Week 1: Assignment 1: Solutions

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

2. 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

3. Which of the following are examples of unsupervised learning?

A) Group news articles based on text similarity

B) Make clusters of books on similar topics in a library

C) Filter out spam emails

D) Segment online customers into two classes based on their age group – below 25 or above 25
Answer: A, B

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

5. Validation set is used for testing the generalization performance of a learning algorithm:

A) T

B) F
Answer: A

6. Which one of the following functions has the highest bias?

A. Linear model
B. Quadratic model
C. Decision tree
Answer. A

7. The variance in an existing model can be reduced by _____ training data. (Choose the appropriate option for filling the
blank)

A. Decreasing
B. Increasing
Answer. B

8. We always wish to make sure that our model performs well in the real world when presented with cases that it has not encountered before. Hence we make sure the _____ is low. (Choose the appropriate option for filling the blank):

A. Bias
B. Variance
Answer: B

9. 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

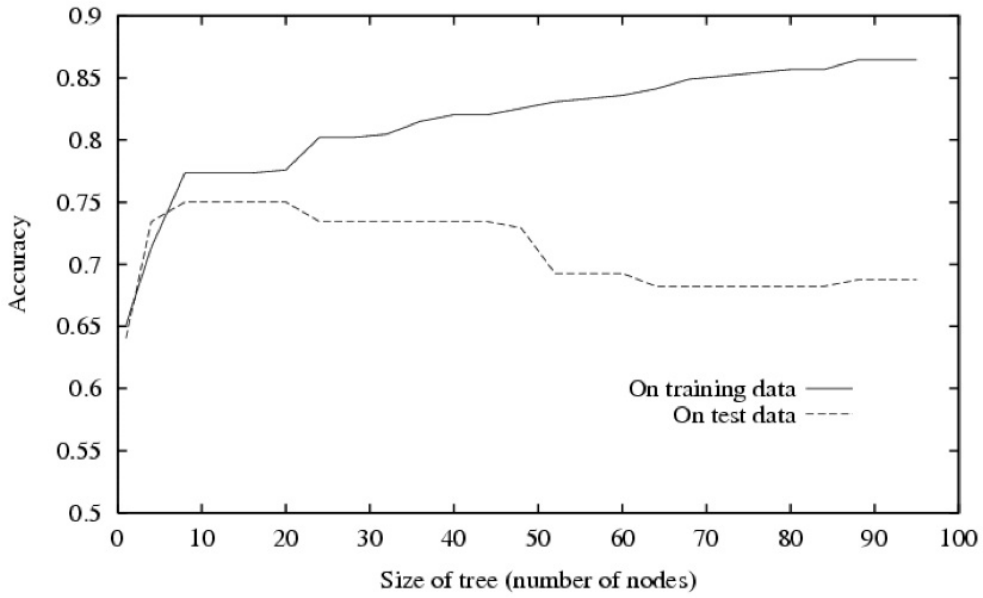
a. First, cluster the users into communities of like-minded people and
b. Second, train separate models for each community to predict which product category (e.g. electronic gadgets, cosmetics, etc) would be the most relevant to that community.

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

10. At what value of number of nodes does overfitting set in?



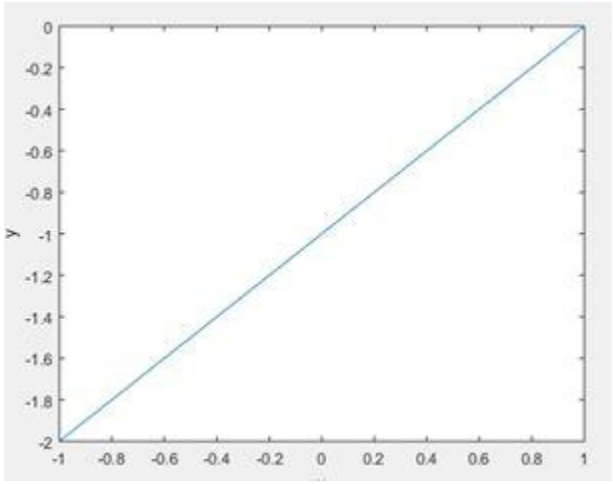
Choose a number from among the following that is the closest to what you expect:

A. 5
B. 20
C. 30

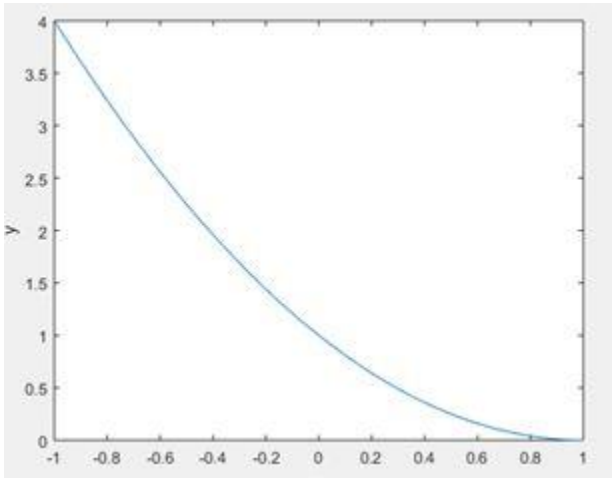
Answer. B

11. Choose the function that has the maximum variance:

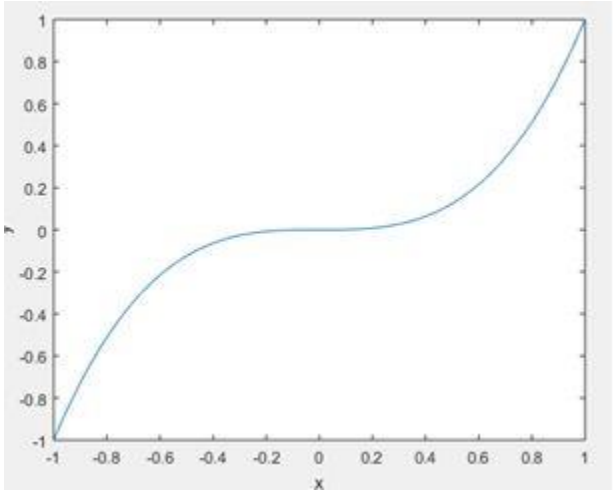
A.


0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-2
y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

B.


4
3.5
3
2.5
2
1.5
1
0.5
0
y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

C.


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
x

Answer. C