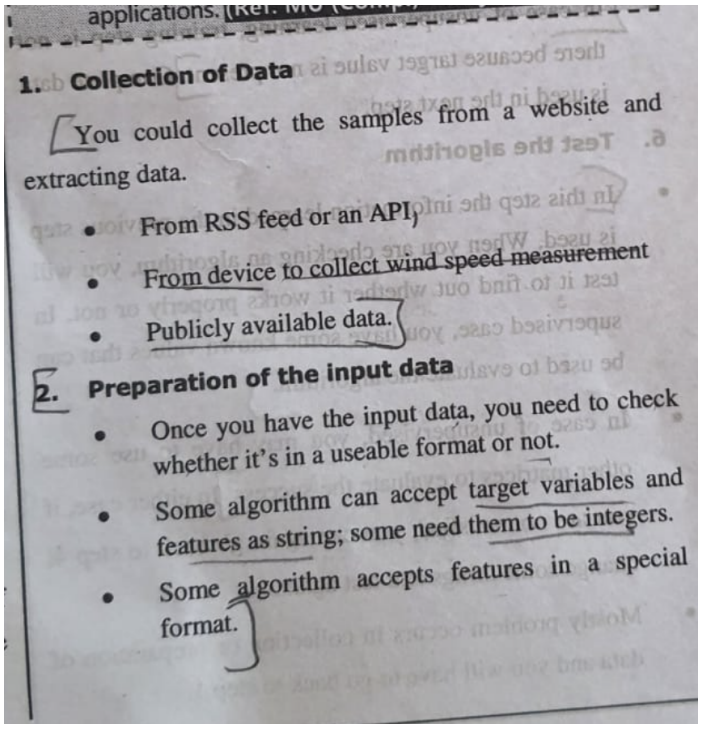
ML QUESTION BANK IA-1

1. **Define the concept of classification.**

Ans) A classification algorithm is a supervised learning technique that uses data training to determine data into different classes. Classification predictive modeling is trained using data or observations, and new observations are categorized into classes or groups. Classification predictive modeling is the task of a mapping function (f) from input variables (x) to discrete output variables (y). In this approach, the algorithm generates a probability score and assigns this score to the input. For example, email service providers use classification to generate probability scores for email identification to determine if the email is in the spam class or not.

2. **How will you design a machine learning system? Steps of developing machine learning.**

Ans)



applications.

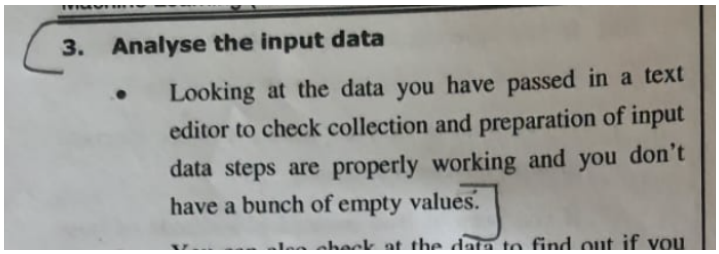
1. Collection of Data

You could collect the samples from a website and extracting data.

- From RSS feed or an API,
- From device to collect wind speed measurement
- Publicly available data.

2. Preparation of the input data

- Once you have the input data, you need to check whether it's in a useable format or not.
- Some algorithm can accept target variables and features as string; some need them to be integers.
- Some algorithm accepts features in a special format.



3. Analyse the input data

- Looking at the data you have passed in a text editor to check collection and preparation of input data steps are properly working and you don't have a bunch of empty values.

4. The importance of this step is that it makes you understand that you don't have any garbage value coming in.

5. **Train the algorithm**

- Good clean data from the first two steps is given as input to the algorithm. The algorithm extracts information or knowledge. This knowledge is mostly stored in a format that is readily useable by machine for next 2 steps.
- In case of unsupervised learning, training step is not there because target value is not present. Complete data is used in the next step.

6. **Test the algorithm**

- In this step the information learned in the previous step is used. When you are checking an algorithm, you will test it to find out whether it works properly or not. In supervised case, you have some known values that can be used to evaluate the algorithm.
- In case of unsupervised, you may have to use some other matrices to evaluate the success. In either case, if

7. **Use it**

In this step a real program is developed to do some task, and once again it is checked if all the previous steps worked as you expected. You might encounter some new data and have to revisit step 1-5.

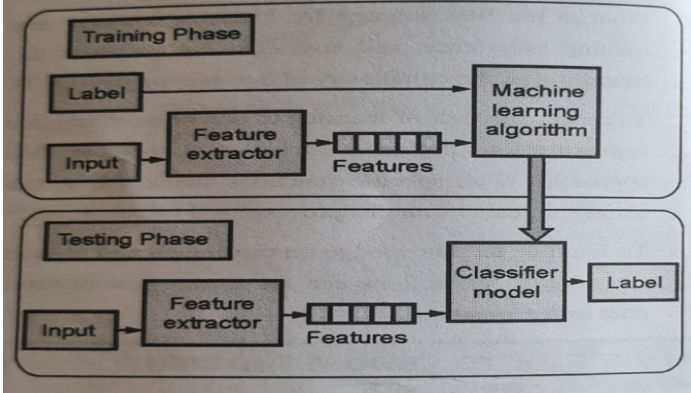


Fig. 1.6.1 : Typical example of Machine Learning Application

3. **What are real life applications of machine learning?(explain on own)(pg 1-16,1-17)**

Ans) 1. Learning Association: (market-basket analysis)

2. Classification: loan given by bank

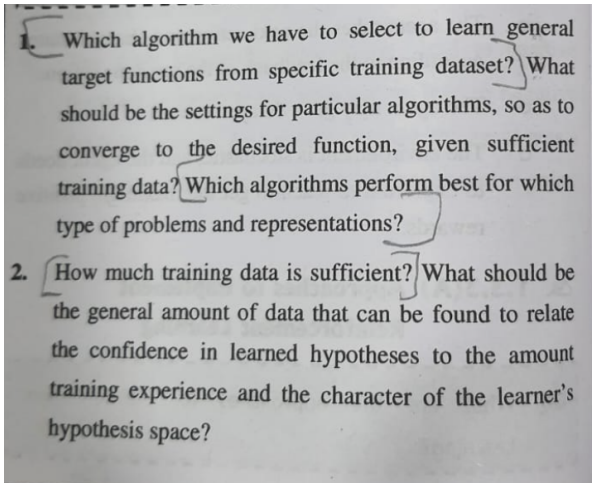
3. Regression: (Price/score prediction)

4. Unsupervised Learning: (clustering,to find outliers, create your own labeled data,check patterns,etc)

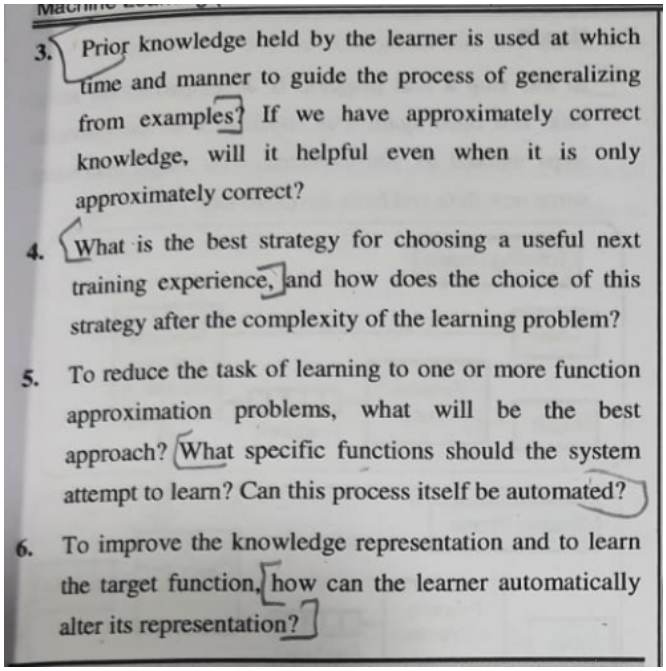
5. Reinforcement Learning: (Chess game)

4. **List and explain issues in machine learning.**

Ans)



1. Which algorithm we have to select to learn general target functions from specific training dataset? What should be the settings for particular algorithms, so as to converge to the desired function, given sufficient training data? Which algorithms perform best for which type of problems and representations?
2. How much training data is sufficient? What should be the general amount of data that can be found to relate the confidence in learned hypotheses to the amount training experience and the character of the learner's hypothesis space?



3. Prior knowledge held by the learner is used at which time and manner to guide the process of generalizing from examples? If we have approximately correct knowledge, will it helpful even when it is only approximately correct?
4. What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy after the complexity of the learning problem?
5. To reduce the task of learning to one or more function approximation problems, what will be the best approach? What specific functions should the system attempt to learn? Can this process itself be automated?
6. To improve the knowledge representation and to learn the target function, how can the learner automatically alter its representation?

5. **Calculate eigen vector of a given matrix**

A = 1 2 -3
2 4 -6
-1 -2 3

6. **What are the performance measures to analyze the quality of the model?**

Ans)The performance of a machine learning model can be evaluated using several performance measures. The choice of performance measure depends on the problem type and the goals of the model. Some common performance measures include:

- Accuracy: The percentage of correctly classified instances in the test set.

$$\text{Accuracy} = \frac{\sum \text{TP} + \text{TN}}{\sum \text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

- Precision: The ratio of true positives to the total number of predicted positives.Precision measures the model's ability to identify only the relevant instances.

$$\text{Precision} = \frac{\sum \text{TP}}{\sum \text{TP} + \text{FP}}$$

- Recall: The ratio of true positives to the total number of actual positives. Recall measures the model's ability to find all relevant instances.

$$\text{Recall} = \frac{\sum \text{TP}}{\sum \text{TP} + \text{FN}}$$

- F1-score: The harmonic mean of precision and recall. F1-score is a balanced measure that combines precision and recall.

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

- Mean Squared Error (MSE): The average squared difference between the predicted and actual values. MSE is used for regression problems.

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

Where:

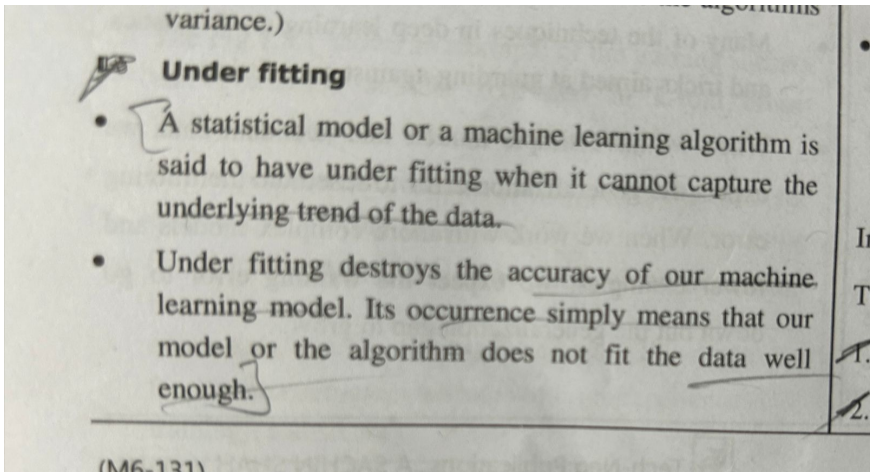
- y_i is the i^{th} observed value.
- $\hat{y}_i$ is the corresponding predicted value.
- n = the number of observations.

- Root Mean Squared Error (RMSE): The square root of the mean squared error. RMSE is used for regression problems.
  RSME= root of MSE
- Confusion Matrix: A matrix that summarizes the number of true positives, false positives, true negatives, and false negatives. Confusion matrix is used to calculate other performance measures such as accuracy, precision, recall, and F1-score.

		Actual Values: Positive (1)	Actual Values: Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

7. Explain overfitting and underfitting of models.

Ans)


variance.)

Under fitting

- A statistical model or a machine learning algorithm is said to have under fitting when it cannot capture the underlying trend of the data.
- Under fitting destroys the accuracy of our machine learning model. Its occurrence simply means that our model or the algorithm does not fit the data well enough.

model will probably

- Under fitting can be avoided by using more data and also reducing the features by feature selection.

In a nutshell, Under fitting - High bias and low variance

Techniques to reduce under fitting :

1. Increase model complexity
2. Increase number of features, performing feature engineering
3. Remove noise from the data.
4. Increase the number of epochs or increase the duration of training to get better results.

Over fitting

- A statistical model is said to be over fitted, when we train it with a lot of data. When a model gets trained with so much of data, it starts learning from the noise and inaccurate data entries in our data set. Then the model does not categorize the data correctly, because of too many details and noise.

- A solution to avoid over fitting is using a linea algorithm if we have linear data or using th parameters like the maximal depth if we are usir decision trees.

In a nutshell, **Overfitting - High variance and low bias**

Techniques to reduce overfitting:

1. Increase training data.
2. Reduce model complexity.

8. Calculate SVD of given matrix

A = 1 0 1
-2 1 0

9. Diagonalize the given matrix A as A= XDX-1

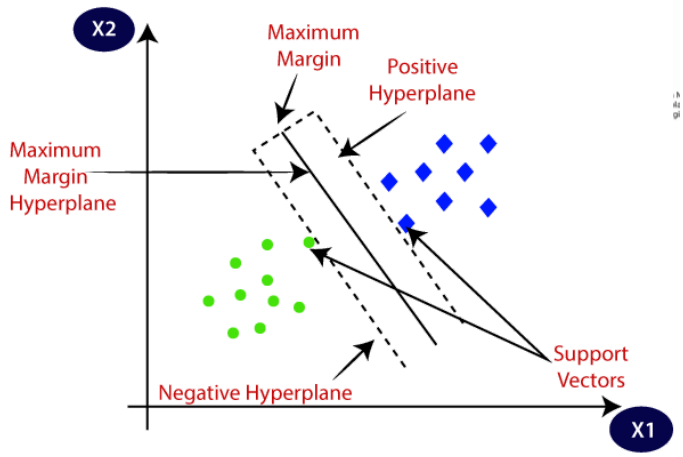
A = 111
111
111

10. Explain support vector machines.

Ans)Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence algorithm is termed as Support Vector Machine.



Example: Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as the support vector creates a decision boundary between these two data (cat and dog) and chooses extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat or dog.

11. What is regularized regression?

Ans)

Regularised regression is a type of regression where the **coefficient** estimates are constrained to zero.

The magnitude (size) of coefficient, as well as the magnitude of the **error term** are **penalised**. Complex models are discouraged to avoid **overfitting**.

"Regularisation" is a way to give a penalty to certain models, generally complex ones.

Ridge regression belongs to the set of L2 regularisation tools. L2 regularisation adds penalty called an L2 penalty, and is equal to the square of the magnitude of coefficients.

All coefficients are shrunk by the same factor, hence all coefficients remain in the same model.

The strength of the penalty term is controlled by a tuning parameter. When this tuning parameter (λ) is set to zero, ridge regression equals least square regression.

When $\lambda = \infty$, all coefficients are shrunk to zero.

Hence the ideal penalty is therefore somewhere in between 0 and ∞.

3.5.1 Types of Regularised Regression

Two commonly used types of regularised regression methods are :

(1) Ridge Regression

It is a way to create a **parsimonius** model when the number of predictor variables in a set exceeds the number of observations, or when a data set has multicollinearity (correlation between predictor variables).

(2) Lasso Regression

It is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean.

This type is very useful when there is a high level of multicollinearity or when one wants to automate certain parts of model selection, like variable selection/ parameter elimination.

12. Explain the norm of a vector.

Ans) The length of the vector is referred to as the vector norm or the vector's magnitude.

The length of a vector is a nonnegative number that describes the extent of the vector in space, and is sometimes referred to as the vector's magnitude or the norm.

For example, we have vector v1=[-2,1]

Norm of a vector or L would be,

L= $\sqrt{(-2)^2 + 1^2} = 2.236$

13. Explain supervised machine learning.

Ans) Supervised learning is the type of machine learning in which machines are trained using well "labeled" training data, and on the basis of that data, machines predict the output. The labeled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

Supervised learning can be further divided into two types of problems:

- **Classification:**
  Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.
- **Regression:**
  Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

14. Explain unsupervised machine learning.

Ans) Unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the

human brain while learning new things. It can be defined as: Unsupervised learning is a type of machine learning in which models are trained using an unlabeled dataset and are allowed to act on that data without any supervision.

The unsupervised learning algorithm can be further categorized into two types of problems:

- Clustering: Clustering is a method of grouping the objects into clusters such that objects with most similarities remain into a group and have less or no similarities with the objects of another group.
- Association: An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occur together in the dataset.

Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs. The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset. The task of the unsupervised learning algorithm is to identify the image features on their own. Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.

15. Find vectors that are orthogonal to [1,2,3]. Explain why we can have an infinite number of such vectors.

16. Explain least squares method for supervised machine learning technique.

Ans) Least-square method is the curve that best fits a set of observations with a minimum sum of squared residuals or errors. Let us assume that the given points of data are (x1, y1), (x2, y2), (x3, y3), …, (xn, yn) in which all x's are independent variables, while all y's are dependent ones. This method is used to find a linear line of the form y = mx + b, where y and x are variables, m is the slope, and b is the y-intercept. The formula to calculate slope m and the value of b is given by:

$m = (n\sum xy - \sum y\sum x)/n\sum x^2 - (\sum x)^2$

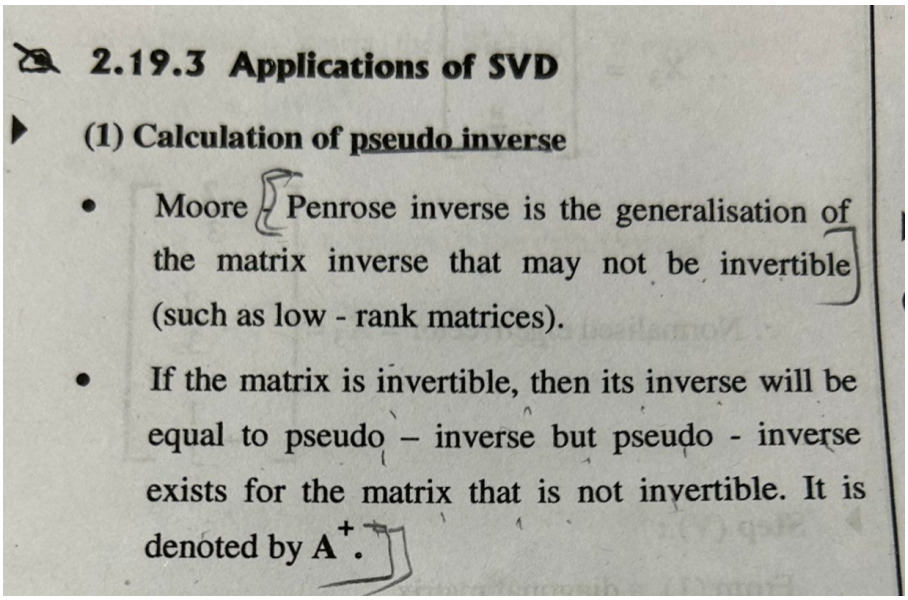
$b = (\sum y - m\sum x)/n$

Here, n is the number of data points.

For example,
To illustrate, consider the case of an investor considering whether to invest in a gold mining company. The investor might wish to know how sensitive the company's stock price is to changes in the market price of gold. To study this, the investor could use the least squares method to trace the relationship between those two variables over time onto a scatter plot. This analysis could help the investor predict the degree to which the stock's price would likely rise or fall for any given increase or decrease in the price of gold.

17. **Solve the linear system -x1+ x2+2x3 = 2, 3x1-x2+x3= 6, -x1+3x2+4x3= 4.**

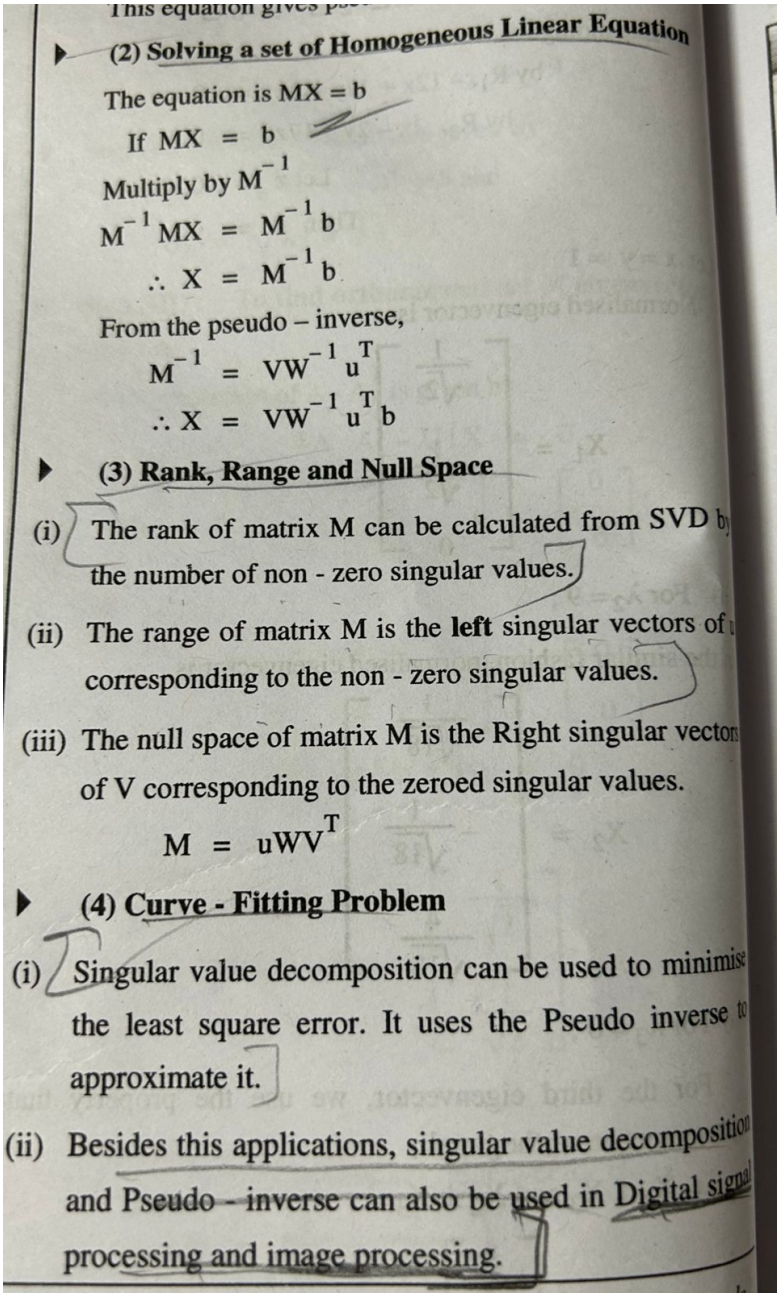
18. **What are the applications of singular value decomposition (SVD)**



2.19.3 Applications of SVD

(1) Calculation of pseudo inverse

- Moore Penrose inverse is the generalisation of the matrix inverse that may not be invertible (such as low - rank matrices).
- If the matrix is invertible, then its inverse will be equal to pseudo – inverse but pseudo - inverse exists for the matrix that is not invertible. It is denoted by A^+.



This equation gives ps...

(2) Solving a set of Homogeneous Linear Equation

The equation is $MX = b$

If $MX = b$

Multiply by M^{-1}

$$M^{-1}MX = M^{-1}b$$

$$\therefore X = M^{-1}b$$

From the pseudo – inverse,

$$M^{-1} = VW^{-1}u^{T}$$

$$\therefore X = VW^{-1}u^{T}b$$

(3) Rank, Range and Null Space

(i) The rank of matrix M can be calculated from SVD by the number of non - zero singular values.

(ii) The range of matrix M is the **left** singular vectors of u corresponding to the non - zero singular values.

(iii) The null space of matrix M is the Right singular vectors of V corresponding to the zeroed singular values.

$$M = uWV^{T}$$

(4) Curve - Fitting Problem

(i) Singular value decomposition can be used to minimise the least square error. It uses the Pseudo inverse to approximate it.

(ii) Besides this applications, singular value decomposition and Pseudo - inverse can also be used in Digital signal processing and image processing.

*NOTE: Matrix and linear equations may change.

Formulas:

Note

If $A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$ then

the characteristic equation is

$$|A - \lambda I| = 0$$

$$\begin{vmatrix} a_1 - \lambda & a_2 & a_3 \\ b_1 & b_2 - \lambda & b_3 \\ c_1 & c_2 & c_3 - \lambda \end{vmatrix} = 0$$

If we expand the determinant on the L. H.S, we get

$$\lambda^3 - (a_1 + b_2 + c_3)\lambda^2 + \left(\begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_3 \\ c_1 & c_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\right)\lambda - |A| = 0$$

We can directly write cubical equation in the problem without solving determinant

Eigen Vector:

$$[A - \lambda I]X = 0 \text{ gives}$$

Use cramers rule,

Example,

By crammer's rule

$$\frac{x_1}{\begin{vmatrix} -8 & -2 \\ -4 & -2 \end{vmatrix}} = \frac{-x_2}{\begin{vmatrix} 7 & -2 \\ 4 & -2 \end{vmatrix}} = \frac{x_3}{\begin{vmatrix} 7 & -8 \\ 4 & -4 \end{vmatrix}}$$

$$\frac{x_1}{8} = \frac{-x_2}{-6} = \frac{x_3}{4}$$

$$\therefore \quad \frac{x_1}{4} = \frac{x_2}{3} = \frac{x_3}{2}$$

$\therefore$ corresponding to eigen value $\lambda = 1$,

the eigen vector is $\begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$