JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-3 EXAMINATION- JUNE -2016

M.Tech IV Semester

COURSE CODE: 11M1WCI431   MAX. MARKS: 35

COURSE NAME: Advanced Web Mining

COURSE CREDITS: 03   MAX. TIME: 2 HRS

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

1. What is distance based and probability based model for machine learning. Explain K-NN and
    Naive Bayes in this regards.   [2+2+3+3]

Doc 1 new home sales top forecasts

Doc 2 home sales rise in July

Doc 3 increase in home sales in July

Doc 4 July new home sales rise

Doc 5 breakthrough drug for schizophrenia

Doc 6 new schizophrenia drug

Doc 7 new approaches for treatment of schizophrenia

Doc 8 new hopes for schizophrenia patients

2. Using cosine similarity and k-means algorithm cluster the above documents into 2 groups.
                                                            [6]

3. Consider the documents {1, 2, 3, 5, 6, 7} as training documents classify {4, 8} using k-nearest
    neighbor method.   [4]

4. What is a linear classifier? Derive the equation of an optimal linear classifier for any given data
    sets.(assume that they are linearly separable)   [2+3]

5. Explain the layered architecture for semantic web. What is the role of RDF, RDFS, OWL and
    OWLS in formation of semantic web?   [2+2+2+2+2]