# CAT Advanced Task :

1. What is the rank of a matrix? Why is it important in machine learning?

Num of linearly independent rows or columns(non-zero rows or columns) which is the dimension of the vector space and it is <=m or <=n based on the smaller number.

Can be used in ML in feature selection as when feature correlation matrix have low rank this shows that some features aren’t important.

In Pca rank of the data matrix plays a key role in determining how many components can be retained without losing significant information.

In solving linear systems it determines whether a unique solution exists in linear regression and other optimization problems.

1. Explain eigenvalues and eigenvectors. How are they used in PCA?

**Eigenvalues and Eigenvectors** are the scalar and vector quantities associated with matrices used for linear transformations.

A⋅v=λ⋅v in this eqution : v is eigen vector and lamda is eigenvalue

This show the direction and magnitude of matrix action.

IN PCA:

The eigenvector corresponding to the largest eigenvalue tells us the direction in the data where the maximum variance occurs and the smaller eigenvalue second principal component (perpendicular to the first one) and corresponds to the next largest variance.

1. What is a singular value decomposition (SVD), and why is it important?

SVD is representation of any matrix in terms of singular values : A=U Σ V transpose

Where :

U : **left singular vectors** of A

**Σ**: Contains the **singular values** of the matrix, which are the square roots of the non-zero eigenvalues

**V transpose** : Contains the **right singular vectors** of A.

It is used in PCA , Noise Reduction to keep the largest singular values and discard the smaller ones to filter out noise,Image Compression where only the most significant singular values and their corresponding vectors are kept.

1. How would you compute the inverse of a matrix? What is its relevance in Machine Learning?

Inverse of a matrix is the square matrix to the power of -1 .

Done by steps of:

1. Getting determinant of square matrix
2. Getting the adjoint matrix
3. Apply the formula which is 1/det(matrix) \* adj(matrix)

Inverse can also be done by gaussian elimination(reducing matrix to I)

It is used in ML to solve linear regression equations (computing wieghts) or modifying weights in ridge regularization and solving linear equations

1. What is the difference between correlation and causation?

Correlation measure how two variables move together and measured between -1 to 1 , this process help in defining relevant featires for feature selection.

Causation it shows how a variable effect another variable and requires proven evidence , it help insure the model’s decision isn’t made on irrelevant links.

1. How does gradient descent work? Explain its importance in ML.

It works on optimizing the loss function. The main idea is keep adjusting the parameters in in the direction that leads to reaching the optima solution.

steps:

1-Model parameters are initialized (w,b)

2-Partial derivatives of loss function with respect to each parameter is computed

3-Parameter is updated so that W(new)=W(old)-(learning rate \* dx/dw)

b(new)=b(old)-(learning rate \* dx/db)

4-process is repeated until the loss function converges

1. What are the types of ML? Discuss each type and give examples on them.
2. **Supervised:** model is trained on known labels this include classification (ex.animal classification)and regression models(ex.house price prediction)
3. **Unsupervised** : mode; trained on unlabeled data and it discover patterns and relationships to create clusters (ex. Clusters:grouping similar users on website, dimentionallity reduction using pca )
4. **Semi supervised:** model is trained on a combination of labeled and unlabeled data.

(ex.google photos grouping a group of photos of same person the you labelling them)

1. **Reinforcement Learning**: agent learns by interacting with an environment. Model use reward and penalties as way of learning (ex games using alphago or smart vacuum systems
2. Explain this code and provide a way to optimize it:A screenshot of a computer screen

   Description automatically generated

The codes shows that fom table Sales S1, Total sales column is computed for each distinct product so that the subquery s2 calcutes the total amount of each product id .

Optimization can be done by using Group by so that

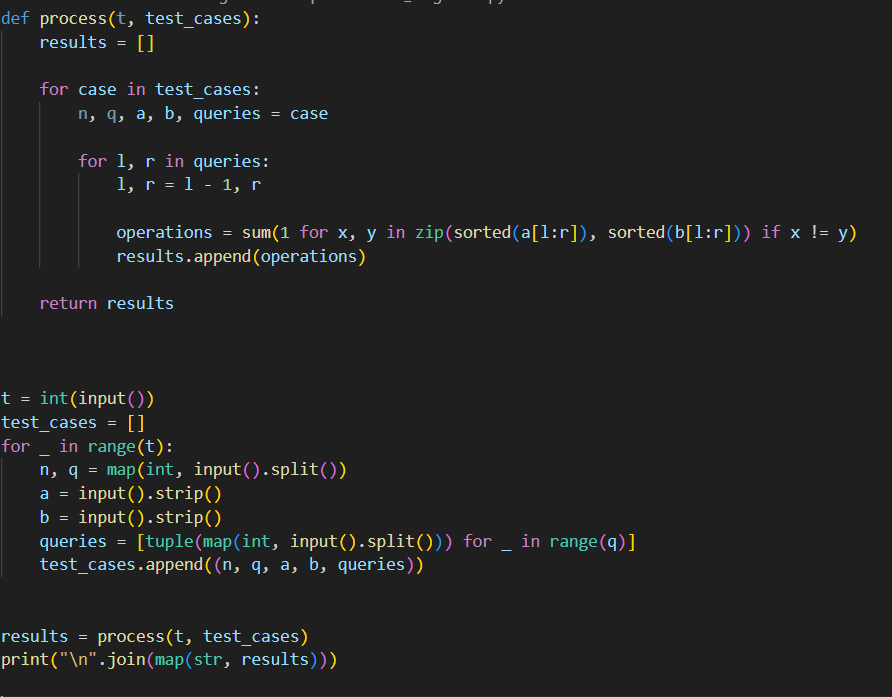
SELECT ProductID,

SUM(amount) as TotalSales

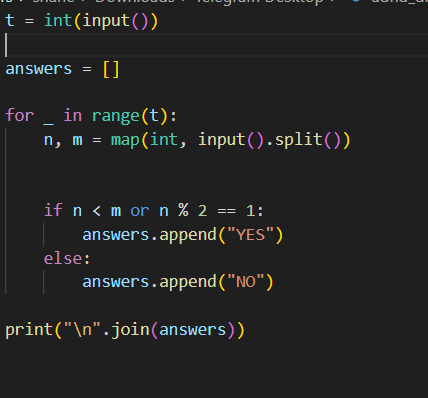
FROM Sales

GROUP BY ProductID;

Problem 1:



Problem 2:



Sql problem1:

SELECT

s.student\_id,

s.student\_name,

sub.subject\_name,

COUNT(e.student\_id) as attended\_exams

FROM

Students s

CROSS JOIN Subjects sub

LEFT JOIN Examinations e ON s.student\_id = e.student\_id

AND sub.subject\_name = e.subject\_name

GROUP BY

s.student\_id,

s.student\_name,

sub.subject\_name

ORDER BY

s.student\_id,

sub.subject\_name;

SQL problem2:

WITH UserTweetCounts AS (

SELECT user\_id, COUNT(\*) as tweet\_count

FROM tweets

WHERE YEAR(tweet\_date) = 2022

GROUP BY user\_id

)

SELECT

tweet\_count as tweet\_bucket,

COUNT(\*) as users\_num

FROM UserTweetCounts

GROUP BY tweet\_count

ORDER BY tweet\_bucket;