**I. Pen-and-paper**

1. First, we transformed the design matrix using the basis function

Then, we will calculate the vector W that minimizes the square-error loss function, this vector is given by the following expression

So, for the training dataset we obtain that

1. To measure the differences between the observed values and the predictions given by the model, in order to test it, we’ll use the RMSE

For that we’ll need to calculate the predictions of our model:

Each prediction is given by the following polynomial regression model

Therefore, the RMSE is:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | 0 | N |
|  | 1 | N |
|  | 1 | N |
|  | 0 | N |
|  | 1 | P |
|  | 0 | P |
|  | 0 | P |
|  | 1 | P |

|  |  |
| --- | --- |
|  | 0.25 |
|  | 0.5 |
|  | 0.25 |
|  | 0.5 |
|  | 0.75 |
|  | 0 |
|  | 0.5 |
|  | 0.25 |
|  | 1 |

|  |  |
| --- | --- |
|  | 0.25 |
|  | 0.375 |
|  | 0.375 |
|  | 1 |
|  | 0.666667 |
|  | 0.666667 |
|  | 0 |
|  | 0.333333 |
|  | 0.333333 |

|  |  |
| --- | --- |
|  | 0.5 |
|  | 0.5 |
|  | 0.5 |
|  | 0.5 |
|  | 0.5 |
|  | 0.5 |

1. Firstly, we computed the equal depth binarization of , in which we used the median of its train values as the criteria for the binarization, , and the class targets , so we obtained the following table:  
    To learn a decision tree using ID3, we need to calculate the information gain (IG) of each variable, which is given by , and the entropy.  
   Probabilities of the training dataset:  
     
     
     
     
     
     
   -  
   -  
   -Information Gain: -

Diagram

Description automatically generated  
We can conclude that the variable that will be used as root for the decision tree is as it is the variable with the highest IG, and always leads to , and to uncertain (?). So, we will need to study now only the cases that , which are on the following table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | y2 | y3 | t |
| x1 | 1 | 0 | N |
| x2 | 1 | 1 | N |
| x4 | 2 | 0 | N |
| x6 | 1 | 0 | P |

From this, we can conclude that is uncertain, and that leads to . Since the information gain of and is the same (using analogous calculus from above), we choose the add to the decision tree, which leaves us with the final table:

|  |  |  |
| --- | --- | --- |
|  | y3 | t |
| x1 | 0 | 0 |
| x2 | 1 | 0 |
| x6 | 0 | 1 |

Here, we can conclude that when , then , and with analogous methods, it is clear that when the information gain will be 0, so that will lead to uncertainty. With all this information, we are now able to create the decision tree:

1. Using the “test” dataset on the decision tree obtained earlier ( and ) we get a prediction , (by following the branch), opposed to the true values of , , so, the accuracy will be:

**II. Programming and critical analysis**

1. Answer 5
2. Answer 6
3. Answer 7
4. Answer 8

**III. APPENDIX**

Paste your programming code here using Consolas 9pt or 10pt.

Use **highlighting** or colored text to facilitate the analysis by your faculty hosts.

**END**