**COMP 3710 Applied Artificial Intelligence**

**Seminar/Lab 8.**

**Decision tree, and k-Nearest Neighbor (kNN) algorithm**

1. **Objectives**

* How to construct decision tree and how to use it
* Use of k-Nearest Neighbor (kNN) algorithm

1. **Decision tree**

Here is the training data set.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play** |
| *Sunny* | *Hot* | *High* | *False* | *No* |
| *Sunny* | *Hot* | *High* | *True* | *No* |
| *Overcast* | *Hot* | *High* | *False* | *Yes* |
| *Rainy* | *Mild* | *High* | *False* | *Yes* |
| *Rainy* | *Cool* | *Normal* | *False* | *Yes* |
| *Rainy* | *Cool* | *Normal* | *True* | *No* |
| *Overcast* | *Cool* | *Normal* | *True* | *Yes* |
| *Sunny* | *Mild* | *High* | *False* | *No* |
| *Sunny* | *Cool* | *Normal* | *False* | *Yes* |
| *Rainy* | *Mild* | *Normal* | *False* | *Yes* |
| *Sunny* | *Mild* | *Normal* | *True* | *Yes* |
| *Overcast* | *Mild* | *High* | *True* | *Yes* |
| *Overcast* | *Hot* | *Normal* | *False* | *Yes* |

1. Construct a decision tree with the above table. You should show how the tree is constructed, by computing information gains and entropies. (You need to show how to compute Gain for all the attributes at each level.)

Gain(Outlook) = 1 – w-entropy(Sunny) – w-entropy(Overcast) – w-entropy(Rainy)

= 1 – 5/13 \* H(Sunny) – 4/13 \* H(Overcast) – 4/13 \* H(Rainy)

= 1 – 5/13 \* (-2/5 log2(2/5) – 3/5 log2(3/5))

– 4/13 \* H(Overcast)

– 4/13 \* H(Rainy)

…

…

1. Answer the next two queries using the above decision tree.

* (Sunny, Cold, High, True)
* (Overcast, Hot, High, True)

1. **k-Nearest Neighbor (kNN) algorithm**

Here is the training data set for class grades.

|  |  |  |
| --- | --- | --- |
| Height | Weight | T Shirt Size |
| 158 | 58 | Small |
| 158 | 59 | Small |
| 160 | 59 | Medium |
| 160 | 60 | Medium |
| 163 | 60 | Medium |
| 163 | 61 | Medium |
| 165 | 61 | Large |
| 165 | 62 | Large |
| 168 | 62 | Large |
| 158 | 63 | Medium |
| 168 | 63 | Large |
| 170 | 63 | Large |
| 160 | 64 | Large |
| 163 | 64 | Large |
| 170 | 64 | Large |
| 165 | 65 | Large |
| 168 | 66 | XLarge |
| 170 | 68 | XLarge |

* 1. Find the T shirt size for the next two queries using the 3NN algorithm.
* (166, 62)
* (159, 66)

1. **Assignment**
   * The title of the mail should include your name, id, and COMP 3710.

* You need to submit a document file for 2) and 3), not image of hand-written document.
  + Due:
    - 11:59 pm, March 4, 2019 – with bonus 10%
    - 6:00 pm, March 6, 2019 – with the full marks
    - 6:00 pm, March 7, 2019 – with penalty 5%
    - 6:00 pm, March 8, 2019 – with penalty 10%
* Total marks: 10 (= 7 + 3)