**Problem 1: Identify the task, performance measure and training experience for the following learning problems**

1.      Diagnosis of Deadly Diseases

2.      Face Recognition and Tagging Features

3.      Detection of Spam Emails

4.      Displaying the Related Advertisements

5.      Shopping Recommendation   - amazon, Netflix

6.      Ranking web pages

7.      Google translator

8.      A checkers learning problem

9.      Handwriting recognition learning problem

10.  A robot driving learning problem

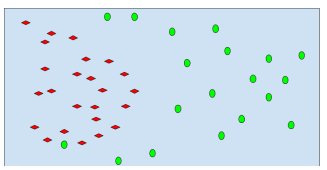
11.  Learning to classify new astronomical structures

**Problem 2:**

**i) Which of the following tasks can be best solved using Clustering, justify your answer**

1. Predicting the amount of rainfall based on various cues (b) Detecting fraudulent credit card transactions (c) Training a robot to solve a maze

**ii) What would be the ideal complexity of the curve which can be used for separating the two classes shown in the image below. justify your answer**



1. Linear (b) Quadratic (c) Cubic

**iii) Map the following examples to one of the below learning methods: Classification, Clustering, Reinforcement Learning, Regression**

* 1. **Movie Recommendation systems**
  2. Predict the age of a person
  3. Predict whether the price of petroleum will increase tomorrow
  4. **Sentiment Analysis**
  5. optical character recognition
  6. face detection
  7. spam filtering
  8. topic spotting: categorize news articles (say) as to whether they are about politics, sports, entertainment, etc.
  9. medical diagnosis
  10. customer segmentation
  11. fraud detection
  12. weather prediction
  13. Learning to ride a bicycle
  14. Group news articles based on text similarity
  15. Make clusters of books on similar topics in a library

iv) I am the marketing consultant of a leading e-commerce website. I have been given a task of making a system that recommends products to users based on their activity on Facebook. I realize that user interests could be highly variable. Hence I decide to a. First, cluster the users into communities of like-minded people and b. Second, train separate models for each community to predict which product category (e.g. electronic gadgets, cosmetics, etc) would be the most relevant to that community. The first task is a/an \_\_\_\_\_\_\_\_\_\_\_\_\_\_ learning problem while the second is a/an \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ problem.

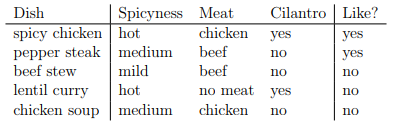
**Problem 3: A computer program made to learn to drive autonomous vehicles, Telsa S Model crashed with truck. Briefly describe any five such failures of Artificial intelligence in Machine learning from 2017 and onwards.**

**Problem 4: Concept Learning**

Bob just met Alice recently at a board game night. After two hours of vivid conversation, Bob invites Alice to have dinner at his apartment the next day. Unfortunately, Alice is a very picky eater and doesn’t like some ingredients and certain combinations of flavors. Bob is unable to ask Alice about her preferences because he is too afraid, but he manages to get hold of Alice’s roommate on his way home. He asks her about the five last dishes that she made for Alice and also whether Alice liked them or not. Because Bob has just finished reading his first introductionary book on machine learning, he decides to apply a more rigorous approach to find out whether Alice will like a recipe. He describes all recipes by the following three criteria:

* Spicyness S: A dish can be either mild, medium or hot.
* Meat M: A recipe can call for no meat, beef or chicken.
* Cilantro C: A binary variable indicating whether cilantro is part of the recipe or not

The next table T shows the information Bob was able to get from Alice’s roommate. The last column indicates whether Alice liked the dish or not.



Stated more formally, Bob wants to learn a function S × M × C → {yes, no} to make his date a success.

1. Define concept learning.
2. Name or label the target concept.
3. Identify and discuss key attributes that distinguish the concept in question from other similar concepts.
4. What is the size of the instance space X?
5. Assuming that the hypothesis space H consists of all possible functions on S ×M × C → {yes, no}, what is the size of H?
6. Initially, Bob thinks that Alice would like any hot dish. Is this hypothesis h consistent with table T above? Why or why not?
7. < Mild ,? , no> whether the hypothesis is true or not.

**Problem 5-Subsetselection**

Filter methods are generally used as a preprocessing step. The selection of features is independent of any machine learning algorithms. Instead, features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. This becomes even more important when the number of features are very large. You need not use every feature at your disposal for creating an algorithm. You can assist your algorithm by feeding in only those features that are really important. Apply Pearson’s co-efficient for the following data set.

|  |  |  |
| --- | --- | --- |
| SUBJECT | AGE X | GLUCOSE LEVEL Y |
| 1 | 43 | 99 |
| 2 | 21 | 65 |
| 3 | 25 | 79 |
| 4 | 42 | 75 |
| 5 | 57 | 87 |
| 6 | 59 | 81 |