**COMP 3710 Applied Artificial Intelligence**

**Seminar/Lab 7.**

**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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Film** | ***Country*** | ***Big Star*** | ***Genre*** | **Success** |
| Film 1 | *USA* | *Yes* | *SF* | False |
| Film 2 | *USA* | *No* | *Comedy* | False |
| Film 3 | *USA* | *Yes* | *Comedy* | False |
| Film 4 | *Europe* | *No* | *Comedy* | True |
| Film 5 | *Europe* | *Yes* | *SF* | True |
| Film 6 | *Europe* | *Yes* | *Romance* | False |
| Film 7 | *Other* | *Yes* | *Comedy* | True |
| Film 8 | *Other* | *No* | *SF* | False |
| Film 9 | *Europe* | *Yes* | *Comedy* | False |
| Film 10 | *USA* | *Yes* | *Comedy* | True |

1. Construct a decision tree with the above table. You should show how in the tree is constructed, by computing information gains and entropies.

Gain(Country) = 1 – w-entropy(USA) – w-entropy(Europe) – w-entropy(Other)

= 1 – 4/10 \* H(USA) – 4/10 \* H(Europe) – 2/10 \* H(Other)

= 1 – 4/10 \* (-3/4 log2(3/4) – 1/4 log2(1/4)) – 4/10 \* H(Europe) – 2/10 \* H(Other) …

…

…

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

* (China, Yes, SF)
* (USA, No, Comedy)

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

Here is the training data set for class grades.

|  |  |  |  |
| --- | --- | --- | --- |
| MATH 1650 | COMP 2230 | COMP 3710 | Grade level |
| 4 | 3 | 4 | Excellent |
| 3 | 4 | 3 | Good |
| 3 | 4 | 4 | Excellent |
| 2 | 3 | 3 | Okay |
| 3 | 3 | 2 | Good |
| 2 | 3 | 2 | Okay |

* 1. Find the grade levels for the next two queries using the 3NN algorithm.
* (3, 2, 4)
* (3, 4, 3)

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.
  + Total marks: 10 (= 7 for 2) + 3 for 3))
  + Due: 6:00 PM, November 1, 2017
* Any late submission will NOT be accepted.