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

1. Diagnosis of Deadly Diseases

Task : Identify deadly diseases

Performance Measure : Percentage of correct identification

Training Experience : Patients health record

2.      Face Recognition and Tagging Features

Task : classifying camera images of faces of various people in various poses and tagging them

Performance Measure : Percentage of images recognized and tagged correctly

Training Experience : A database of images with various poses

3.      Detection of Spam Emails

Task : Classification of mails as spam emails

Performance Measure : Percentage of correct spam emails identified out of total received mails

Training Experience : Set of emails with spam and non spam keywords

4.      Displaying the Related Advertisements

Task : Showing related advertisements

Performance Measure : Percentage of related advertisement out of total advertisements

Training Experience : Set of advertisements based on the topics or key words

5.      Shopping Recommendation   - amazon, Netflix

Task : Recommend products based on shopping history

Performance Measure : Percentage of relavent recommendation out of total recommendation

Training Experience : order history

6.      Ranking web pages

7.      Google translator

Task : Translate the given word or sentence to a specific language

Performance Measure : Percentage of correct translation out of total translation

Training Experience : Train on data dictionary

8.      A checkers learning problem

Task : Playing checkers

Performance Measure : Percentage of games won against the opponents

Training Experience : Playing practice games again itself

9.      Handwriting recognition learning problem

Task : Recognizing and classifying hand written words with in images

Performance Measure : Percentage of words correctly classified

Training Experience : A database of handwritten words with classification

10.  A robot driving learning problem

Task : Driving on public 4-lane highways using vision sensors

Performance Measure : Average distance travelled before an error (as judged by human overseer)

Training Experience : A sequence of images and steering commands while observing a human driver

11.  Learning to classify new astronomical structures

Task : Classify new astronomical structures

Performance Measure : Correct classification of celestial objects

Training Experience : Classification of the objects using sky servey

**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

Regression

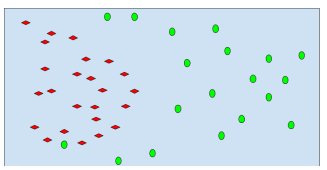
1. Detecting fraudulent credit card transactions

Clustering

1. Training a robot to solve a maze

Re-inforcement learning

**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**

**Clustering**

* 1. Predict the age of a person

**Regression**

* 1. Predict whether the price of petroleum will increase tomorrow

**Classification**

* 1. **Sentiment Analysis**

**Classification**

* 1. optical character recognition

Classification

* 1. face detection

Regression/ Classification

* 1. spam filtering
     1. Classification
  2. topic spotting: categorize news articles (say) as to whether they are about politics, sports, entertainment, etc.

Clustering

* 1. medical diagnosis
     1. Classification
  2. customer segmentation

Clustering

* 1. fraud detection

Classification

* 1. weather prediction
     1. Classification
  2. Learning to ride a bicycle
     1. Reinforcement
  3. Group news articles based on text similarity
     1. Clustering
  4. Make clusters of books on similar topics in a library
     1. Clustering

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.

Unsupervised, Supervised

**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.**

[**https://www.techrepublic.com/article/the-10-biggest-ai-failures-of-2017/**](https://www.techrepublic.com/article/the-10-biggest-ai-failures-of-2017/)

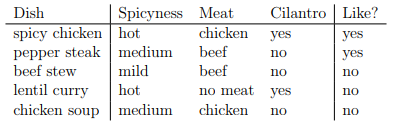
1. **Google Translate shows gender bias in Turkish-English translations**
2. **Facebook chatbots shut down after developing their own language**
3. **Google Allo suggested man in turban emoji as response to a gun emoji**
4. **Face ID beat by a mask**
5. **Alexa brings the party with her in Germany**
6. **Google Home outage causes near 100% failure rate**
7. **Google Home Minis spied on their owners**

**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 |