

Assignment -1

1. Which ONE of the following are regression tasks?

- A) Predict the age of a person**
- B) Predict the country from where the person comes from
- C) Predict whether the price of petroleum will increase tomorrow
- D) Predict whether a document is related to science

Answer: A (Age is a continuous variable but other options are discrete)

2. Which of the following is a supervised learning problem?

- A) Grouping people in a social network.
- B) Predicting credit approval based on historical data**
- C) Predicting rainfall based on historical data**
- D) all of the above

Answer: B and C (A does not have labels to indicate the groups. B and C have the correct answers for the examples in the dataset.

3. Which of the following are classification tasks? (Mark all that apply)

- A) Find the gender of a person by analyzing his writing style**
- B) Predict the price of a house based on floor area, number of rooms etc.
- C) Predict whether there will be abnormally heavy rainfall next year**
- D) Predict the number of copies of a book that will be sold this month

Answer: A, C

In (C), the amount of rain fall is a continuous variable. But, we are predicting whether there will be abnormally heavy rainfall next year or not. So it is a Classification task. Similarly, the number of classes in gender identification (A) is discrete. So, it's a classification task. In other options, the output variable is a continuous class, so these are regression tasks.

4. Which of these are categorical features?

A) Height of a person

B) Price of petroleum

C) Mother tongue of a person

D) Amount of rainfall in a day

Answer: C (Categorical variables represent types of data which may be divided into groups. All other features are continuous)

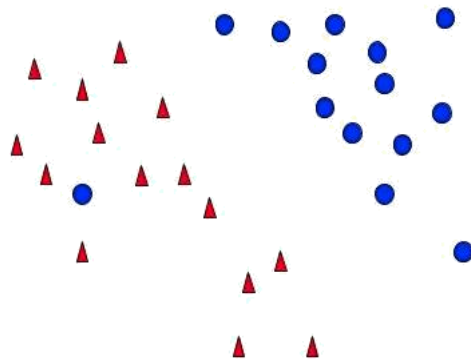
5. What would be the ideal complexity of the curve which can be used for separating the two classes shown in the image below?

A) Linear

B) Quadratic

C) Cubic

D) insufficient data to draw conclusion



Answer: A

(The blue point in the red region is an outlier (most likely noise). The rest of the data is linearly separable.)

6. Occam's razor is an example of:

- (a) Inductive bias
- (b) Preference bias

Answer: A (Prefer simplest hypothesis than complex one)

7. How does generalization performance change with increasing size of training set?

- A) Improves**
- B) Deteriorates
- C) No Change
- D) None

Answer: A (From lecture slides)

8. One of the most common uses of Machine Learning today is in the domain of Robotics. Robotic tasks include a multitude of ML methods tailored towards navigation, robotic control and a number of other tasks. Robotic control includes controlling the actuators available to the robotic system. An example of this is control of a painting arm in automotive industries. The robotic arm must be able to paint every corner in the automotive parts while minimizing the quantity of paint wasted in the process. Which of the following learning paradigms would you select for training such a robotic arm?

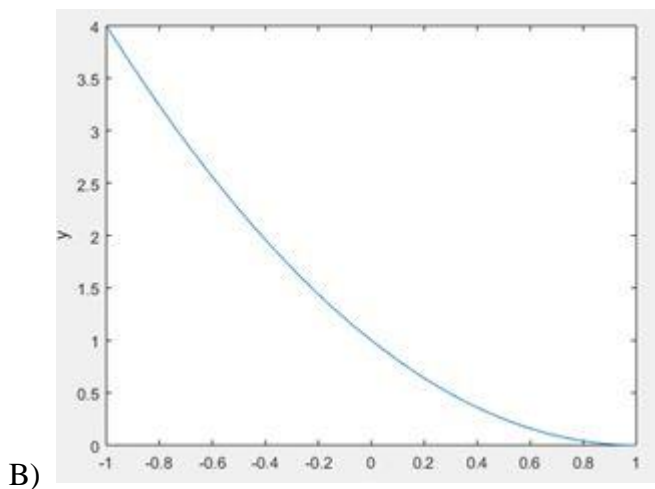
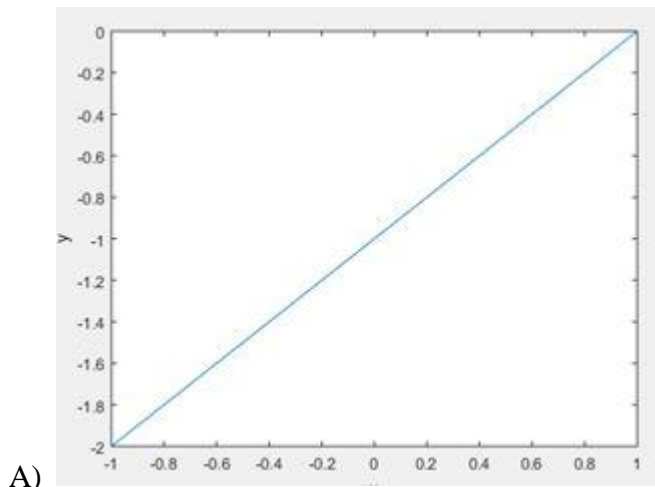
- A) Supervised learning
- B) Unsupervised learning
- C) Combination of supervised and unsupervised learning
- D) Reinforcement learning**

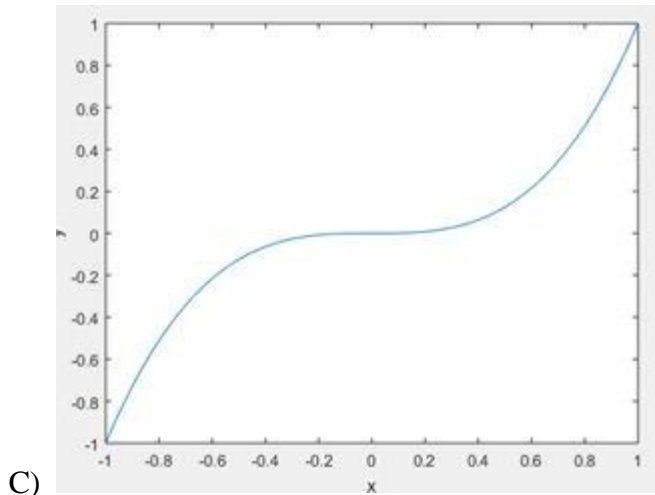
Answer: D (This kind of a learning problem warrants the use of Reinforcement Learning.)

We see that the robotic arm has to cover every corner, i.e. maximize the area covered and all the while minimizing the quantity of paint wasted in the process. One can design a primitive reward signal that takes into account the area covered and paint wasted (normalized to some extent) and use it to train a reinforcement learning agent.

Supervised Learning cannot be used in this setting as there will not be any kind of training data available for the agent. Similarly, Unsupervised Learning cannot be used as it is not a pattern discovery problem.)

9. Choose the function that has the maximum variance:





Answer: C (The function with highest variance has more curved structure)

10. 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.

Choose from the options:

- A) Supervised and unsupervised
- B) Unsupervised and supervised**
- C) Supervised and supervised
- D) Unsupervised and unsupervised

Answer: B (From the definition of supervised and unsupervised learning)