AMERICAN INTERNATIONAL UNIVERSITY–BANGLADESH (AIUB)

Faculty of Science and Technology

Fall 2024-2025

Section: A

Course: Machine Learning

MID TERM ASSIGNMENT

Supervised By

DR. MD. SAEF ULLAH MIAH

Prepared By

Name : MD Abdul Aziz

ID :   22-47013-1

Section :   A

Course: Machine Learning

QUIZ-1

Question:

1. What is the primary Goal of machine learning?
2. Define Supervised Learning and give an example.
3. Why is gradient descent important in linear regression?
4. Explain the difference between passive and active learning in machine learning.
5. Describe how logistic regression can be used as a binary classifier and the role of the sigmoid function.

Correct Answer	Answer Given by me
1. The primary goal of machine learning is to teach computers to learn from data so they can make decisions or predictions without being directly programmed. It trains a machine for specific tasks by giving some of the earlier output as the input so that the machine can observe and analyze those old cases and performs more efficiently in future.	1. The primary goal of machine learning is to train a machine for specific tasks by giving some of the earlier output as the input so that the machine can observe and analyze those old cases and performs more efficiently in future.
2. Supervised machine learning is a fundamental approach for machine learning and artificial intelligence. It involves training a model using labeled data, where each input comes with a corresponding correct output.	2. Supervised Learning: Supervised learning is a supervised machine learning which models linear relationship between an independent variable and a dependent variable and creates

supervised machine learning involves training a model on labeled data to learn patterns and relationships, which it then uses to make accurate predictions on new data. Spam detection is an example of a supervised learning model. Using supervised classification algorithms, organizations can train databases to recognize patterns or anomalies in new data to organize spam and non-spam-related correspondence effectively.
 	
<u>Reference:</u>
<u>https://www.geeksforgeeks.org/supervised-machine-learning/</u>
	a fitting straight line. The fitting Straight line indicates the maximum minimization of error in the model.
	Supervised learning is two types:
	   1. Classification.
	   2. Regression.

3. Gradient descent is important in linear regression because it helps find the best-fit line by minimizing the error (cost function) between the predicted values and the actual values. A linear regression model can be trained using the gradient descent optimization algorithm, which iteratively adjusts the model's parameters to minimize the mean squared error (MSE) on the training dataset. <u>Reference:</u> <u>https://www.geeksforgeeks.org/gradient-descent-in-linear-regression/</u>	3. Gradiant descent is important in linear regression because it gives accurate results in each iteration. We know $J = \frac{1}{n}\sum (\hat{y}_i - y_i)^2$; we can replace $\hat{y}_i$ by $\theta_1 - \theta_2 x_i$. After that we use gradient in each iteration so that we can work on the smallest part of the equation in order to get an accurate output after each iteration.

4. In passive learning, the machine is given a fixed set of data and uses it to build a model or make predictions. The process is : Data → Learner → Model. In active learning, the machine can ask questions or request specific data to improve its learning.	4. <u>Passive Machine Learning:</u> Most of the machine learning models uses passive learning method. In this method , machine firstly analyze the data then it observes the patterns and characteristics to improve its performance and according to that it gives an output.
5. Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is commonly used as a binary classifier to predict one of two possible outcomes, such as "yes" or "no," "1" or "0," or "true" or "false." It works by estimating the probability that a given input belongs to a specific class. Reference: https://www.geeksforgeeks.org/understanding-logistic-regression/#how-does-logistic-regression-work	5. Logistic regression is different from linear regression. Because in linear regression we often get an output but logistic regression doesn't give us a direct output instead it gives us a logic or function to work on. It gives us a sigmoid function. The job of sigmoid function is to convert the initial output to 0 or 1. This is why logistic regression can be used as binary classifier.

Quiz – 2

Question :

1. A new customer participated in the satisfaction survey with following details :

Age : 33 , Gender Female , Product Quality Rating : 9

Using the trained model , predict whether this customer is satisfied or Not Satisfied . Provide a step by step explanation of your prediction , including the probabilities involved and the reasoning behind your decision using Naïve Bayes.

Age	Gender	Product Quality (1-10)	Class (satisfied/Not Satisfied)
28	Male	8	Satisfied
35	Female	7	Satisfied
42	Male	6	Not Satisfied
29	Male	9	Satisfied
36	Female	7	Not Satisfied
41	Male	8	Satisfied
27	Female	6	Not Satisfied
34	Male	9	Satisfied
31	Female	7	Not Satisfied
37	Male	8	Satisfied
25	Female	6	Not Satisfied
32	Male	9	Satisfied
39	Female	7	Not Satisfied
26	Male	8	Satisfied
33	Female	6	Not Satisfied
40	Male	9	Satisfied
28	Female	7	Not Satisfied
35	Male	8	Satisfied
30	Female	6	Not Satisfied
38	Male	9	Satisfied
27	Female	7	Not Satisfied
36	Male	8	Satisfied
31	Female	6	Not Satisfied
39	Male	9	Satisfied
29	Female	7	Not Satisfied
37	Male	8	Satisfied
32	Female	6	Not Satisfied
45	Male	9	Satisfied
33	Female	7	Not Satisfied
46	Male	8	Satisfied

Answer to The Question No -1

Correct Answer	Answer Given by me
Step 1: Prior class : p(satisfied)= 16/30 = 0.53 p(Not Satisfied) = 14/30 = 0.47 Age: (Age :33 | satisfied) = 0/16 (Age :33 | Not satisfied) = 2/14 Gender : (Female | satisfied) = 1/16 (Female |Not satisfied) = 13/14 Product Quality Rating: (Product Q:9 | satisfied) = 7/16 (Product Q:9 | Not satisfied) = 0/14 p(satisfied| Age :33, Female, Product Q:9) =(0/16) * (1/16) * (8/16) * 0.53 = 0 (Zero frequency problem) Zero frequency problem occurs . so , Now we have to apply Laplace Smoothing to solve the problem. p(satisfied)= (16+1)/(30+2) = 0.53 (Age :33 | satisfied) =(0+1)/(16+20)=0.02 (Female | satisfied) = (1+1)/(16+2)=0.11 (Product Q:9 | satisfied) = (7+1)/(16+4)=0.4 p(satisfied| Age :33, Female, Product Q:9) =0.53 * 0.02 *0.11*0.4 = 0.0004664	Step 1: Prior class : p(satisfied)= 16/30 = 0.53 p(Not Satisfied) = 14/30 = 0.47 features likelihood (Step - 2) : (Age :33 | satisfied) = 0/16 (Female | satisfied) = 1/16 (Product Q:9 | satisfied) = 8/16 (Age :33 | Not satisfied) = 2/14 (Female |Not satisfied) = 13/14 (Product Q:9 | Not satisfied) = 0/14 p(satisfied| Age :33, Female, Product Q:9) =(0/16) * (1/16) * (8/16) * 0.53 = 0 p(Not satisfied| Age :33, Female, Product Q:9) =(2/14) * (13/14) * (0/14) * 0.47 =0 The following result can't be 0. So, (Age :33 | satisfied) =(0+1)/(16+2) (Female | satisfied) = (1+1)/(16+2) (Product Q:9 | satisfied) = (8+1)/(16+2) (Age :33 | Not satisfied) = (2+1)/(14+2) (Female |Not satisfied) = (13+1)/(14+2) (Product Q:9 | Not satisfied) = (0+1)/(14+2) p(satisfied| Age :33, Female, Product Q:9) =(1/18) * (2/18) * (9/18) * 0.53 = 0.001636

P(Not Satisfied) = (14+1)/(30+2) = 0.46875
(Age :33 | Not satisfied) = (2+1)/(14+20)=0.08
(Female |Not satisfied) = (13+1)/(14+2)=0.875
(Product Q:9 | Not satisfied) = (0+1)/(14+4)=0.05

p(Not satisfied| Age :33, Female, Product Q:9)
=0.46875 * 0.08 * 0.875 * 0.05
=0.001645

According to the prediction of Naïve Bayes The customer is not satisfied because
p(Not satisfied| Age :33, Female, Product Q:9) is Higher.

So, The new Customer is Not satisfied

p(Not satisfied| Age :33, Female, Product Q:9)
=(3/16) * (4/16) * (1/16) * 0.47
=0.00138

According to the prediction of Naïve Bayes The customer is not satisfied because
p(Not satisfied| Age :33, Female, Product Q:9) is Higher.

So, The new Customer is Not satisfied.

Question :

2. Your task is to classify a new fruit based on its weight and sweetness level into one of two classes : "Apple" or "Orange". Use the K-Nearest Neighbors(KNN) algorithm **with K=3 and K=5**. Predict the pairs where **"?"** marks are put.

Weight (grams)	Sweetness level (1-10)	Fruit Type
140	8	Apple
160	7	Apple
135	9	Apple
145	6	Apple
150	8	?
155	7	Orange
130	9	Orange
170	5	Orange
165	6	Orange
180	4	Orange

Answer To The Question No -2

Correct Answer	Answer Given by me
$X_1 = (140, 8)$	$X_1 = (140, 8)$
$X_2 = (160, 7)$	$X_2 = (160, 7)$
$X_3 = (135, 9)$	$X_3 = (135, 9)$
$X_4 = (145, 6)$	$X_4 = (145, 6)$
$X_5 = (155, 7)$	$X_5 = (155, 7)$
$X_6 = (130, 9)$	$X_6 = (130, 9)$
$X_7 = (170, 5)$	$X_7 = (170, 5)$
$X_8 = (165, 6)$	$X_8 = (165, 6)$
$X_9 = (180, 4)$	$X_9 = (180, 4)$
$X_? = (150, 8)$	$X_? = (150, 8)$
Distance from $X_?$:	Distance from $X_?$:
To $X_1 = \sqrt{(150-140)^2+(8-8)^2} = 10$	To $X_1 = \sqrt{(150-140)^2+(8-8)^2} = 10$
To $X_2 = \sqrt{(150-160)^2+(8-7)^2} = 10.05$	To $X_2 = \sqrt{(150-160)^2+(8-7)^2} = 10.05$
To $X_3 = \sqrt{(150-135)^2+(8-9)^2} = 15.03$	To $X_3 = \sqrt{(150-135)^2+(8-9)^2} = 15.03$
To $X_4 = \sqrt{(150-145)^2+(8-6)^2} = 5.39$	To $X_4 = \sqrt{(150-145)^2+(8-6)^2} = 5.39$
To $X_5 = \sqrt{(150-155)^2+(8-7)^2} = 5.099$	To $X_5 = \sqrt{(150-155)^2+(8-7)^2} = 5.099$
To $X_6 = \sqrt{(150-130)^2+(8-9)^2} = 20.02$	To $X_6 = \sqrt{(150-130)^2+(8-9)^2} = 20.02$
To $X_7 = \sqrt{(150-170)^2+(8-5)^2} = 20.22$	To $X_7 = \sqrt{(150-170)^2+(8-5)^2} = 20.22$
To $X_8 = \sqrt{(150-165)^2+(8-6)^2} = 15.13$	To $X_8 = \sqrt{(150-165)^2+(8-6)^2} = 15.13$
To $X_9 = \sqrt{(150-180)^2+(8-4)^2} = 30.27$	To $X_9 = \sqrt{(150-180)^2+(8-4)^2} = 30.27$
When K=3 :	When K=3 :
When K=3 , the three nearest neighbors of $X_?$ are X_5 , X_4, X_1.	When K=3 , the three nearest neighbors of $X_?$ are X_5 , X_4, X_1.
The fruit type of X_5 is Orange	The fruit type of X_5 is Orange
The fruit type of X_4 is Apple	The fruit type of X_4 is Apple
The fruit type of X_1 is Apple	The fruit type of X_1 is Apple

So, The occurrence of orange fruit type is 1 and the occurrence of Apple fruit type is 2. The majority occurrence is of Apple.
So, with K=3 , KNN predicts that the **"?"** fruit type is **Apple .**

When K=5 :
When K=5 , the five nearest neighbors of $X_?$ are X_5 , X_4, X_1 , X_2 , X_3

The fruit type of X_5 is Orange
The fruit type of X_4 is Apple
The fruit type of X_1 is Apple
The fruit type of X_2 is Apple
The fruit type of X_3 is Apple

So, The occurrence of orange fruit type is 1 and the occurrence of Apple fruit type is 4. The majority occurrence is of Apple.
So, with K=5 , KNN predicts that the **"?"** fruit type is **Apple .**

Quiz -3

Question:

1. Explain how the ID3 algorithm uses entropy and information gain to build a decision tree. Provide an example to illustrate your explanation.

2. Differentiate between pre-pruning and post-pruning in decision trees. How do these techniques prevent overfitting?

3. How does boosting sequentially improve weak learners , and why is this approach effective for reducing bias compared to bagging?

4.Why is random feature selection within Random Forests essential for reducing overfitting? How does it enhance model diversity?

5. Discuss the significance of precision, recall and the ROC curve in evaluating classification models. How do these metrices complement each other in assessing performance?

Correct Answer	Answer Given by me
1. The ID3 algorithm constructs a decision tree by using entropy and information gain to determine the best attribute to split the data at each step. Entropy is a measure of the impurity or disorder in a dataset, with 0 representing a completely pure dataset and 1 representing maximum uncertainty in a binary classification problem. Information gain quantifies the reduction in entropy achieved when the dataset is split on a particular attribute. At each step, the algorithm calculates the entropy of the dataset and then computes the information gain for each attribute. The attribute with the highest information gain is selected for the split. This process is repeated recursively, partitioning the dataset into subsets until all subsets are pure (i.e., they contain data points from a single class) or no more attributes are available for splitting. For example, in a dataset predicting whether to play tennis based on attributes like "Outlook," "Temperature," "Humidity," and "Wind," the algorithm starts by calculating the entropy of the dataset and the information gain for each attribute. If "Outlook" has the highest information gain, the dataset is split based on its values	Couldn't Attend the 3rd Quiz

(e.g., "Sunny," "Overcast," "Rain"), and the process continues for each subset. This results in a decision tree that efficiently classifies data by progressively narrowing down the options at each node.

Reference : https://www.saedsayad.com/decision_tree.htm

2. Pre-Pruning occurs while the tree is still growing. it essentially put limits on how far the tree is allowed to go before it stops. It is used for keeping the tree from becoming overly complex by cutting off branches before they get too big.
Post-Pruning, on the other hand, allows the tree to grow as much as it wants. You only come in and start trimming after the tree is fully grown. Pre-pruning tends to result in faster training times. post-pruning offers more optimized results because it lets the tree grow fully before deciding what's worth keeping.

Overfitting Prevention By Pre-Pruning:
pre-pruning helps reduce overfitting by keeping the decision tree simple right from the start. Overfitting happens when the tree becomes too detailed and starts learning random patterns or noise in the training data instead of focusing on the main trends. By stopping the tree from growing too much, pre-pruning prevents it from making unnecessary splits based on small or

Couldn't Attend the 3rd Quiz

irrelevant differences in the data. pre-pruning simplifies the tree, making it better at generalizing and less likely to fit the noise in the data.

<u>Overfitting Prevention By Post-Pruning:</u>
post-pruning helps find the right balance between underfitting and overfitting. Underfitting happens when the model is too simple to capture important patterns in the data, while overfitting occurs when the model becomes too complex and starts capturing noise or irrelevant details. With post-pruning, the tree is allowed to grow fully at first, so it doesn't risk missing any important patterns (avoiding underfitting). Afterward, unnecessary parts of the tree are removed based on evaluation, reducing its complexity and preventing it from overfitting.

Reference :
https://tinyurl.com/bddzhhzs

3. Boosting is an ensemble learning technique, but it aims to improve the performance of weak learners by combining them in a sequential manner. The core idea behind boosting is to give more weight to misclassified instances during the training process, enabling subsequent learners to focus on the mistakes made by their predecessors. In each iteration, a new weak learner is trained on the data, and the weights of misclassified instances are

Couldn't Attend the 3rd Quiz

increased. This allows the subsequent learner to pay more attention to the previously misclassified examples.

Boosting Approach is more effective for reducing bias compared to bagging because bagging Primarily reduces variance by averaging predictions from multiple models, making it effective for models with high variance but boosting Addresses both bias and variance, with a focus on reducing bias by sequentially correcting mistakes made by weak learners.

Reference :
https://medium.com/@roshmitadey/bagging-v-s-boosting-be765c970fd1

4. Random feature selection in Random Forests is crucial for reducing overfitting and enhancing model diversity. By considering only a random subset of features at each split, Random Forests prevent individual trees from becoming overly complex and sensitive to noise in the training data. This randomness ensures that each tree captures different patterns, leading to a more robust and generalized model. This approach not only reduces overfitting but also improves the model's generalization capability.

Random feature selection in Random Forests enhances model diversity by ensuring that each decision tree is trained on a different subset of features. This

Couldn't Attend the 3rd Quiz

randomness prevents individual trees from focusing too heavily on specific features, which could lead to overfitting. As a result, the ensemble of trees captures a broader range of patterns in the data, improving the model's generalization capability.

Reference :
https://www.geeksforgeeks.org/why-do-we-pick-random-features-in-random-forest/

https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

5. precision, recall, and the ROC curve are essential tools for evaluating classification models. Precision and recall offer detailed insights into the model's performance concerning the positive class, while the ROC curve provides an overall view across all thresholds. Together, they enable a comprehensive assessment of a model's effectiveness, allowing practitioners to make informed decisions based on the specific requirements of their application.

Precision, recall, and the Receiver Operating Characteristic (ROC) curve are essential metrics for evaluating classification models, each providing unique insights into performance. Precision assesses the accuracy of positive predictions, recall measures the model's ability to identify all positive instances, and the ROC curve evaluates the trade-off

Couldn't Attend the 3rd Quiz

between true positive and false positive rates across various thresholds. These metrics complement each other by offering a comprehensive view of a model's performance, especially in scenarios with imbalanced datasets. For instance, while the ROC curve may present an overly optimistic view in such cases, precision-recall curves focus on the performance with respect to the positive class, providing a more informative picture.

Reference :
1. https://medium.com/@juanc.olamendy/choosing-the-right-metrics-recall-precision-pr-curve-and-roc-curve-explained-682259961cbe
2. https://ftp.cs.wisc.edu/machine-learning/shavlik-group/davis.icml06.pdf