JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- APRIL 2019

M.Tech IV Semester

COURSE CODE: 11M1WCI431   MAX. MARKS: 25

COURSE NAME: Advanced Web Mining

COURSE CREDITS: 03   MAX. TIME: 1Hr 30 Min

Note: *All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.*

1. Write differences between: [CO3- 6 marks]
a) Supervised and unsupervised learning.
b) Distance based model and probabilistic model.
c) Boolean indexing and TFIDF indexing.

2. Explain the K-means clustering algorithm, include convergence and initialization issue of the algorithm.

[CO4- 6 marks]

3. Consider the documents (D1-D9) given below, assume K=3 and show the output of K-means clustering with respect a suitable distance measure chosen by you.   [CO4- 6 marks]

4. Consider the documents given below, assume   [CO4- 7 marks]

   {D1, D2, D3} = class news, {D4, D5, D6} = class Chinese, and {D7, D8, D9} = class home, show the output of Naïve Bayes classifier for documents {D10, D11, D12}.

   D1: Covers breaking news from around the world.
   D2: International news, analysis and commentary.
   D3: Internationally published free city news paper.
   D4: Chinese Beijing Chinese
   D5: Chinese Chinese Shanghai
   D6: Chinese Macao
   D7: new home sales top forecasts
   D8: home sales rise in july
   D9: increase in home sales in july

   D10: july new home sales rise
   D11: Chinese Chinese Tokyo Japan
   D12: news home city world