### Class 21BIT – Term II/2023-2024

### Course: CS300 – Artificial Intelligence

Homework 04

***Submission Notices:***

* *Conduct your homework by filling answers into the placeholders given in this file (in Microsoft Word format). Questions are shown in black color, instructions/hints are shown in italic and blue color, and your content should use any color that is different from those.*
* *After completing your homework, prepare the file for submission by exporting the Word file (filled with answers) to a PDF file, whose filename follows the following format,*

*<StudentID-1>\_<StudentID-2>\_HW02.pdf (Student IDs are sorted in ascending order)*

*E.g.,* ***2152001\_2152002\_HW02.pdf***

*and then submit the file to Moodle directly WITHOUT any kinds of compression (.zip, .rar, .tar, etc.).*

* *Note that you will get zero credit for any careless mistake, including, but not limited to, the following things.*
  1. *Wrong file/filename format, e.g., not a pdf file, use “-” instead of “\_” for separators, etc.*
  2. *Disorder format of problems and answers*
  3. *Conducted not in English.*
  4. *Cheating, i.e., copying other students’ works or let the other student(s) copy your work.*

**Problem 1. (3pts)** For each of the learning types mentioned in the below table, present an application by briefly addressing the following points.

* The problem to be solved in the application
* How is the learning model built from available data?
* A list of references

Note that an application here can be either a system in practical or an academic research work.

|  |  |
| --- | --- |
| **Types of learning** | **Application** |
| Supervised learning | *Application: Email Spam Classification*   * **Problem**: Classify emails as spam or non-spam. * **Model**: Built from labeled data using algorithms like SVM or Naive Bayes. * **References**: Sahami et al. (1998), Androutsopoulos et al. (2000). |
| Unsupervised learning | *Application: Customer Segmentation for Marketing*   * **Problem**: Segment customers based on behavior and demographics. * **Model**: Built from unlabeled data using clustering algorithms like k-means. * **References**: Jain & Dubes (1988), Han et al. (2011) |
| Reinforcement learning | *Application: Autonomous Driving*   * **Problem**: Develop a self-driving car. * **Model**: Learns driving decisions through interaction with the environment using algorithms like Q-learning. * **References**: Sutton & Barto (2018), Silver et al. (2016). |

**Problem 2. (4pts)** Consider the following dataset of five examples, in which **EnjoySport** is the target attribute.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sky** | **AirTemp** | **Humidity** | **Wind** | **Water** | **Forecast** | **EnjoySport** |
| sunny | warm | normal | strong | warm | same | yes |
| sunny | warm | high | strong | warm | same | yes |
| rainy | cold | high | strong | warm | change | no |
| sunny | warm | high | strong | cool | change | yes |
| sunny | warm | normal | weak | warm | same | no |

1. (3pts) Show calculations to train an ID3 decision tree classifier from the given dataset. Attributes are evaluated using information gain measure *(precision to 3 decimal places).*

*(1pt) Calculate the Average Entropy and Information Gain for each of the following attributes*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| H(D) = **0,971** | Sky | AirTemp | Humidity | Wind | Water | Forecast |
| Average Entropy (AE) | **0,971** | **0,971** | **0,971** | **0,971** | **0,971** | **0,971** |
| Information Gain (IG) | **0,25** | **0,02** | **0,02** | **0,02** | **0,119** | **0,02** |

*(2pts) There is a three-way tie between Sky, AirTemp, and Wind for the choice of attribute at the root. For each case, continue to calculate AE and IG for the remaining attributes*

|  |  |  |  |
| --- | --- | --- | --- |
| **Root** | **Current Tree** | **Calculations** | **Final Tree** |
| Sky | *Draw the tree here*  H(Sky=Sunny)= **-1,068** | |  |  |  | | --- | --- | --- | |  | AE | IG | | AirTemp | **0,918** | **0,031** | | Humidity | **0,918** | **0,031** | | Wind | **0,918** | **0,031** | | Water | **0,918** | **0,151** | | Forecast | **0,918** | **0,031** | | *Draw the tree here* |
| AirTemp | *Draw the tree here*  H(Sky=Sunny)= **-1,068** | |  |  |  | | --- | --- | --- | |  | AE | IG | | AirTemp | **0,971** | **0,118** | | Humidity | **0,971** | **0,02** | | Wind | **0,971** | **0,02** | | Water | **0,971** | **0,119** | | Forecast | **0,971** | **0,02** | | *Draw the tree here* |
| Wind | *Draw the tree here*  H(Wind=Strong)=**-1,068** | |  |  |  | | --- | --- | --- | |  | AE | IG | | Sky | **0,971** | **0,118** | | AirTemp | **0,918** | **0,031** | | Humidity | **0,918** | **0,031** | | Water | **0,918** | **0,151** | | Forecast | **0,918** | **0,031** | | *Draw the tree here* |

1. (1pt) Which class will be assigned to each of the following examples?

|  |  |
| --- | --- |
| **Example** | **EnjoySport** |
| [Sky = rainy, AirTemp = cold, Humidity = normal, Wind = weak, Water = cool, Forecast = change] | **no** |
| [Sky = sunny, AirTemp = cold, Humidity = high, Wind = strong, Water = cool, Forecast = same] | **yes** |
| [Sky = rainy, AirTemp = warm, Humidity = normal, Wind = weak, Water = warm, Forecast = same] | **yes** |

**Problem 3. (3pts)** In the network shown below, all the units have binary inputs (0 or 1), unipolar step functions and binary outputs (0 or 1). The weights for this network are w31 = 1, w32 = 1, w41 = 1, w42 = 1 and w43 = −2. The threshold of the hidden unit (3) is 1.5 and the threshold of the output unit (4) is 0.5. The threshold of both input units (1 and 2) is 0.5, so the output of these units is the same as the input.

A diagram of a diagram

Description automatically generated

Which Boolean functions can be computed by this network? Justify your answer by showing detailed calculations.

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 2 | 3 | 4 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |