JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATIONS-2022

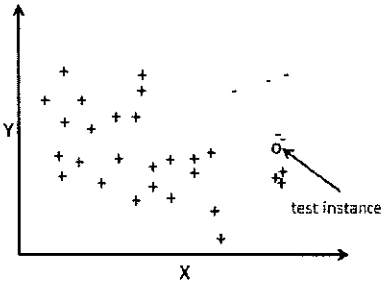
M.Tech. - I Semester (CS/IT/DS)

COURSE CODE (CREDITS): 22M1WCI133 (3) MAX. MARKS: 15

COURSE NAME: Introduction to Statistical Learning

COURSE INSTRUCTOR: Dr. Nancy Singla MAX. TIME: 1 Hour

Note: *All questions are compulsory. Marks are indicated against each question in square brackets.*

<table>
<tr><td>Q1.</td><td>To study the relationship between the marks obtained in Statistics (x) and marks in Economics (y) of the students of a school, a sample of ten students is taken and the following information is obtained.
$\Sigma(x-\bar{x})(y-\bar{y}) = 120,\ \Sigma(x-\bar{x})^2 = 80,\ \Sigma(y-\bar{y})^2 = 500$
Find the correlation coefficient between x and y.</td><td>[2] CO1</td></tr>
<tr><td>Q2.</td><td>Suppose you are working on weather prediction, and you would like to predict whether or not it would be raining at 5pm tomorrow. You want to use a learning algorithm for this. Would you treat this as a classification or a regression problem and why?</td><td>[2] CO2</td></tr>
<tr><td>Q3.</td><td>What is the Bias-Variance tradeoff?</td><td>[3] CO2</td></tr>
<tr><td>Q4.</td><td>A KNN classifier assigns a test instance the majority class associated with its K nearest training instances. Distance between instances is measured using Euclidean distance. Suppose we have the following training set of positive (+) and negative (-) instances and a single test instance (o). All instances are projected onto a vector space of two real-valued features (X and Y). Answer the following questions. Assume "unweighted" KNN (every nearest neighbor contributes equally to the final vote).
 

(a) What would be the class assigned to this test instance for K=3 and K=5?
(b) Setting K to a large value seems like a good idea. We get more votes! Given this particular training set, would you recommend setting K = 11? Why or why not?
(c) How to choose optimal value of K in KNN Algorithm?</td><td>[2+1+2] CO1</td></tr>
<tr><td>Q5.</td><td>The monthly sale of different types of laptops (in hundred units) and its profit (in lakh) for the last six months for a company is given below:
<table><tr><td>Months</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>No. of laptops sold (in hundred units) X</td><td>5</td><td>7</td><td>5</td><td>12</td><td>8</td><td>3</td></tr><tr><td>Profit (lakh) Y</td><td>8</td><td>9</td><td>10</td><td>15</td><td>10</td><td>6</td></tr></table>Obtain the regression line of Y on X. Also find the error in estimating Y for X =7.</td><td>[3] CO1</td></tr>
</table>