# Machine, Data, Learning: Quiz 2 (Set A)

Duration: 45 minutes

1. No clarifications during the quiz.
2. Make *reasonable assumptions* and *clearly state* them to answer *ambiguous* questions.
3. Use the common default interpretation of any concept/method as discussed in class.
4. Show your steps / reasons. Be concise and organized.
5. Floating point answers should be upto 2 significant decimal places.
6. Calculators allowed. Sharing of calculators not allowed.

1) Consider the following data. [15]

| Record# | Colour | Type | Origin |
|---|---|---|---|
| 1 | Red | Sports | Domestic |
| 2 | Red | Sports | Imported |
| 3 | Red | SUV | Domestic |
| 4 | Yellow | Sports | Domestic |
| 5 | Red | Sports | Imported |
| 6 | Yellow | SUV | Imported |
| 7 | Yellow | SUV | Imported |
| 8 | Yellow | SUV | Domestic |
| 9 | Red | SUV | Imported |
| 10 | Red | Sports | |

(i) What is the distance between record 3 and 8 (upto 2 significant decimal places), using the technique discussed in class for categorical variables?

(ii) What are the 3 nearest neighbours of record 10 (if weights of the features (colour, type, origin) are (0.5, 0.3, 0) respectively) _____

(iii) What is the predicted Origin of record 10 (if weights of the features (colour, type, origin) are (0.5, 0.3, 0) respectively) using k-NN algorithm with k=3 _____

(iv) What is the predicted Origin of record 10 (if weights of the features (colour, type, origin) are (0.5, 0.3, 0) respectively) using k-NN algorithm with k=5 _____

2) Data: {(Ram,65,60),(Shyam,60,62),(Gita,58,70),(Mohan,68,70)}. Run 2 iterations of k-means algorithm using euclidean distance and k=2. Choose Shyam and Gita as initial means. The clusters after 2 iterations are: _____ and _____
[10]

3) List the order in which nodes are visited in the tree below for each of the following three search strategies (choosing leftmost branches first in all ambiguous cases). The cost/weight of each edge from a node $i$ to node $j$ is equal to the integer value of $j$. For e.g. the cost of edge to node (6) is 6. [15]

   (a) Breadth-first search
   (b) Depth-limited search with depth limit = 3; Treat root as level 1
   (c) Greedy Best-first search

Also, list the nodes present in the *fringe / open-list* in step 2 of each strategy (i.e. just after the 2nd node is expanded in that strategy). Assume that a good estimate of reaching a goal from each node $i$ is equal to the integer value of $i$. For e.g. the estimate of reaching a goal from node (6) is 6.