COURSE CODE (CREDITS): 18B1WCI634 & 18B11BI611 (2)   MAX. MARKS: 35

COURSE NAME: Machine Learning & Machine Learning for Bio-Informatics

COURSE INSTRUCTORS: **Mr. Praveen Modi**, Dr. Aman Sharma,   MAX. TIME: 02:00Hr

Dr. Shubham Goel and Mr. Kapil Rana

Note: (a)All questions are compulsory, (b) Marks are indicated against each question in square brackets,(c)The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems,(d) Write the answer of the question belonging to the same part in the same order.

Q1. (a) Consider the following ANN model with back propagation algorithm. Let the initial learning rate is 0.9. Weight and bias are given in the table. Find the updated weights and bias for one iteration? Use Error function (Err = O(1-O)(T-O), where O is the output of the neuron and T is the target output) and an activation function is sigmoid.   [7] (CO5)
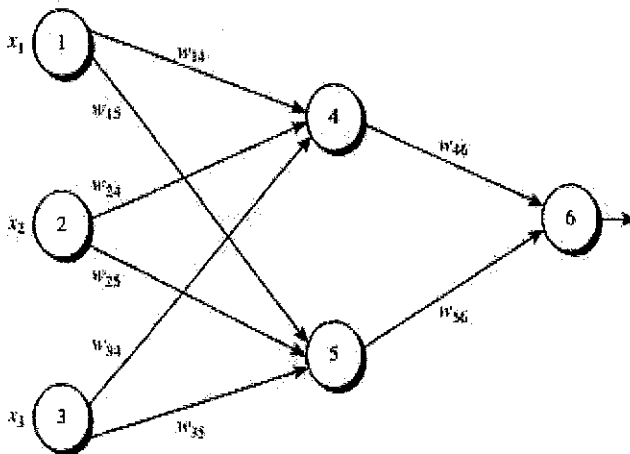
Weight/ bias	values	Weight/ bias	Values
x_1	1	w_{34}	-0.5
x_2	0	w_{35}	0.2
x_3	1	w_{46}	-0.3
w_{14}	0.2	w_{56}	-0.2
w_{15}	-0.3	θ_1	-0.4
w_{24}	0.4	θ_2	0.2
w_{25}	0.1	θ_3	0.1

(b) Write only the two **main drawbacks** of K-Means clustering algorithm?   [2] (CO5)

Q2. (a) Identify the class of instance (20, 35) using K-NN algorithm, where K = 5, using following dataset instances? What method should be applied to improve the accuracy of KNN?   [4+1] (CO4)

A	B	Class Label
40	20	Red
50	50	Blue
60	90	Blue
10	25	Red
70	70	Blue
60	10	Red
25	80	Blue

(b) Compare the cross-fold validation with boot-strapping? [3] (CO1)

Q3. (a) Consider the problem of predicting whether the university will be closed on a particular day, using following dataset. What would be the effect on the selecting the weekend attribute for the decision tree? (Use Information gain ratio method) [6](CO2)

Snowstorm	Weekend	Official Holiday	University closed
yes	No	No	Yes
Yes	Yes	No	Yes
No	Yes	No	Yes
No	No	Yes	Yes
No	No	No	No
Yes	No	Yes	Yes
No	Yes	Yes	Yes
yes	yes	Yes	Yes

(b) Justify the requirement of logistic regression, if there is already a binary classification technique? [2]CO3

Q4. (a) Write the equation for the quadratic contained optimization problem related to SVM classifier. (Clearly state the each variable of the equations) [2+2] (CO4)

1) Hard-Margin SVM

2) Soft-Margin SVM

(b) Referring the Genetic algorithms, Consider the initial population containing four hypothesis: $CR_1=$ 01101, $CR_2=$11000, $CR_3=$01000, $CR_4=$10011.The x is function for conversion of CR_i to its decimal value e.g. $x(CR_1)=$13.We define a function $f(x)=x^2$.The task is to maximize the value of function $f(x)$.

1) Calculate the fitness function for each hypothesis. [1+2+2+1] (CO5)

2) Calculate the probability of each hypothesis using fitness proportionate selection.

3) Perform crossover of the pairs $CR_1 \leftrightarrow CR_2$, $CR_3 \leftrightarrow CR_4$ using single point crossover with mask 11100.

4) Perform the mutation on new population generated in the step3 above.