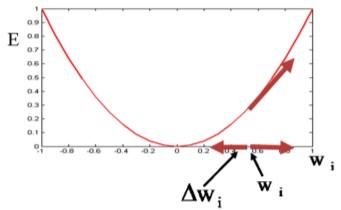
1. Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram.

**Answer:**

Training can be posed as an optimization problem, in which the goal is to optimize a function (usually to minimize a cost function E) with respect to a number of free variables, usually weights wi. The gradient decent algorithm begins from an initialization of the weights (e.g., a random initialization) and in an iterative procedure updates the weights wi by a quantity Δwi, where Δwi = –α\*(∂E / ∂wi) and (∂E / ∂wi) is the gradient of the cost function with respect to the

weights, while α is a constant which takes small values in order to keep the updates low and avoid oscillations.



1. What are the four types of Machine Learning? Explain each type and give suitable examples.

**Answer:**

**Supervised learning:** When an algorithm learns from example data (labels) already provided by user. E.g.: We label a given fruit as “Apple” based on its shape, size, colour, etc. We label a given fruit as “Banana” saying it’s long, thin and yellow. The algorithm checks a new fruit for the above properties and classifies it as “Apple” or “Banana”.

**Unsupervised learning:** When an algorithm learns without any answers (labels) and determines the data patterns on its own. E.g.: A bunch of cat and dog pictures are fed to the algorithm without telling it which pictures belong to which animal. The algorithm makes predictions based on factors like shape, size, colour, height, texture, etc and categorizes the pictures into “0” (cats) and “1” (dogs). The algorithm checks the new picture for the above properties and puts it into either the cats or dogs category.

**Reinforcement learning:** When an algorithm takes suitable action to maximize reward (best possible path) in a particular situation. E.g.: The goal of the robot is to get the reward (diamond) and avoid the obstacles (fire). The robot learns by trying all the possible paths and then choosing the path which gives it the reward with the least hurdles. Each right step will give the robot a reward and each wrong step will subtract the reward of the robot.

**Semi-supervised learning:** When an algorithm is trained upon a combination of labelled and unlabelled data. E.g.: Speech Analysis (Labelling audio files)

1. Write down 2 advantages and 2 disadvantages of simple linear regression and multiple linear regression.

**Answer:**

**Advantages of Simple linear Regression:**

1. Easy to understand because they are built upon basic statistical principles such as correlation and least square error

2. Easy to implement, it’s easier to implement and gives a very satisfactory result

**Disadvantages of Simple linear Regression:**

1. Correlation doesn’t mean causation: Even if there is a connection between 2 variables doesn’t necessarily mean one should be the cause of other.

2. As the no of variable increases the reliability of the regression decreases, regression models works better if you have small number of variables

**Advantages of Multi linear Regression:**

1. Fits linearly separable datasets almost perfectly and is often used to find the nature of the relationship between variables.

2. Overfitting can be reduced by regularization [Overfitting is a situation that arises when a machine learning model fits a dataset very closely and hence captures the noisy data as well.]

**Disadvantages of Multi linear Regression:**

1.Prone to underfitting that arises when a machine learning model fails to capture the data properly. This typically occurs when the hypothesis function cannot fit the data well.

2.Sensitive to outliers, where outliers of a data set are anomalies or extreme values that deviate from the other data points of the distribution. Data outliers can damage the performance

1. Describe the following terms in detail with examples:
   * + - 1. Classification
         2. Regression
         3. Clustering

**Answer:**

**Classification:** A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”. Classification models include logistic regression, decision tree, Naive Bayes, etc. Given one or more inputs a classification model will try to predict the value of one or more outcomes. E.g., Filtering emails as “spam” or “not spam”, Transaction data as “fraudulent”, or “authorized”.

**Regression:** It is a representation of a linear equation that combines a specific set of input values (x) to predict the output value (y). The two types of regression are: Simple Linear Regression (one input) and Multiple Linear Regression (multiple inputs). E.g., Predicting the weather, predicting the house price sales, etc.

**Clustering:** It is an unsupervised learning method. It is the task of dividing the data points into a number of groups such that data points in the same groups are more similar to each other and dissimilar to data points in other groups. E.g., For clustering shares in marketing, books in libraries and proteins in Biology.

1. Write down the primary equations with numerical examples of the following models:
   * + - 1. Multiple Linear Regression
         2. Logistic Regression

**Answer:**

**Multiple Linear Regression:** *y = B0 + (B1\*x1 + B2\*x2 + B3\*x3 + …… + Bnxn).* E.g., If there are 2 features and the values are assumed to be B0 = 1, B1= 0.5, B2 = 0.2. Then the equation becomes – y = 1 + 0.5x1 + 0.2x2. If x1 = 2 and x2 = 5, then for those values:

y = 1 + 0.5(2) + 0.2(5)

y = 1 + 1 + 1 = 3

Hence, Y = 3.

**Logistic Regression**: *ln(p/1-p) = b0 + b1\*x*. E.g., X = 150cm (height of a person). The categories are - 0 if p(male) < 0.5 and 1 if p(male) >= 0.5.

y = e^(b0 + b1\*X) / (1 + e^(b0 + b1\*X))

y = e^(-100 + 0.6\*150) / (1 + e^(-100 + 0.6\*X))

y = 0.0000453978687

Hence, X belongs to female category.

1. Draw a Decision tree with the following conditions: You should have only 3 input features (they can be any) and your tree should have 4 output classes which are - Car, Auto, Bus, Tractor.

**Answer:**

My input features are: 3-Wheeler, Heavy Duty and Carries passengers.

**3-Wheeler**

/ \

Yes No

/ \

**Auto** **Heavy Duty**

/ \

Yes No

/ \

**Carries passengers** **Car**

/ \

Yes No

/ \

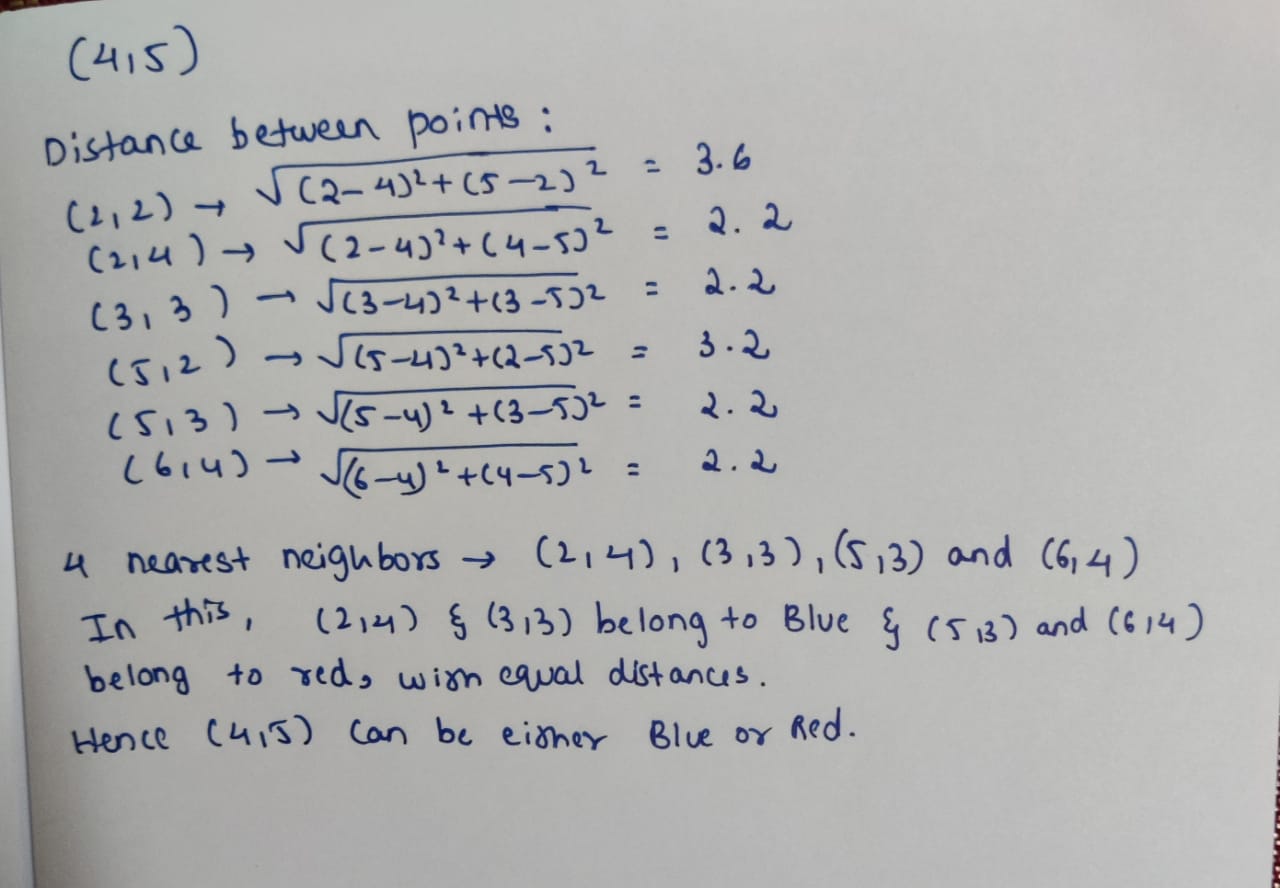
**Bus Tractor**

1. Consider the following data:

|  |  |
| --- | --- |
| Blue | Red |
| (2,2) | (5,2) |
| (2,4) | (5,3) |
| (3,3) | (6,4) |

Using KNN classification, for K=4, determine which category (blue or red), will (4,5) belong to. Show all the steps and calculations.

**Answer:**

****

1. Differentiate between hierarchical based methods and portioning methods in clustering and give examples.

**Answer:**

**Hierarchical Based Methods**: It is an iterative process where, it identifies two clusters that are closest together, and merges them (continues till all the clusters are merged together) E.g.: Agglomerative Clustering

**Partitioning Methods**: It partitions the objects into k clusters and each partition forms one cluster (for similarity measures) E.g., K-Means Clustering

1. Explain the six steps involved in K-Means Clustering and write a short note on the Elbow method.

**Answer:**

**Steps:**

Step-1: Select the number K to decide the number of clusters

Step-2: Select random K points or centroids

Step-3: Assign each data point to their closest centroid

Step-4: Calculate the mean and place a new centroid of each cluster

Step-5: Reassign each datapoint to the new closest centroid

Step-6: Continue 4 and 5 until the model is ready

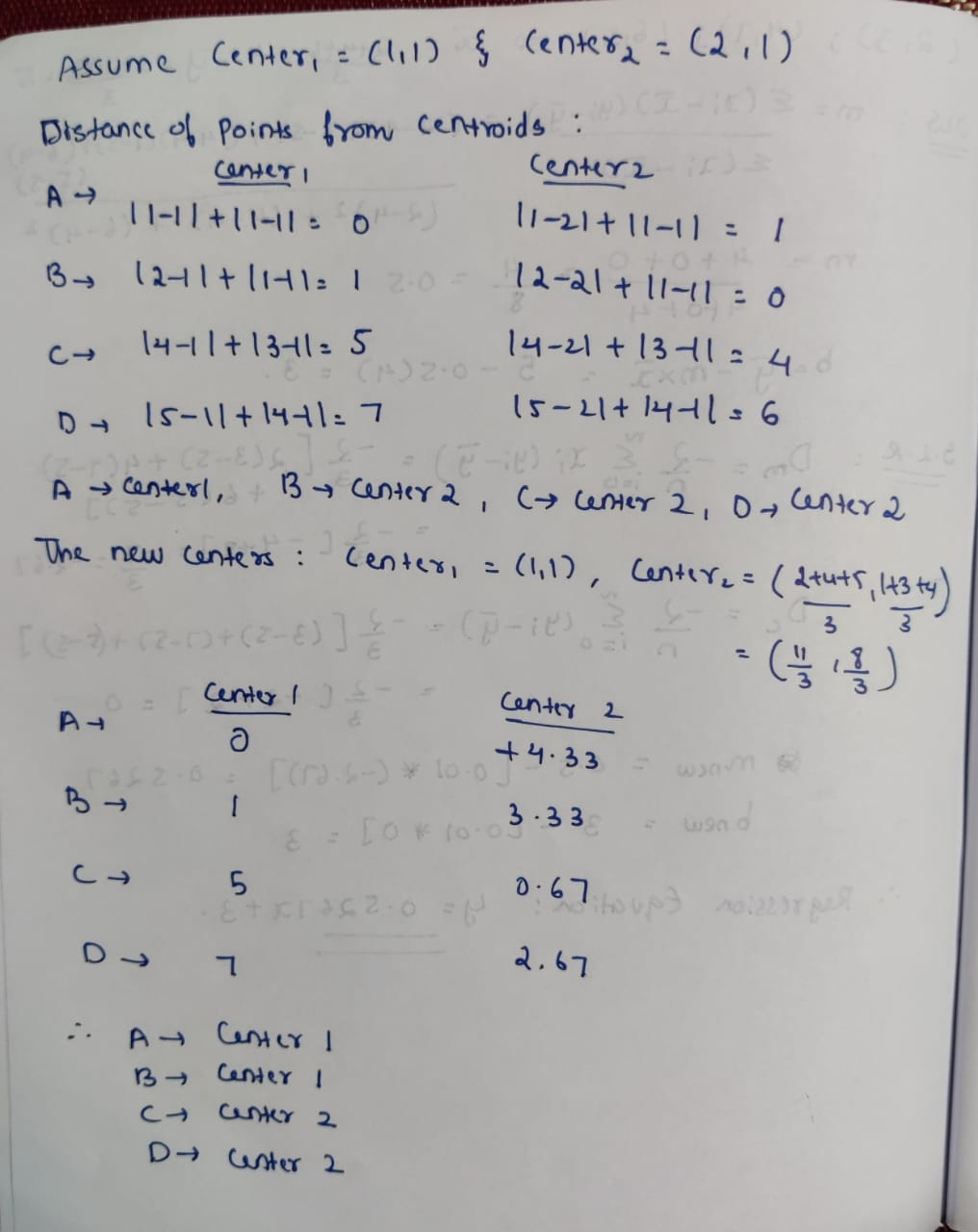
**Elbow Method:** It determines the optimal value of K. It executes the K-means clustering on a given dataset for different K values (ranges from 1-10). For each value of K, calculates the WCSS value. Plots a curve between calculated WCSS values and the number of clusters K. The sharp point of bend or a point of the plot which looks like an arm, is considered as the best value of K.

1. Consider the following data:

|  |  |
| --- | --- |
| Medicine | Data Point |
| A | (1,1) |
| B | (2,1) |
| C | (4,3) |
| D | (5,4) |

Perform K Means clustering on the given data by grouping them into K=2 clusters. Determine which cluster (1 or 2) do Medicine A, B, C and D belong to. Show all the steps and calculations.

**Answer:**

****

1. Explain the following terms in detail:
   * + - 1. Bias
         2. Variance
         3. Overfitting
         4. Underfitting

**Answer:**

**Bias:** The inability of a machine learning algorithm to capture the true relationship is called as bias – Training data

**Variance:** The difference in fits between datasets is called variance – Testing data. Bias and variance are inversely proportional (Increasing the bias will decrease the variance).

**Underfitting:** (High bias and low variance) It means that an ML model or the algorithm does not fit the data well enough. It usually happens when we have fewer data in our dataset. The model will perform poorly on training and testing data.

**Overfitting:** (High variance and low bias) It happens when we train an ML model with a lot of data. The model does not categorize the data correctly, because of too many details and noise. The model will perform well on training data and badly on testing data.

1. Explain the three types of Cross Validation techniques.

**Answer:**

**Hold-Out Method:** The entire dataset is divided into 2 sets – train set and test set. The proportion of training data has to be larger than the test data. The data split happens randomly.

**Leave One Out Cross-Validation:** In 1st observation, we select a single observation as test data, and everything else is labeled as training data and the model is trained. Now the 2nd observation is selected as test data and the model is trained on the remaining data. This process continues ‘n’ times and the average of all these iterations is calculated and estimated as the test set error.

**K-Fold Cross-Validation:** The whole data is divided into k sets of almost equal sizes. The first set is selected as the test set and the model is trained on the remaining k-1 sets. This process continues for all the k sets. The mean of errors from all the iterations is calculated as the test error estimate.

1. Differentiate between Bagging and Boosting with suitable examples.

**Answer:**

**Bagging (Bootstrap Aggregating):** The idea behind bagging is combining the results of multiple models (for instance, all decision trees) to get a generalized result. Each model is built independently. Aim to decrease variance, not bias. E.g., The Random Forest model uses Bagging.

**Boosting:** A sequential process, where each subsequent model attempts to correct the errors of the previous model. The succeeding models are dependent on the previous model. When an input is misclassified, its weight is increased so that next hypothesis is more likely to classify it correctly. Aim to decrease bias, not variance. The AdaBoost uses Boosting techniques.

1. What is PCA? Explain the different steps involved in Principal Component Analysis (PCA)?

**Answer:**

**PCA:** It is a dimensionality-reduction method used to reduce the dimensionality of large datasets. It does so by transforming a large set of features into smaller ones that still contain most of the information in the original dataset.

**Steps:**

Step 1: Standardize/Normalize the dataset.

Step 2: Calculate the covariance matrix for the features in the dataset.

Step 3: Calculate the eigenvalues and eigenvectors for the covariance matrix.

Step 4: Sort eigenvalues and their corresponding eigenvectors.

Step 5: Pick k eigenvalues and form a matrix of eigenvectors. (k is no. of principal components)

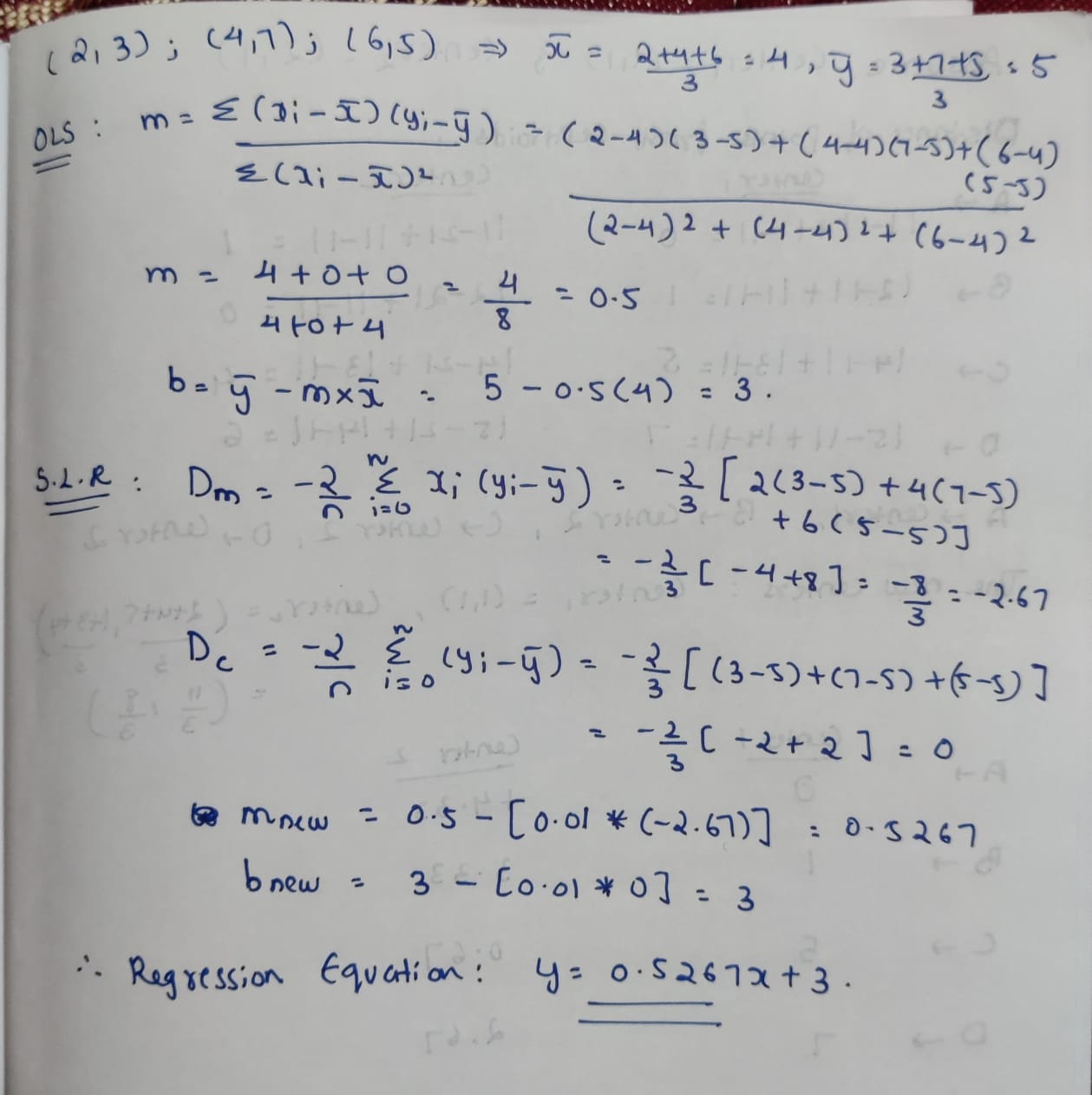
Step 6: Transform the original matrix.

1. Consider the following data:

|  |  |
| --- | --- |
| X | Y |
| 2 | 3 |
| 4 | 7 |
| 6 | 5 |

Perform OLS to find out the initial regression coefficient m and slope b. Then perform Simple Linear Regression to find out the regression line equation. Show all the steps and calculations.

**Answer:**

****