum number of selected features, and vice-versa.

Common discounts:
- No justification (0.0val);
- Only one justification (0.75val);
- Repetition of justification (0.75val).

7) [1.5v] Select a specific depth. Justify

Solution note: for the collected results in (5), a depth of 5 produces the maximum test accuracy, suggesting a good balance between underfitting and overfitting risks.

Different seeds can provide moderately different results. Answers should be in agreement with evidence gathered in (5).

Common discount: not in accordance with exercise 5) (0.0val).

END