

Siddaganga Institute of Technology. Tumkur  
(An Autonomous institution under Visvesvaraya Technological University, Belgaum)  
Department of Electronics and Communication Engineering  
**Question Bank for Machine Learning (ECE20)**

Unit 1: **Introduction**

- I
1. A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E.  
Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. What would be a reasonable choice for P?  
i) None of these, ii) The probability of it correctly predicting a future date's weather, iii) The process of the algorithm examining a large amount of historical weather data, and iv) The weather prediction task.
  2. Suppose you are working on weather prediction, and you would like to predict whether or not it will 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?  
i) Regression and ii) Classification
  3. Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will declare bankruptcy within the next 7 days (by training on data of similar companies that had previously been at risk of bankruptcy). Would you treat this as a classification or a regression problem?  
i) Classification and ii) Regression
  4. Some of the problems below are best addressed using a supervised learning algorithm, and the others with an unsupervised learning algorithm. Which of the following would you apply supervised learning to? (Select all that apply.) In each case, assume some appropriate dataset is available for your algorithm to learn from.  
i) Take a collection of 1000 essays written on the US Economy, and find a way to automatically group these essays into a small number of groups of essays that are somehow "similar" or "related".  
ii) Examine a large collection of emails that are known to be spam email, to discover if there are sub-types of spam mail.  
iii) Given genetic (DNA) data from a person, predict the odds of him/her developing diabetes over the next 10 years. Examine the statistics of two football teams, and predict which team will win tomorrow's match (given historical data of teams' wins/losses to learn from).
  5. Which of these is a reasonable definition of machine learning?  
i) Machine learning is the field of allowing robots to act intelligently, ii) Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed, iii) Machine learning learns from labeled data and iv) Machine learning is the science of programming computers.
  6. Mention applications of machine learning.
  7. Compare supervised learning and unsupervised learning.

8. Compare classification and regression.
9. Name at least four measures used to evaluate performance in a machine learning system.
10. Mention different learning algorithms used in building a machine learning application.
11. Give three computer applications for which machine learning approaches seem appropriate and three for which they seem inappropriate. Pick applications that are not already mentioned in this chapter, and include a one-sentence justification for each.
12. Pick some learning task. Describe it informally in a paragraph in English. Now describe it by stating as precisely as possible the task, performance measure, and training experience. Finally, propose a target function to be learned and a target representation. Discuss the main tradeoffs you considered in formulating this learning task.
13. In online debate forums, people debate issues, express their preferences, and argue why their viewpoint is right. For example, a debate can be which mobile phone is better: iPhone or Blackberry, or which OS is better: Windows vs. Linux vs. Mac? Given a debate forum, how can you use machine learning to: i) Detect the hot debate topics, ii) Identify the points of contention within the debate, and iii) For a given topic, recognize which stance a person is taking in an online debate posting. For each of the task above, please specify what type of machine learning problem it is (regression, classification, density estimation, etc). Identify what will be the training data, features and labels (if any), and what would be the output of the algorithm.
14. A secure access system is to designed for an institution with about 4000 students and 500 employees on campus. Modeling this as a pattern classification problem and using voice, finger print as metric, Explain the steps involved in designing such a system. Discuss the issues involved.
15. A company plans to design a intelligent car driver assistant using hand gesture and voice as metric. Explain the steps involved in designing such a system under the framework of pattern recognition usign a block diagram. Discuss the issues involved.
16. Assume the probability of a certain disease is 0.01. The probability of test positive given that a person is infected with the disease is 0.95 and the probability of test positive given the person is not infected with the disease is 0.05.(i) Calculate the probability of test positive and ii) Use Bayes Rule to calculate the probability of being infected with the disease given that the test is positive.
17. In online debate forums, people debate issues, express their preferences, and argue why their viewpoint is right. For example, a debate can be which mobile phone is better: iPhone or Blackberry, or which OS is better: Windows vs. Linux vs. Mac? Given a debate forum, how can you use machine learning to: i) Detect the hot debate topics, ii) Identify the points of contention within the debate, and iii) For a given topic, recognize which stance a person is taking in an online debate posting. For each of the task above, please specify what type of machine learning problem it is (regression, classification, density estimation, etc). Identify what will be the training data, features and labels (if any), and what would be the output of the algorithm.

18. In an electronics laboratory, there are identically looking resistors of three makes in  $A_1, A_2$  and  $A_3$  the ratio 4:1:5. It is known that 2% of  $A_1$ , 3% of  $A_2$  and 2% of  $A_3$  are defective. What percentage of resistors in the laboratory are defective? If a resistor picked at random is found to be defective, what is the probability that it is of make i)  $A_1$ , ii)  $A_2$ , and iii)  $A_3$  ?
19. In an electronics laboratory, there are identically looking CMOS FETs of three makes  $C_1, C_2$  and  $C_3$  in the ratio 2:3:4. It is known that 1% of  $C_1$ , 1.5% of  $C_2$  and 2% of  $C_3$  are defective. What percentage of MOSFETs in the laboratory are defective? If a MOSFET picked at random is found to be defective, what is the probability it is of make  $C_2$  ?
20. In a binary communication system, a zero and a one is transmitted with probability 0.6 and 0.4 respectively. Due to error in communication system, a zero becomes a one with a probability 0.1 and a one becomes zero with a probability 0.08. Determine the probability i) of receiving a one and ii) that a one was transmitted when the received message is one.
21. In an automobile industry, there are identically looking vehicles of three makes,  $A_1, A_2$  and  $A_3$  respectively in the ratio 4:1:5. It is known that 2% of  $A_1$ , 5% of  $A_2$  and 4% of  $A_3$  are exported. What percentage of vehicles in the industry are exported? If a vehicle is picked at random is found to be of export quality. What is the probability that it is of make i)  $A_1$ , ii)  $A_2$  and iii)  $A_3$  ?