Assignment Questions

1. Define machine learning and explain the main types of machine learning approaches.
2. Differentiate between supervised and unsupervised learning. Provide examples of each.
3. What are parametric and non-parametric models in machine learning? Give an example of each.
4. How do parametric models differ from non-parametric models in terms of flexibility and complexity?
5. Explain the bias-variance tradeoff in the context of machine learning. How does it impact model performance?
6. Define underfitting and overfitting. How can these issues be addressed in machine learning models?
7. Discuss the importance of model selection in machine learning. What criteria should be considered when choosing a model?
8. Explain the concept of cost functions in the context of machine learning. Why are they crucial for model training?
9. Learning Theory: Bias/Variance Tradeoff, Union and Chernoff Bounds, VC Dimensions
10. Describe the bias-variance tradeoff and its implications on model generalization.
11. Introduction to Gradient Descent
12. What is gradient descent, and how is it used in machine learning optimization?
13. Explain the role of learning rate in the context of gradient descent. How does it affect convergence and model performance?
14. Define Bayesian concept learning and its applications in machine learning.
15. Explain the fundamental principles of Bayesian Decision Theory.
16. Describe the Naïve Bayes algorithm and its assumptions. When is it particularly useful?
17. Discuss the concept of zero probability in Bayesian models. How does Laplacian correction address this issue?
18. What are Bayesian Belief Networks, and how are they used to model probabilistic relationships in data?
19. Suppose you are working on a dataset related to housing prices, aiming to predict the price of a house based on its size (in square feet). You have collected the following data:

House Size (in square feet)	House Price (in thousands of dollars)
1400	250
1600	300
1700	320
1875	375
1100	200

Assuming a simple linear regression model $y = mx + b$ (where y is the house price and x is the house size), answer the following:

a) Calculate the mean of house sizes ($\bar{x}$) and house prices ($\bar{y}$).

b) Calculate the slope (m) and the y-intercept (b) for the linear regression model.

c) Write the equation of the linear regression model based on the values of m and b obtained.

d) Predict the price of a house with a size of 1600 square feet using the linear regression model.

e) Calculate the coefficient of determination (R^2) to assess the goodness of fit for the model.

20. Solve it

Consider a real estate dataset where you want to predict house prices based on multiple features, including the number of bedrooms (x_1), the number of bathrooms (x_2), and the square footage (x_3). The dataset is as follows:

House Size (sq ft)	Bedrooms	Bathrooms	House Price ($1000)
1800	3	2	320
2000	4	3	400
1600	3	2	280
2400	5	3	500
2200	4	2.5	450

a) Perform multiple linear regression to find the coefficients (b_0, b_1, b_2, b_3) for the model $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$.

b) Write the equation of the multiple regression model based on the coefficients obtained.

c) Use the model to predict the house price for a property with 4 bedrooms, 2 bathrooms, and 2000 square feet.

21. Solve it

Consider a binary classification problem where a machine learning model is designed to identify whether an email is spam (positive class) or not (negative class). The model is tested on a dataset of 200 emails, and the confusion matrix is as follows:

Actual / Predicted	Predicted Negative	Predicted Positive
Actual Negative	120	15
Actual Positive	10	55

a) Calculate the accuracy, precision, recall, and F1-score based on the given confusion matrix.

22. Solve it

Imagine a medical diagnostic test designed to identify a rare disease. The test is applied to a sample of 150 individuals suspected of having the disease. The confusion matrix for the test is as follows:

Actual / Predicted	Predicted Negative	Predicted Positive
Actual Negative	90	5
Actual Positive	10	45

a) Calculate the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) based on the given confusion matrix.

23. Solve

Consider a university admissions scenario where two key factors, academic performance and letters of recommendation, influence the admission decision. The Bayesian network is designed as follows:

- **Node 1:** Academic Performance (AP)
  - Low Academic Performance (AP=L)
  - High Academic Performance (AP=H)
- **Node 2:** Letters of Recommendation (LoR)
  - Weak Letters of Recommendation (LoR=W)
  - Strong Letters of Recommendation (LoR=S)
- **Node 3:** Admission Decision (AD)
  - Successful Admission (AD=S)
  - Unsuccessful Admission (AD=U)

The conditional probability distributions are as follows:

$$P(AP = L) = 0.3$$
$$P(AP = H) = 0.7$$

$$P(LoR = W|AP = L) = 0.8$$
$$P(LoR = W|AP = H) = 0.2$$

$$P(LoR = S|AP = L) = 0.2$$
$$P(LoR = S|AP = H) = 0.8$$

$$P(AD = S|AP = L, LoR = W) = 0.1$$
$$P(AD = S|AP = L, LoR = S) = 0.9$$
$$P(AD = S|AP = H, LoR = W) = 0.8$$
$$P(AD = S|AP = H, LoR = S) = 0.99$$

$$P(AD = U|AP = L, LoR = W) = 0.9$$
$$P(AD = U|AP = L, LoR = S) = 0.1$$
$$P(AD = U|AP = H, LoR = W) = 0.2$$
$$P(AD = U|AP = H, LoR = S) = 0.01$$

a) Calculate the probability that a student with low academic performance and weak letters of recommendation is successful in gaining admission ($P(AD = S|AP = L, LoR = W)$).

b) Determine the overall probability distribution for the entire Bayesian network, considering all possible combinations of academic performance, letters of recommendation, and admission decisions.

24. Solve it

Consider a Bayesian Belief Network (BBN) designed to diagnose a medical condition based on two symptoms: fever and cough. The network also incorporates the patient's history of exposure to a particular virus.

- **Node 1:** Fever (F)
  - Low Fever (F=L)
  - High Fever (F=H)
- **Node 2:** Cough (C)
  - Mild Cough (C=M)
  - Severe Cough (C=S)
- **Node 3:** Virus Exposure (VE)
  - No Exposure (VE=N)
  - Exposure (VE=E)
- **Node 4:** Diagnosis (D)
  - Healthy (D=H)
  - Infected (D=I)

The conditional probability distributions are as follows:

$$P(F = L) = 0.6$$
$$P(F = H) = 0.4$$

$$P(C = M|F = L) = 0.8$$
$$P(C = S|F = L) = 0.2$$
$$P(C = M|F = H) = 0.3$$
$$P(C = S|F = H) = 0.7$$

$$P(VE = N) = 0.9$$
$$P(VE = E) = 0.1$$

$$P(D = H|F = L, C = M, VE = N) = 0.95$$
$$P(D = I|F = L, C = M, VE = N) = 0.05$$
$$P(D = H|F = H, C = S, VE = E) = 0.1$$
$$P(D = I|F = H, C = S, VE = E) = 0.9$$

a) Calculate the probability that a patient with low fever, mild cough, and no history of virus exposure is diagnosed as healthy ($P(D = H|F = L, C = M, VE = N)$).

b) Determine the overall probability distribution for the entire Bayesian Belief Network, considering all possible combinations of fever, cough, virus exposure, and diagnosis.

25. Naïve Base classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

A: attributes

M: mammals

N: non-mammals

$$P(A|M)=\frac{6}{7}\times\frac{6}{7}\times\frac{2}{7}\times\frac{2}{7}=0.06$$

$$P(A|N)=\frac{1}{13}\times\frac{10}{13}\times\frac{3}{13}\times\frac{4}{13}=0.0042$$

$$P(A|M)P(M)=0.06\times\frac{7}{20}=0.021$$

$$P(A|N)P(N)=0.004\times\frac{13}{20}=0.0027$$

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

26. Naïve base classifier

Consider the given Dataset ,Apply Naive Baye's Algorithm and Predict that if a fruit has the following properties then which type of the fruit it is
Fruit = {Yellow , Sweet ,long}

Frequency Table:

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

27. Naïve base classifier

Consider the training data in the following table where Play is a class attribute. In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No".

Humidity	Sunny	Wind	Play
L	N	S	No
H	N	W	Yes
H	Y	S	Yes
H	N	W	Yes
L	Y	S	No

What is class label for the following day (Humidity=L, Sunny=N, Wind=W), according to naïve Bayesian classification.

27. Linear Regression

Suppose you have a dataset with the following pairs of house sizes (in square feet) and corresponding prices (in thousands of dollars):

House Size (sq ft)	House Price ($1000)
1400	250
1600	300
1700	320
1875	375
1100	200

You want to fit a simple linear regression model $y = mx + b$ to predict house prices based on size. The initial values for the slope (m) and y-intercept (b) are set to 0. Use the Mean Squared Error (MSE) as the cost function.

The update rules for gradient descent are as follows:

$$m_{\text{new}} = m_{\text{old}} - \alpha \frac{1}{N} \sum_{i=1}^{N} 2x_i(y_i - (mx_i + b_{\text{old}}))$$
$$b_{\text{new}} = b_{\text{old}} - \alpha \frac{1}{N} \sum_{i=1}^{N} 2(y_i - (mx_i + b_{\text{old}}))$$

where N is the number of data points, α is the learning rate, x_i is the house size, and y_i is the corresponding house price.

Initial Values:

$m_{\text{old}} = 0$

$b_{\text{old}} = 0$

$\alpha = 0.01$

Perform one iteration of gradient descent and calculate the updated values for m and b. Repeat this process for a total of three iterations.