Indian Institute of Engineering Science and Technology, Shibpur

B. Tech. Information Technology, 6th Semester Mid Semester Examinations, March 2022

Machine Learning

IT-3205

Full Marks: 30   Answer any five questions   Time: 45 Minutes

1. a) Define a well-posed learning problem. What is the role of the target function in designing a learning system? [3]
   b) When is a hypothesis h_j is more-general-than-or-equal-to h_k? Explain with an example. [3]

2. a) For classification task with missing inputs, explain why the learning algorithm need to learn a set of functions rather than a single classification function. How is this situation handled with the help of probability distribution? [3]
   b) Differentiate between classification and regression. Suppose several instances of a random vector **x** is given to you what type of learning algorithm will be appropriate in such a case? Explain. [3]

3. a) What is meant by an objective function? Explain with examples under what condition you want to i) minimize ii) maximize an objective function. [3]
   b) "Linear models rarely overfit". First explain what is overfitting, then explain why linear models in regression rarely overfit. [3]

4. a) What is logistic regression? What is the optimization criterion in logistic regression? [3]
   b) What is a split in Decision Tree Learning? Why entropy is used in order to evaluate the goodness of a split in Decision Tree Learning? [3]

5. a) In Support Vector Machine, what is meant by margin? Why is it required to optimize the margin? [3]
   b) If the decision boundary is not linear, then how you can derive the non-linear decision boundary? [3]

6. a) Briefly state the distance weighted nearest neighbour algorithm. [3]
   b) What do you understand by the 'curse of dimensionality' in kNN? How can you overcome this problem? [3]

7. a) Differentiate between prior probability and posterior probability used in Bayes learning. [3]
   b) Differentiate between MAP (Maximum A Posteriori) and ML (Maximum Likelihood) hypothesis. [3]