Machine-Learning Lab Report

Name : Jawed Hawari

RegNo.:CSE21073

Course Code: 19CSE305

*Abstract:* The contents of this report try to answer the question related machine-learning lab. This report also explains about rank of an observation matrix, regression vs classification, and building a model to predict the stock price. Data used for this report are Purchase data of fruits and IRCTC stock prices.

# Importance of rank of an observation matrix in model building for classification

Rank of a matrix is the maximum number of linearly independent column or row vectors. Rank of an observation matrix is an important factor to consider when building a model for classification. Rank of an observation matrix can be used to predict behavior of a model. The rank of the observation matrix can tell if it is possible to find a unique solution for the model. If the rank is less than number of columns, then there exists some linear dependence between the matrix columns and rows. This means that the model may not be able to distinguish between certain features, leading to overfitting or under fitting.

It plays a fundamental role in the performance and interpretability of machine learning models. Some of its importance are mentioned below.

### Independence of Features: The rank of the observation matrix determines the number of independent features or predictors in the dataset. Linearly dependent features can lead to issues like multicollinearity, where the effect of one feature on the target variable cannot be separated from the effects of other features. This can make it challenging for a classification model to attribute importance to individual features accurately.

### Model Interpretability: When the rank of the observation matrix is high, meaning there are many independent features, it can be challenging to interpret the model’s coefficients or feature importances. This can obscure the understanding of which features are driving the model’s predictions and make it difficult to draw meaningful insights from the model.

### Overfitting: High rank observation matrices can increase the risk of overfitting, especially when the number of features is close to or exceeds the number of observations. Overfit models perform well on the training data but fail to generalize to new, unseen data. Reducing the rank by removing irrelevant or redundant features can help mitigate overfitting.

# Discuss on regression (Ex: a2) and classification (Ex: a3) tasks . Differentiante between them

Regression and classification are two fundamentals types of tasks in machine learning, each with its own objectives, algorithm, and evaluation matrices.

In machine learning model, Regression is use to predict continuous values such as price, salary, age, etc. On the other hand, Classification is use to predict discrete values such as binary values (True or False, Spam or Ham), index of temperature, breed of animal etc.

In Ex: A2, model vector X is calculated for predicting cost of the products available with the vendor. Here the cost of the products calculated by model is continuous because there is no interval mentioned for the cost of the products. In Ex A3, a model is developed, which is categorizing the customers in RICH or POOR based on certain conditions. In this model, the output labels are either RICH or POOR.

## Difference between Regression and Classification

### Output: Regression model produces a continuous output, typically a real number. The prediction is value within a range or a continuous scale. Classification model produces a discrete output, assigning data points to specific categories or classes. The prediction is a label or a class.

### Evvaluation Metrics: Regression model’s common evaluation metrics include MSE, MAE, RMSE, and R-squared. Evaluation metrics in Classification model includes accuracy, precision, recall, F1-score, and ROC-AUC.

### Output Interpretation: Output in Regression model is numeric value that can be interpreted as an estimate or predication of the target value.The output in Classification model is class label, and model’s prediction represents the probability that the input belongs to a particular class.

### Decision Boundary: There is no decision boundary in regression model because the output is continuous. Classification models have decision boundaries that separates different classes in the feature space.

# Observing the stock data provided, record your suggestions to build a system that may be able to predict the price and Change% into future

To create a system capable of predicting stock prices and percentage changes in the future following steps should be taken into consideration.

### Data Collection: Gather histo4rical stock price data, including relevant features such as trading volume, news sentiment, and economic indicators.

### Feature Engineering: Carefully select and engineer features that could have an impact on stock prices, such as moving averages,technical indicators, and macroeconomic variables.

### Data Preprocessing: Clean and preprocess the data to handle missing values and outliers, and ensure it’s in a suitable format for modeling.

### Model Selection: Choose appropriate machine learning model for regression and classification.

### Training and Validation: Split the data into training and validation sets to train and evaluate your models. Use time-series cross-validation techniques to account for temporal dependencies.

### Evaluation Metrics: Use relevant metrics model to train data on the model. For example use RMSE, MAE, etc for regression model and accuracy, precision, and recall for classification model.

### Regularization: Apply regularization techniques to prevent overfitting, as stock data can be noisy and prone to overfitting.

### Feature Importance Analysis: Analyze feature importance to understand which factors have the most significant impact on stock price and changes%.

### Continuous Updating & Backtesting: Implement a mechanism to contiinuously update the model with a new data to adapt to changing market conditions. Evaluate the model’s performance on historical data to assess its effectiveness in a real world context.

Building a stock price prediction system is a complex task that requires careful data handling, model selection, and ongoing monitoring.