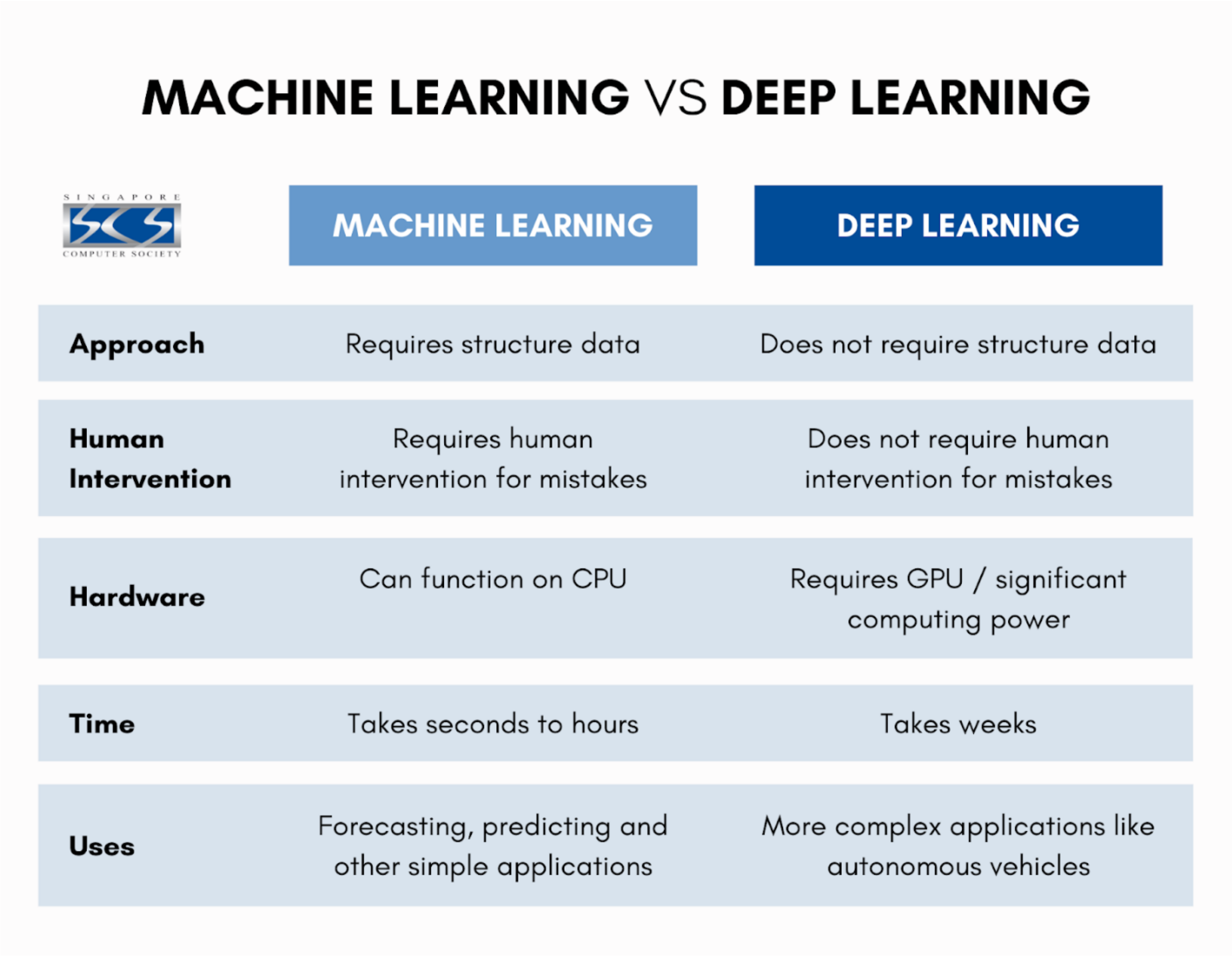
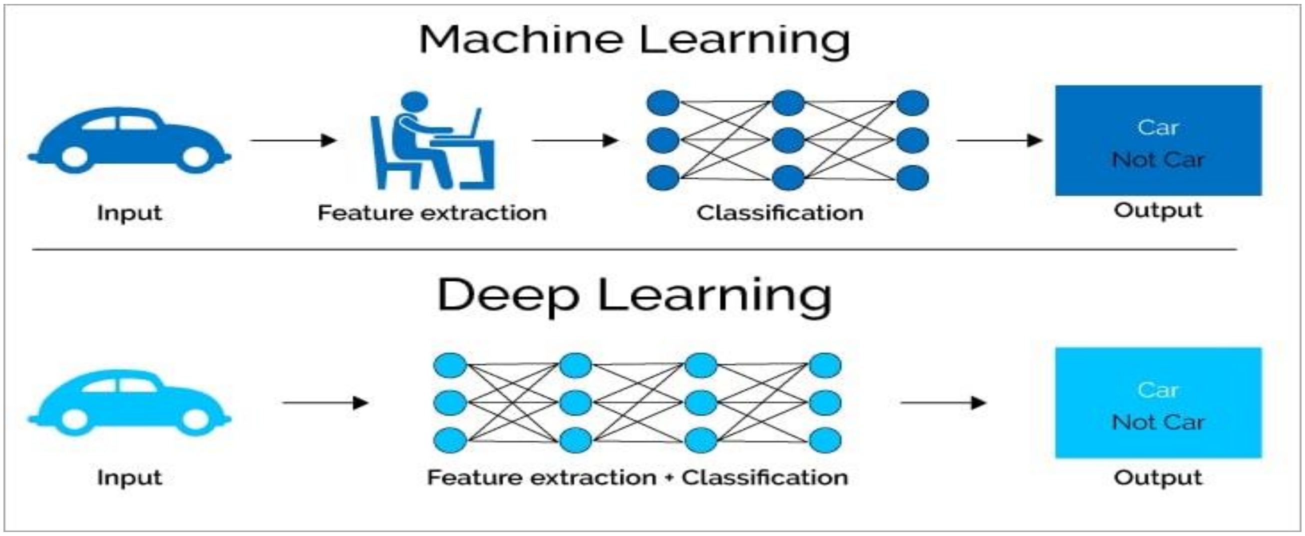
100 days of Machine Learning



MACHINE LEARNING VS DEEP LEARNING

	MACHINE LEARNING	DEEP LEARNING
Approach	Requires structure data	Does not require structure data
Human Intervention	Requires human intervention for mistakes	Does not require human intervention for mistakes
Hardware	Can function on CPU	Requires GPU / significant computing power
Time	Takes seconds to hours	Takes weeks
Uses	Forecasting, predicting and other simple applications	More complex applications like autonomous vehicles



Types of ML:

Supervised Learning	Unsupervised learning	Semi-supervised learning	Reinforcement
- Regression - Classification	- Clustering - Dimensionality - Anomaly detection - Association rule learning		

Supervised ML:

It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. It will have input and output section.

IQ	CGPA	placement
input	input	output

Data type:

Numerical	Categorical
- Age - CGPA - Weight	- Gender - name

Regression: Output will be numerical. EX: house price prediction.

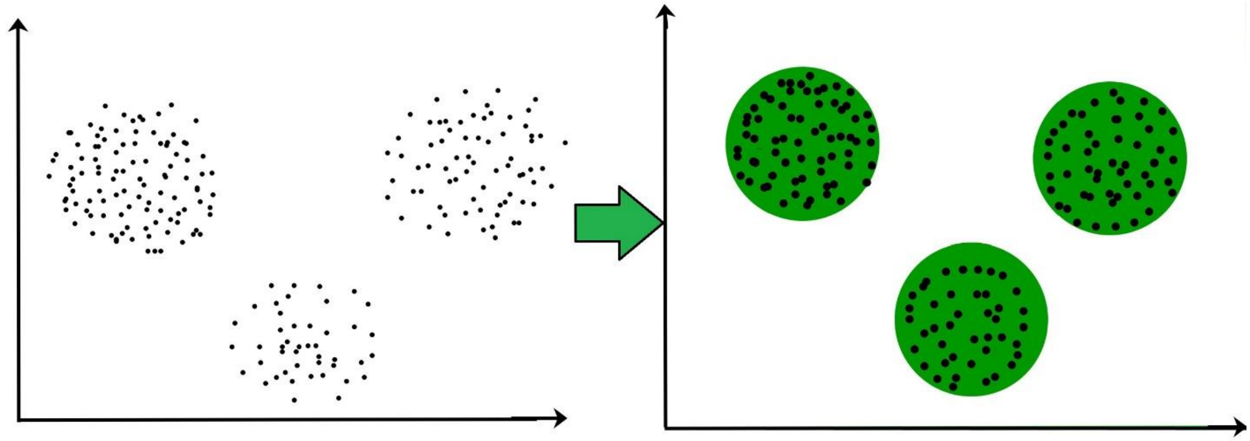
Classification: Output will be categorical. Will it rain today or not.

Unsupervised Learning:

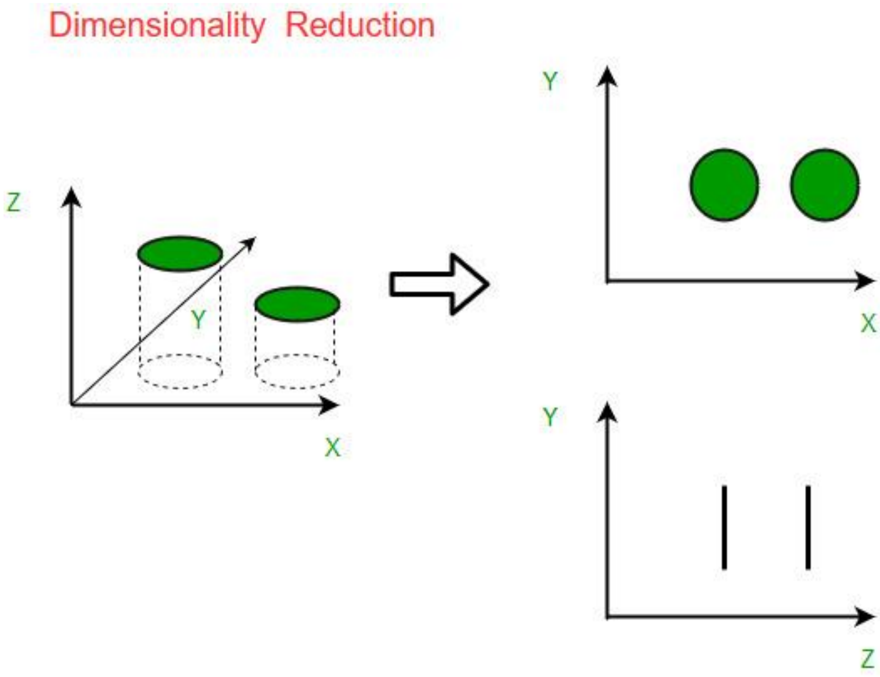
uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. There is no output section.

CGPA	IQ

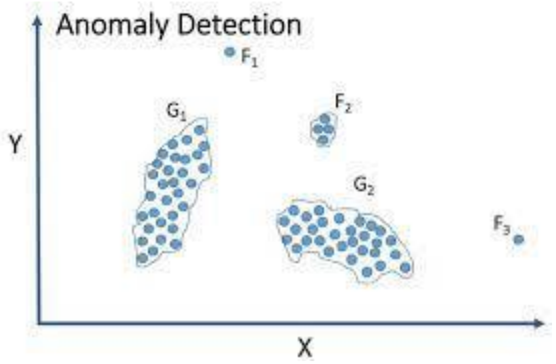
Clustering: making group according to the given dataset.



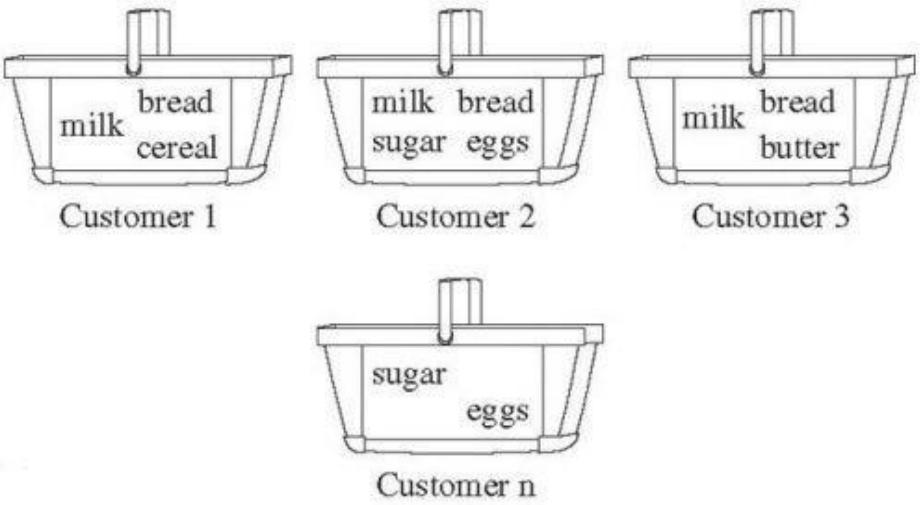
Dimensionality reduction: reduce unnecessary input columns.



Anomaly detection: to catch abnormal data and which are far away from the normal group.



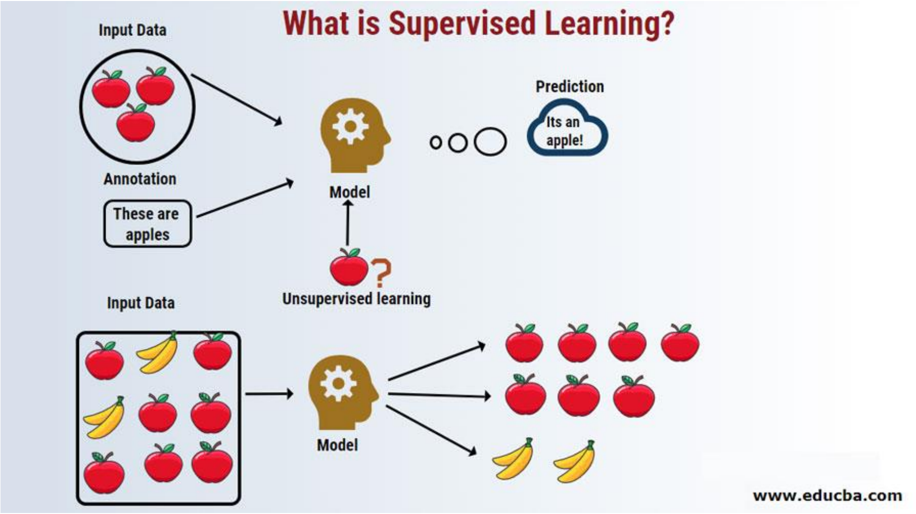
Association rule learning: Association rule mining could be used to identify relationships between items that are frequently purchased together. For example, the rule "If a customer buys bread, they are also likely to buy milk" is an association rule that could be mined from this data set.



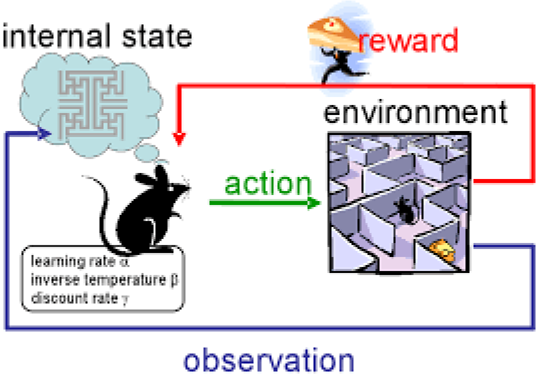
Semi-supervised learning:

In a nutshell, semi-supervised learning (SSL) is a machine learning technique that uses a small portion of labeled data and lots of unlabeled data to train a predictive model. To avoid cost to make data in a category

we use this technique to learn them from given data set. Like facebook auto tag system.



Reinforcement learning: Reinforcement learning (RL) is a subset of machine learning that allows an AI-driven system (sometimes referred to as an agent) to learn through trial and error using feedback from its actions.

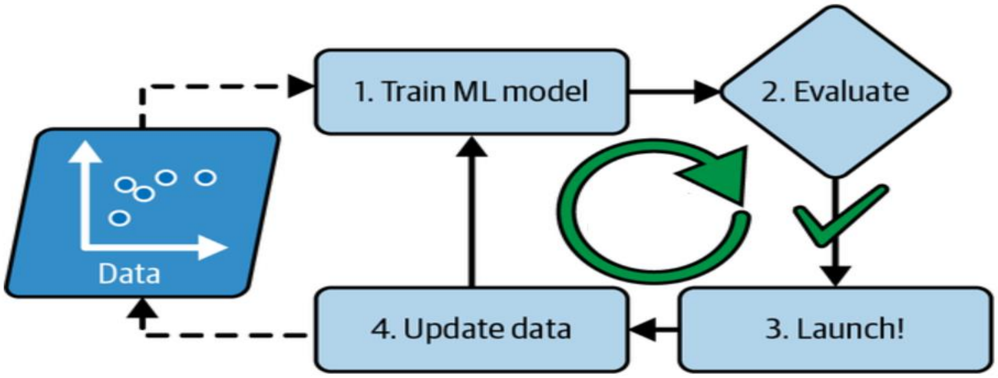


Batch learning:

Batch learning, also known as offline learning, involves training a model on a fixed dataset, or a batch of data, all at once. The model is trained on the entire dataset, and then used to make predictions on new data. We train our model first and then put it into a server. If data is updated then we have to train the offline model again and send the updated model to server again.

Disadvantages of batch learning:

- Lots of data
- Hardware limitations
- Availability

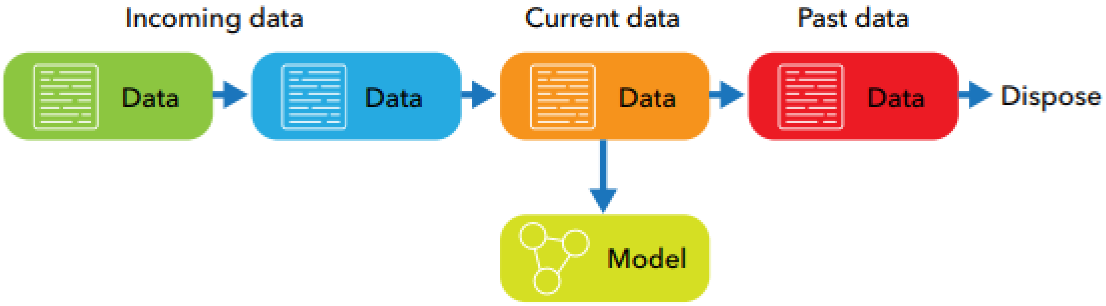


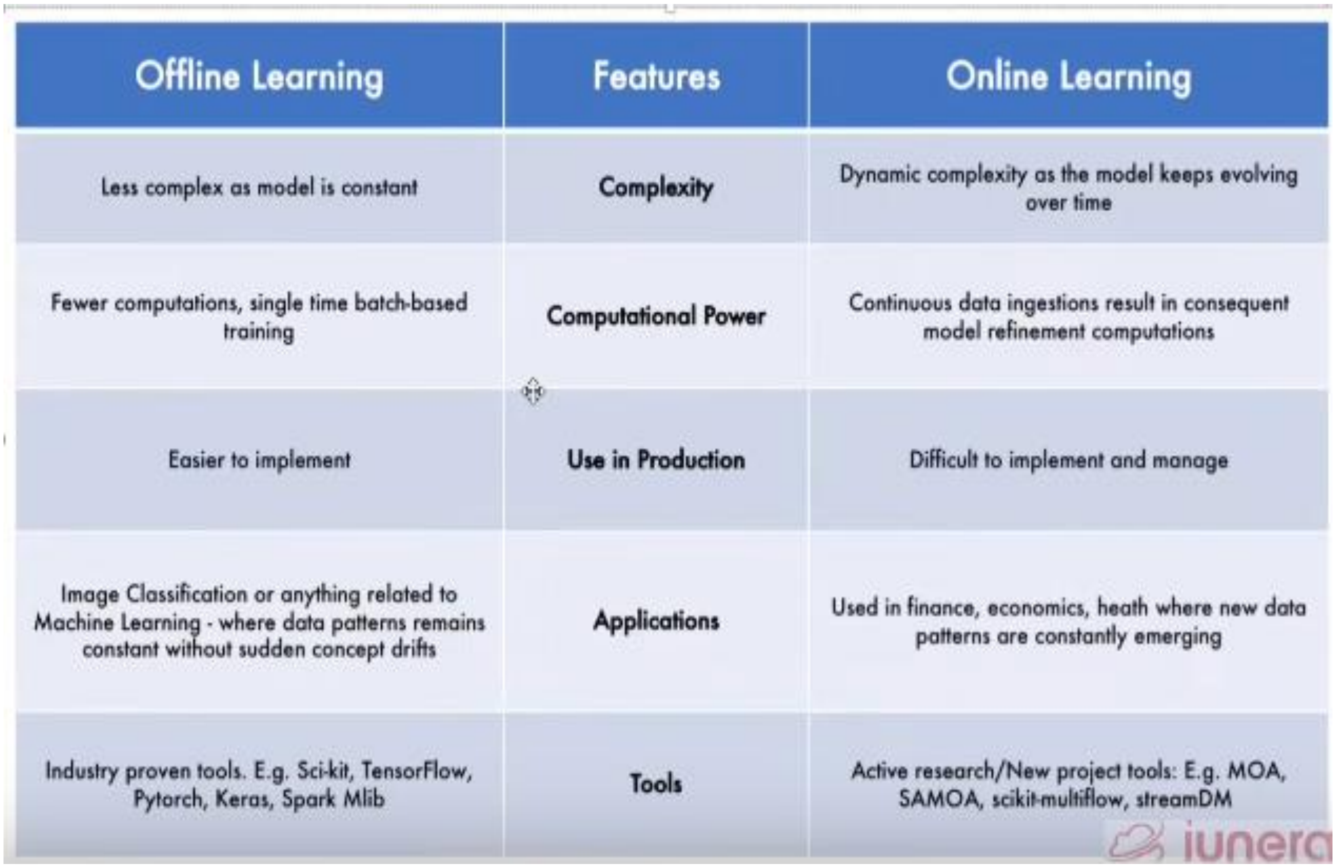
Online machine learning:

Online machine learning is a type of machine learning where data is acquired sequentially and is utilized to update the best predictor for future data at each step. In other words, online machine learning means that learning takes place as data becomes available.

Disadvantages:

- Tricky to use
- Risky (Data is online trained so there can be spam and wrong data can interrupt)

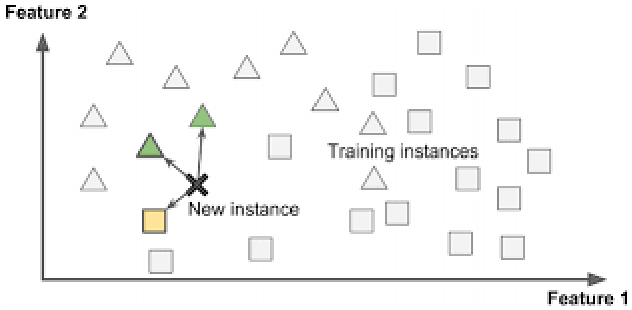




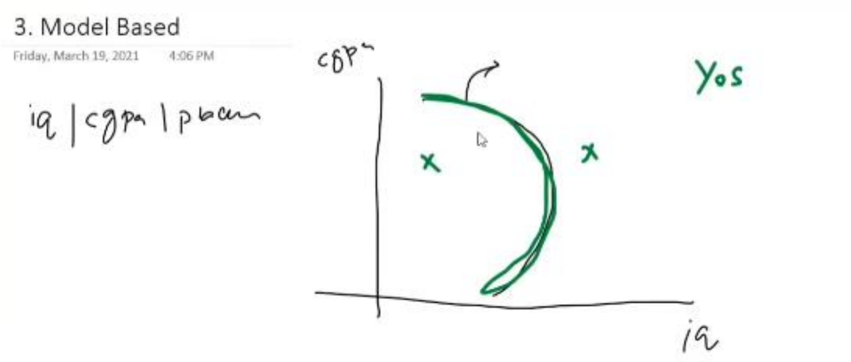
Offline Learning	Features	Online Learning
Less complex as model is constant	Complexity	Dynamic complexity as the model keeps evolving over time
Fewer computations, single time batch-based training	Computational Power	Continuous data ingestions result in consequent model refinement computations
Easier to implement	Use in Production	Difficult to implement and manage
Image Classification or anything related to Machine Learning - where data patterns remains constant without sudden concept drifts	Applications	Used in finance, economics, heath where new data patterns are constantly emerging
Industry proven tools. E.g. Sci-kit, TensorFlow, Pytorch, Keras, Spark Mlib	Tools	Active research/New project tools: E.g. MOA, SAMOA, scikit-multiflow, streamDM

Instance based learning:

Finds the similarity and predicts simply. Instance-based learning refers to a family of techniques for classification and regression, which produce a class label/predication based on the similarity of the query to its nearest neighbor(s) in the training set. It doesn’t train the model. KNN algo is used here.

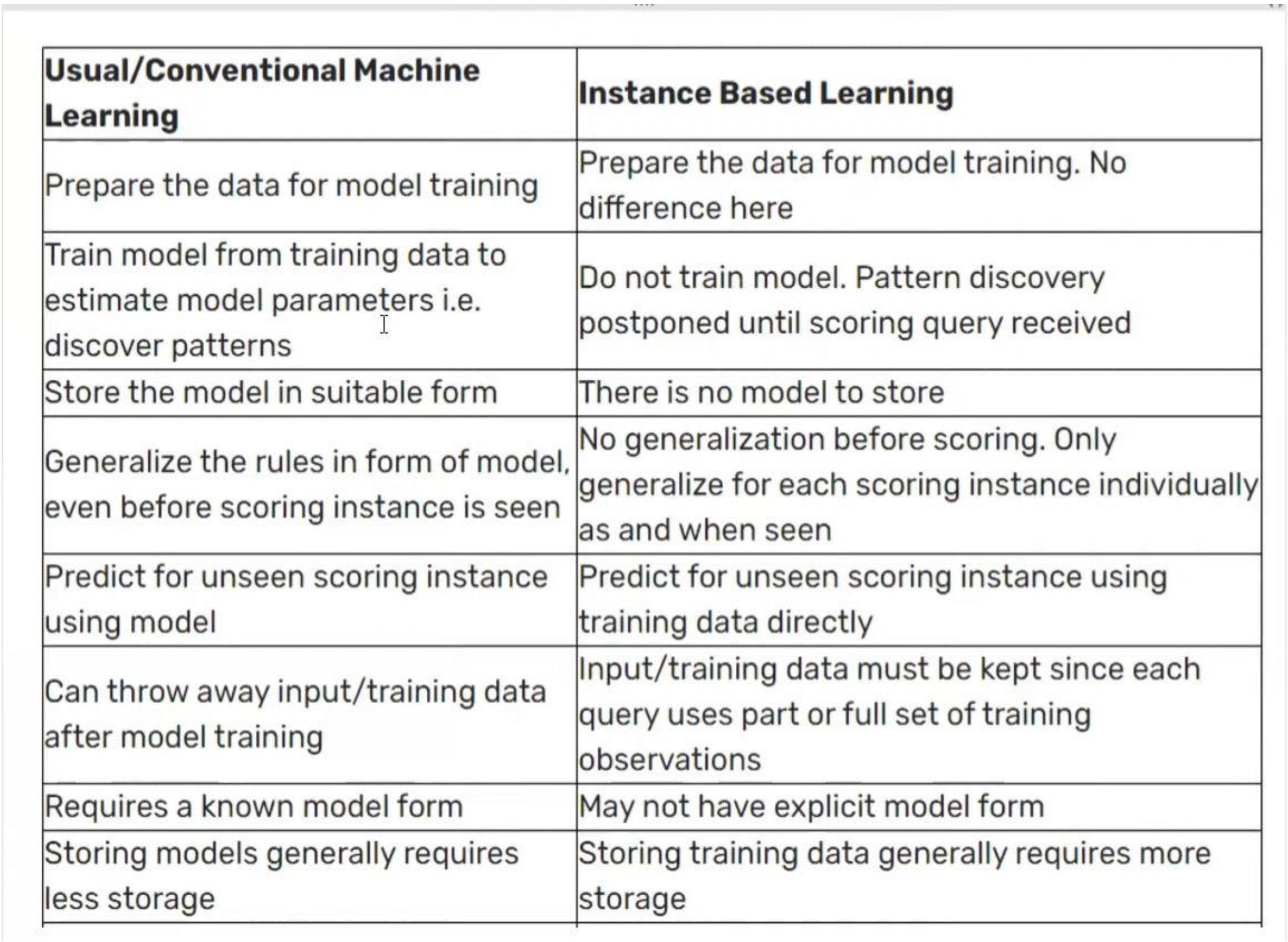


Model based learning: make a decision function and then predicts the output.



3. Model Based

Friday, March 19, 2021 4:06 PM



Usual/Conventional Machine Learning	**Instance Based Learning**
Prepare the data for model training	Prepare the data for model training. No difference here
Train model from training data to estimate model parameters i.e. discover patterns	Do not train model. Pattern discovery postponed until scoring query received
Store the model in suitable form	There is no model to store
Generalize the rules in form of model, even before scoring instance is seen	No generalization before scoring. Only generalize for each scoring instance individually as and when seen
Predict for unseen scoring instance using model	Predict for unseen scoring instance using training data directly
Can throw away input/training data after model training	Input/training data must be kept since each query uses part or full set of training observations
Requires a known model form	May not have explicit model form
Storing models generally requires less storage	Storing training data generally requires more storage

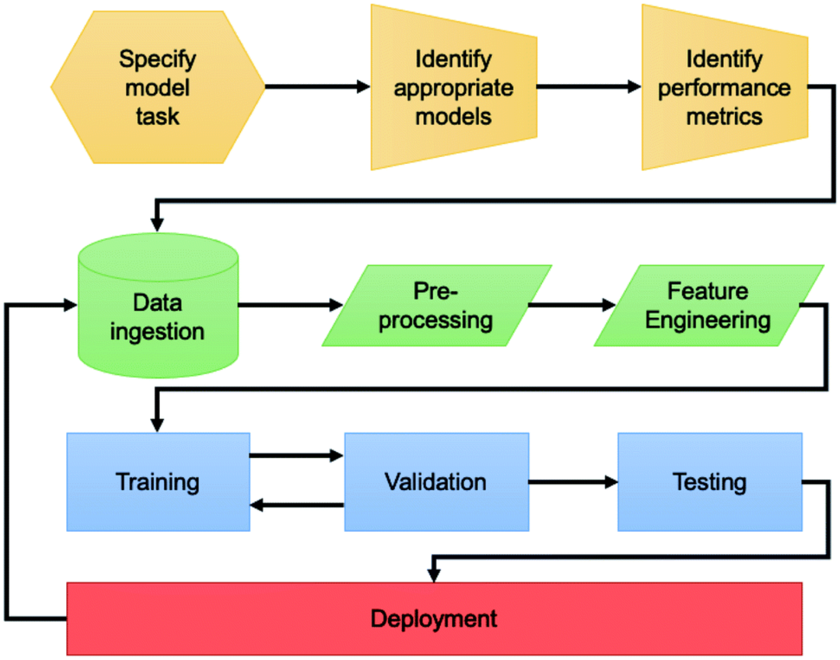
Challenges of machine learning:

- Data collection
- Insufficient data
- Non representative data
- Poor quality data
- Irrelevant features
- Overfitting
- Underfitting
- Software integration
- Offline learning deployment
- Cost involved

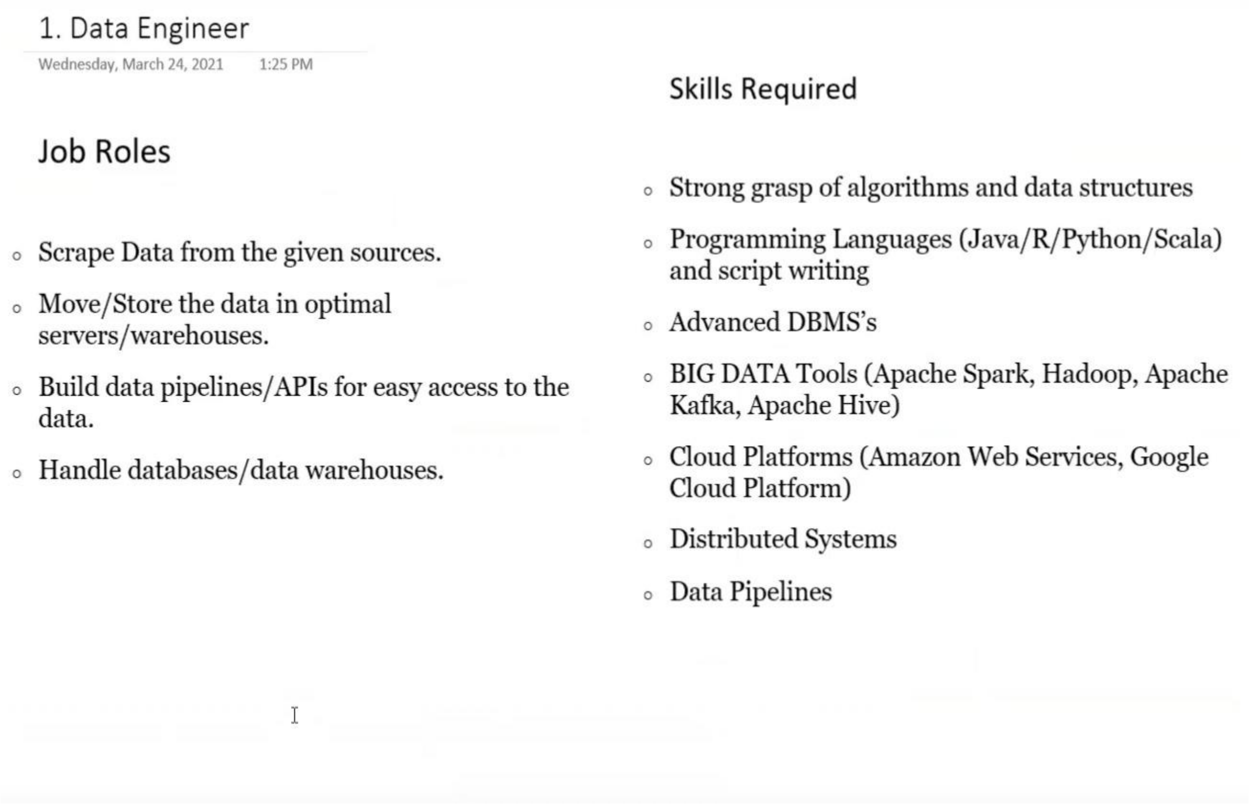
Sentiment analysis:

Sentiment analysis is the process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral. Today, companies have large volumes of text data like emails, customer support chat transcripts, social media comments, and reviews.

Machine learning development life cycle (MLDLC):



Machine learning career:



1. Data Engineer

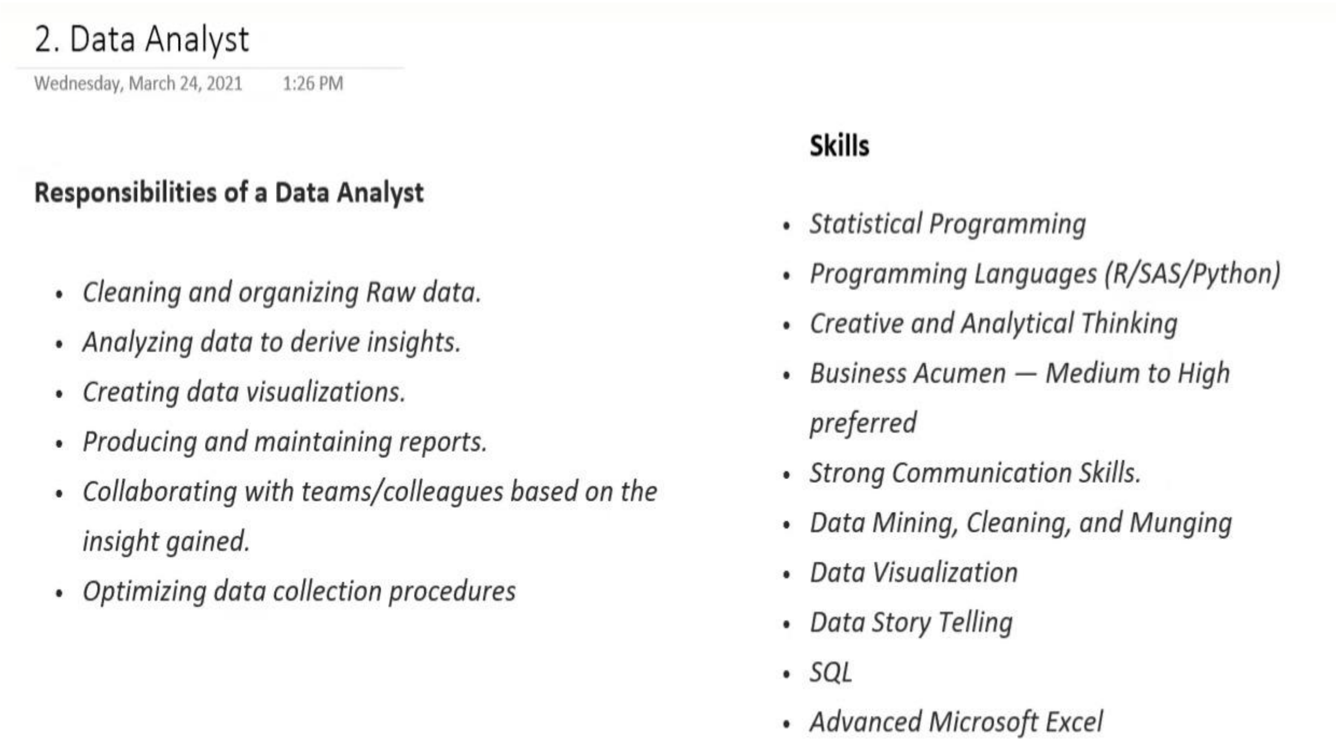
Wednesday, March 24, 2021 1:25 PM

Job Roles

- Scrape Data from the given sources.
- Move/Store the data in optimal servers/warehouses.
- Build data pipelines/APIs for easy access to the data.
- Handle databases/data warehouses.

Skills Required

- Strong grasp of algorithms and data structures
- Programming Languages (Java/R/Python/Scala) and script writing
- Advanced DBMS's
- BIG DATA Tools (Apache Spark, Hadoop, Apache Kafka, Apache Hive)
- Cloud Platforms (Amazon Web Services, Google Cloud Platform)
- Distributed Systems
- Data Pipelines



2. Data Analyst

Wednesday, March 24, 2021 1:26 PM

Responsibilities of a Data Analyst

- *Cleaning and organizing Raw data.*
- *Analyzing data to derive insights.*
- *Creating data visualizations.*
- *Producing and maintaining reports.*
- *Collaborating with teams/colleagues based on the insight gained.*
- *Optimizing data collection procedures*

Skills

- *Statistical Programming*
- *Programming Languages (R/SAS/Python)*
- *Creative and Analytical Thinking*
- *Business Acumen — Medium to High preferred*
- *Strong Communication Skills.*
- *Data Mining, Cleaning, and Munging*
- *Data Visualization*
- *Data Story Telling*
- *SQL*
- *Advanced Microsoft Excel*

3. Data Scientist

Wednesday, March 24, 2021 1:26 PM

“A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician”.

4. ML Engineer

Wednesday, March 24, 2021 1:26 PM

Responsibilities

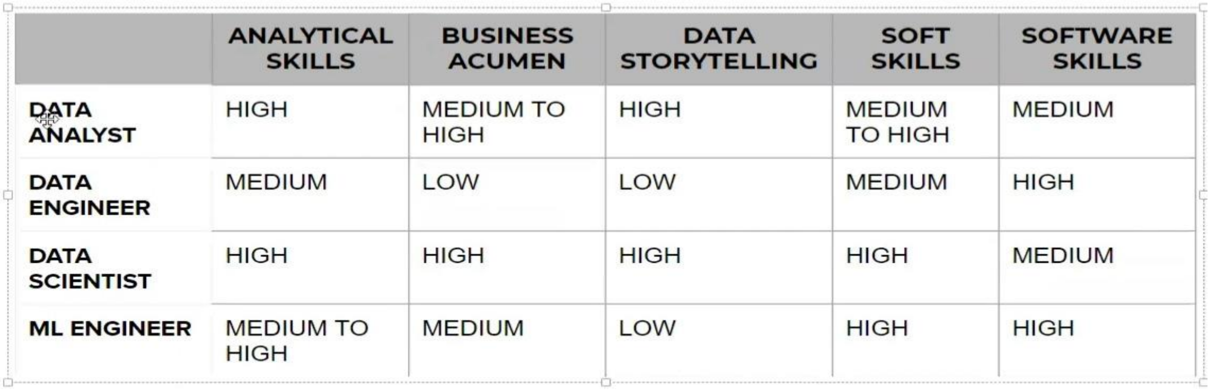
- Deploying machine learning models to production ready environment
- Scaling and optimizing the model for production
- Monitoring and maintenance of deployed models

Skills

- Mathematics
- Programming Languages (R/Python/Java/Scala mainly)
- Distributed Systems
- Data model and evaluation
- Machine Learning models
- Software Engineering & Systems design

5. Comparison

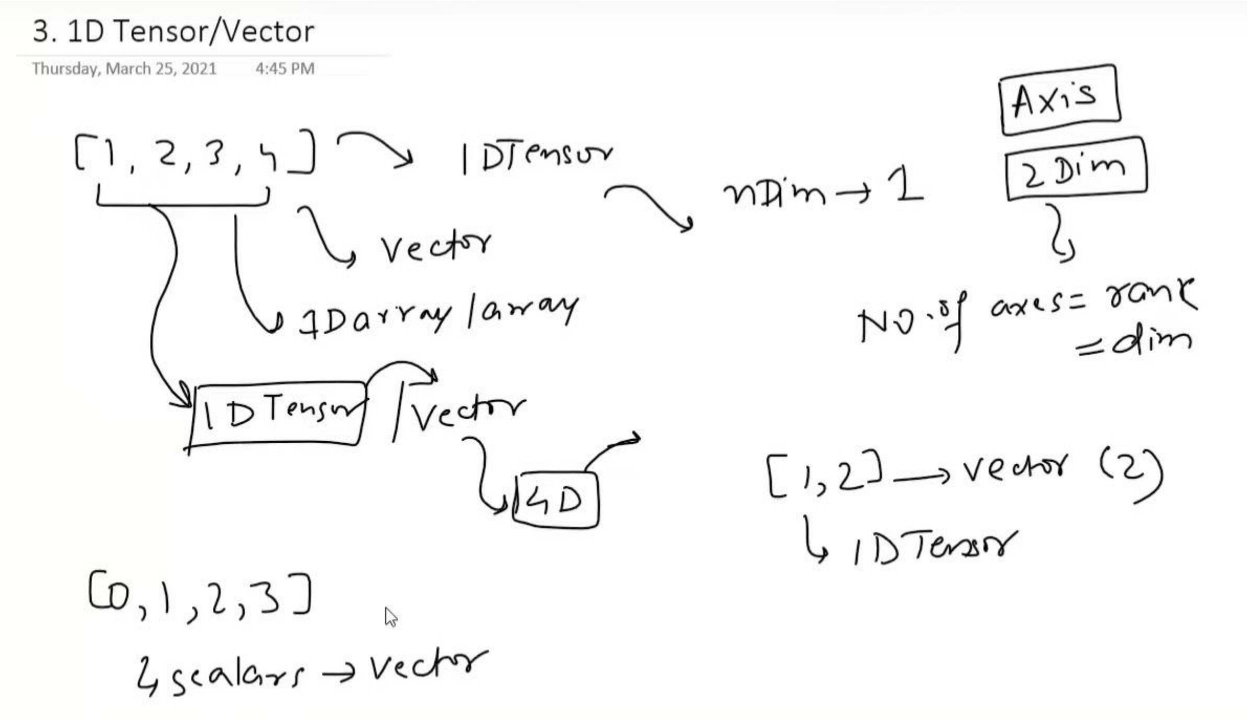
Wednesday, March 24, 2021 1:26 PM



	ANALYTICAL SKILLS	BUSINESS ACUMEN	DATA STORYTELLING	SOFT SKILLS	SOFTWARE SKILLS
DATA ANALYST	HIGH	MEDIUM TO HIGH	HIGH	MEDIUM TO HIGH	MEDIUM
DATA ENGINEER	MEDIUM	LOW	LOW	MEDIUM	HIGH
DATA SCIENTIST	HIGH	HIGH	HIGH	HIGH	MEDIUM
ML ENGINEER	MEDIUM TO HIGH	MEDIUM	LOW	HIGH	HIGH

What is a 1D tensor?

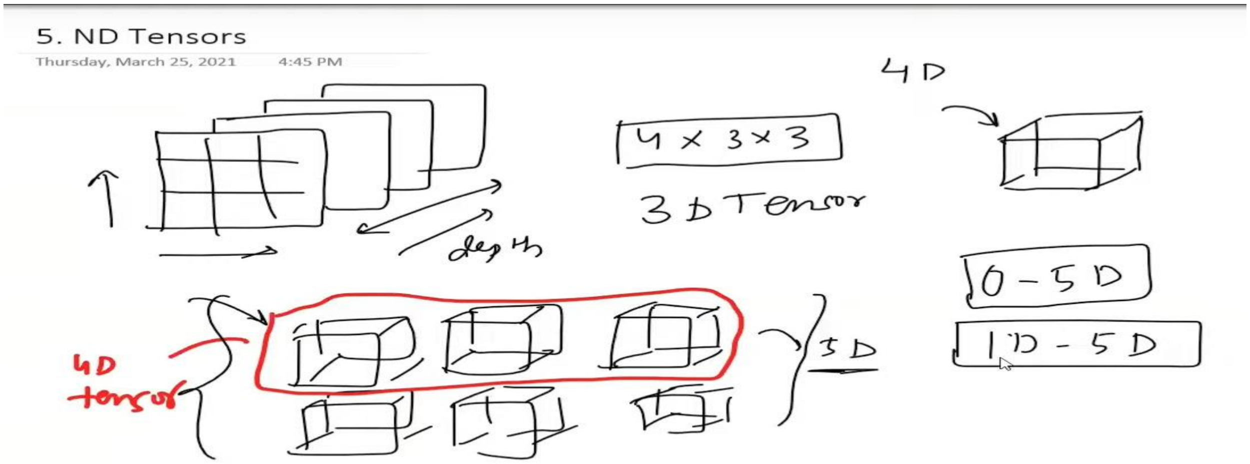
1D-Tensor is similar to 1D- matrix. In one dimensional Tensor have only one row and one column which is known as vector. There is a zero-dimensional tensor also which is known as a scalar.


3. 1D Tensor/Vector

Thursday, March 25, 2021 4:45 PM

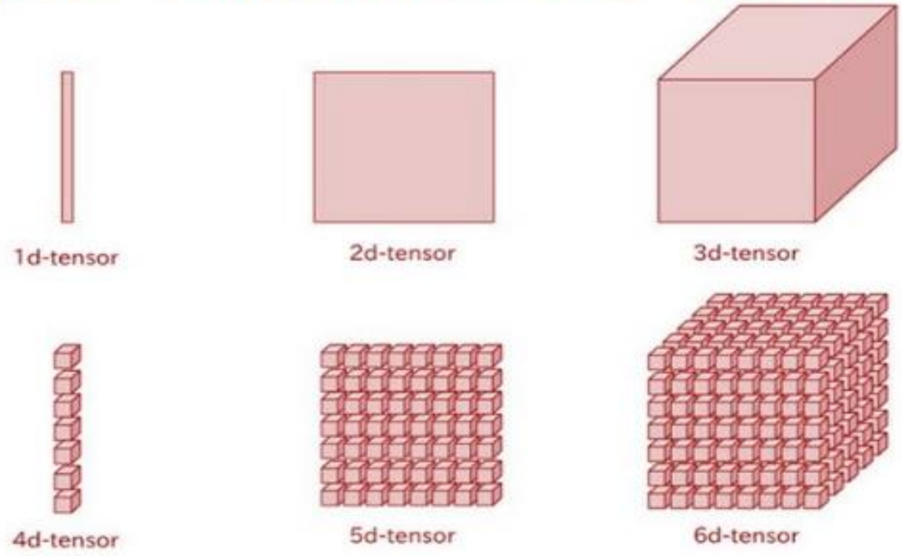
What is 2D tensor?

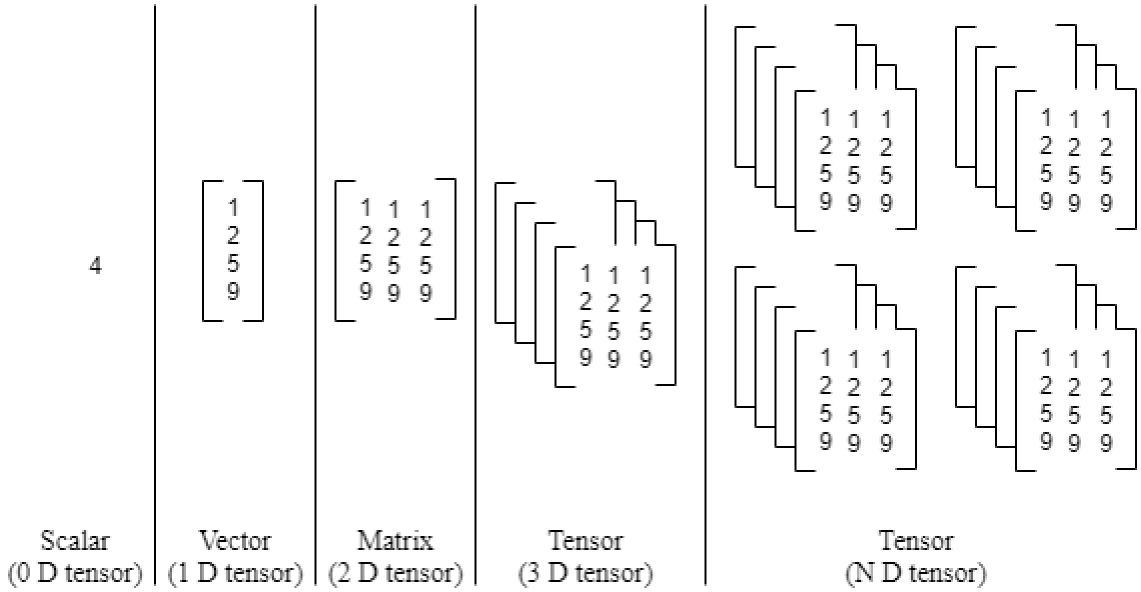
Two-dimensional tensors are analogous to two-dimensional metrics. Like a two-dimensional metric, a two-dimensional tensor also has n number of rows and columns. Let's take a gray-scale image as an example, which is a two-dimensional matrix of numeric values, commonly known as pixels.


5. ND Tensors

Thursday, March 25, 2021 4:45 PM

Simple Tutorial on Tensors

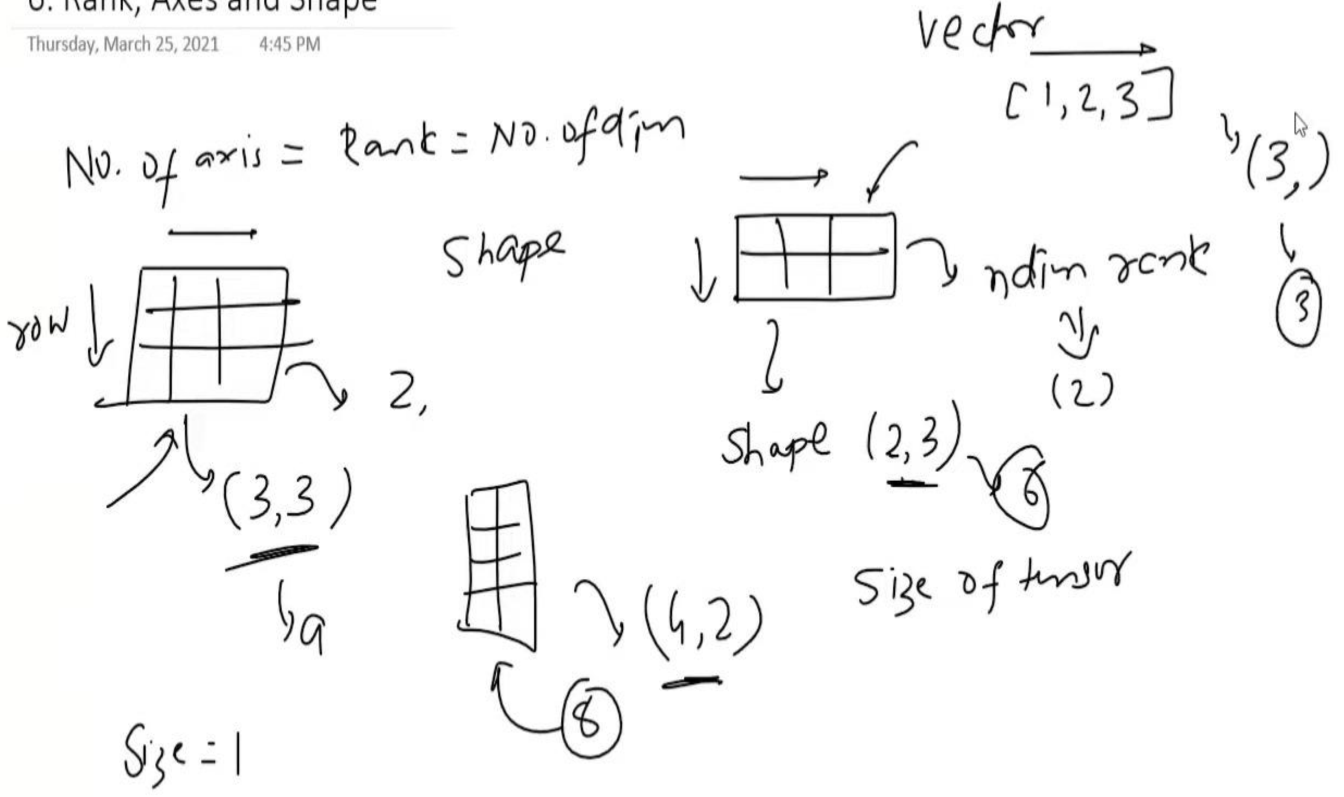




6. Rank, Axes and Shape

Thursday, March 25, 2021 4:45 PM

No. of axis = Rank = No. of dim



7. Example of 1D Tensors

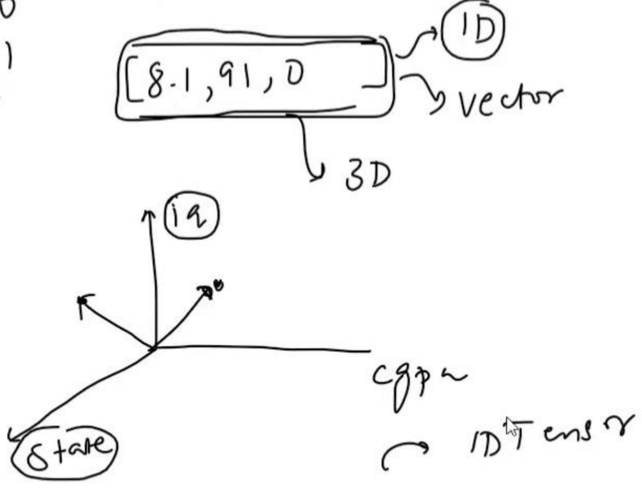
Thursday, March 25, 2021 4:46 PM

Students 10000

WB = 0
KR = 1

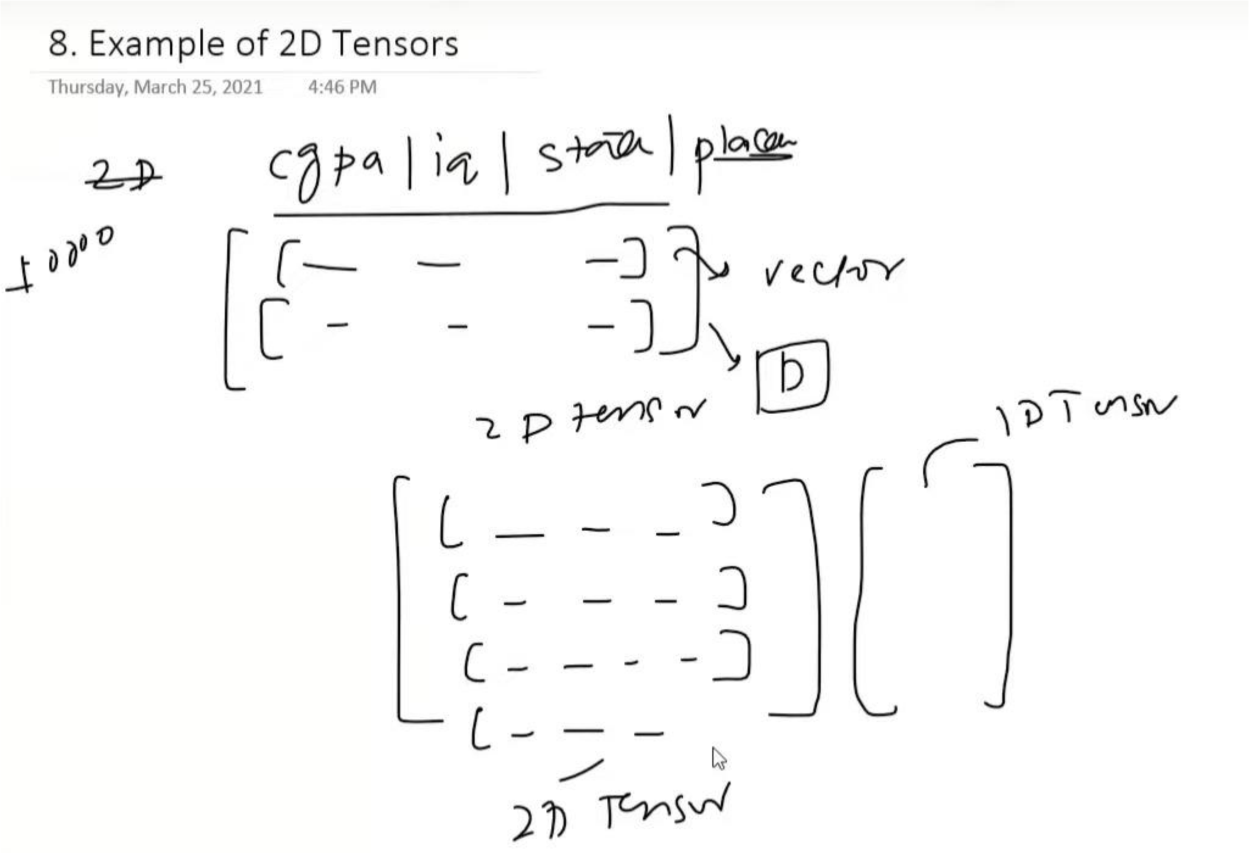
cgpa	iq	state	placement
8.1	91	WB 0	1
[7.2	102	kr]	0
			1
			0
			1
			1
			0

1D Tensor / Vector



[1 0 1 0 1 1 0 1 . . . .]

10000 no.



8. Example of 2D Tensors

Thursday, March 25, 2021 4:46 PM

Data gathering techniques:

CSV file	Comma separated values.
JSON / SQL	JavaScript Object Notation, structured Query Language.
Fetch API	provides an interface for fetching resources (including across the network.
Web scraping	refers to the extraction of data from a website.

CSV: Follow colab notebook

JSON: Follow colab norebook

Fetch API:

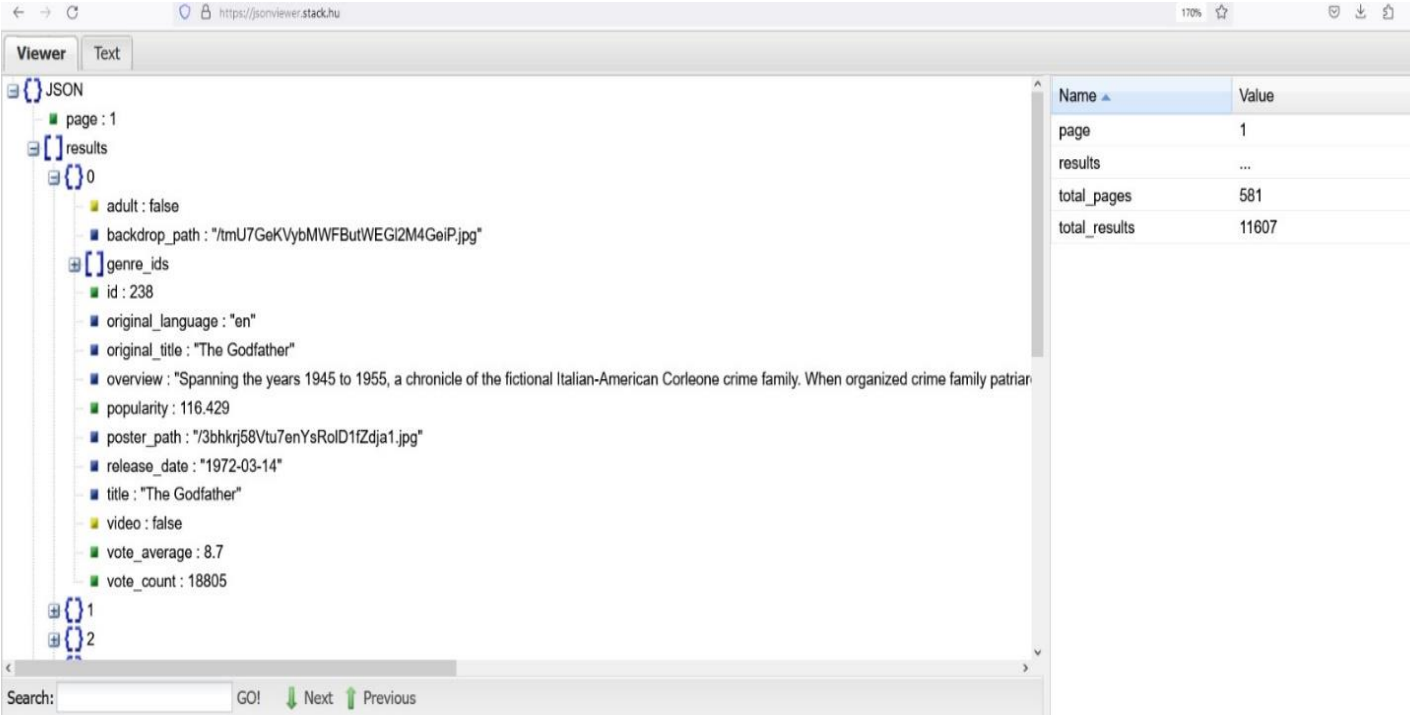
TMDB server: user: nayeem_456 , Pass: 1278

API key: 4b4f1f3b2487c2880ef2924e59e01527

WEB:
https://api.themoviedb.org/3/movie/top_rated?api_key=4b4f1f3b2487c2880ef2924e59e01527

Just change the top_rated word to any other available word to get another data set.

Copy raw data and paste to json online viewer.



```
{} JSON
   page : 1
  [] results
    {} 0
       adult : false
       backdrop_path : "/tmU7GeKVybMWFButWEGl2M4GeiP.jpg"
      [] genre_ids
       id : 238
       original_language : "en"
       original_title : "The Godfather"
       overview : "Spanning the years 1945 to 1955, a chronicle of the fictional Italian-American Corleone crime family. When organized crime family patriar
       popularity : 116.429
       poster_path : "/3bhkrj58Vtu7enYsRolD1fZdja1.jpg"
       release_date : "1972-03-14"
       title : "The Godfather"
       video : false
       vote_average : 8.7
       vote_count : 18805
    {} 1
    {} 2
```

Name ▲	Value
page	1
results	...
total_pages	581
total_results	11607

Follow colab notebook for the rest.

Understanding data and analysis data:

- Ask basic questions (Follow colab notebook)
- EDA univariate analysis
- EDA multivariate analysis
- Pandas profilers

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

Types of EDA:

Univariate: single column

Bivariate: double column

Multivariate: multiple column

Feature engineering:

Feature engineering is the process that takes raw data and transforms it into features that can be used to create a predictive model using machine learning or statistical modeling, such as deep learning.

Types:

Feature Transformation	Feature Construction	Feature selection	Feature extraction
- Missing value imputation - Handling categorical feature - Outlier detection - Feature scaling	Compressing two or more columns to generate new column to get something more efficient for analysis.	Considering only important features.	Combining tow important columns into one column to get benefit from the both. - PCA - LDA

Feature Transformation:

Feature scaling:

Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step. Works through mean standardization.

Standardization	Normalization
	- Min-max scaler - Robust scaler

Feature scaling standardization:

Feature scaling through standardization, also called Z-score normalization, is an important preprocessing step for many machine learning algorithms. It involves rescaling each feature such that it has a standard deviation of 1 and a mean of 0.

- Follow colab notebook.

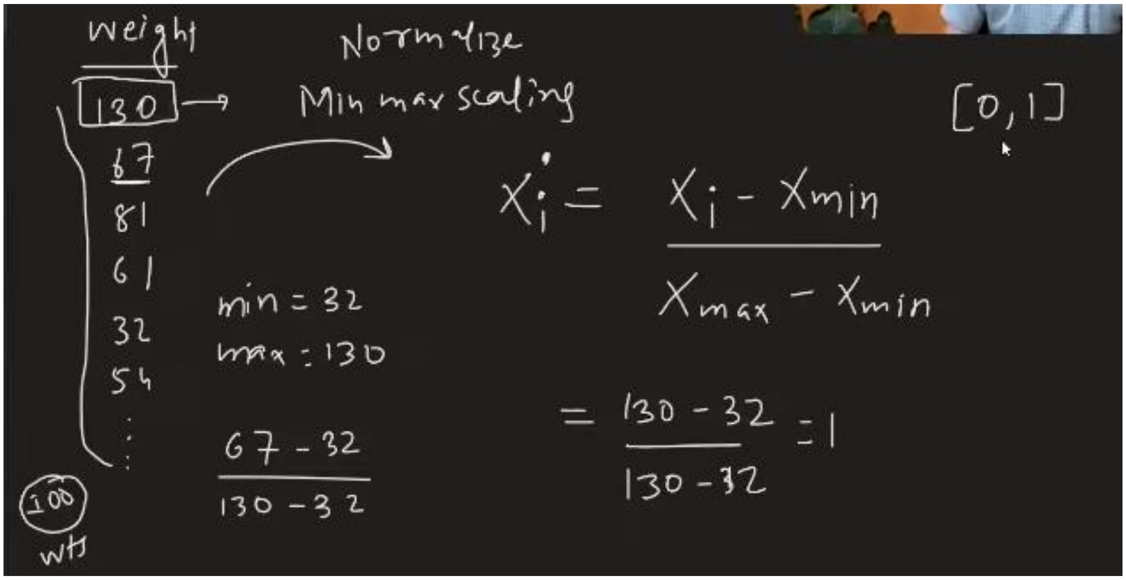
Feature scaling normalization:

Mean Normalization is a way to implement Feature Scaling. What Mean normalization does is that it calculates and subtracts the mean for every feature. A common practice is also to divide this value by the range or the standard deviation.

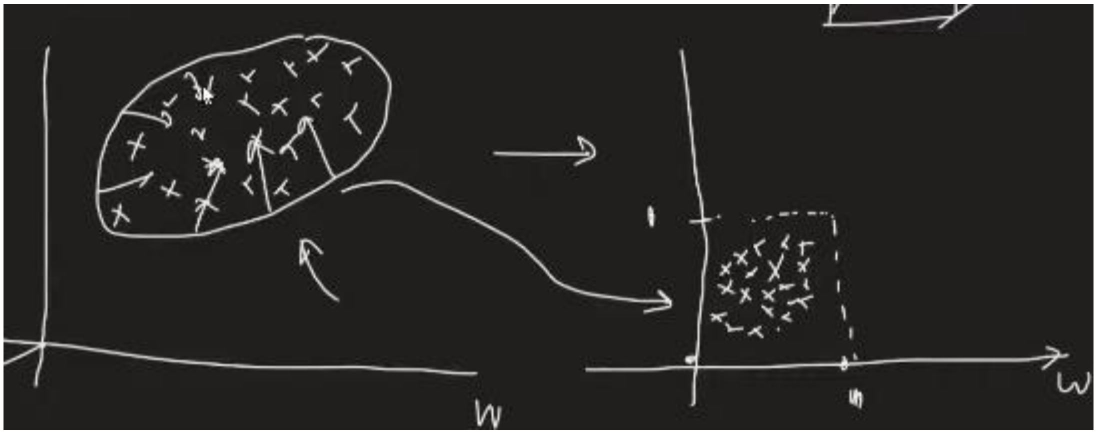
Types:

- Min Max scaling **
- Mean normalization
- Max absolute scaling
- Robust scaling

Min max scaling: values will be in 0 to 1.

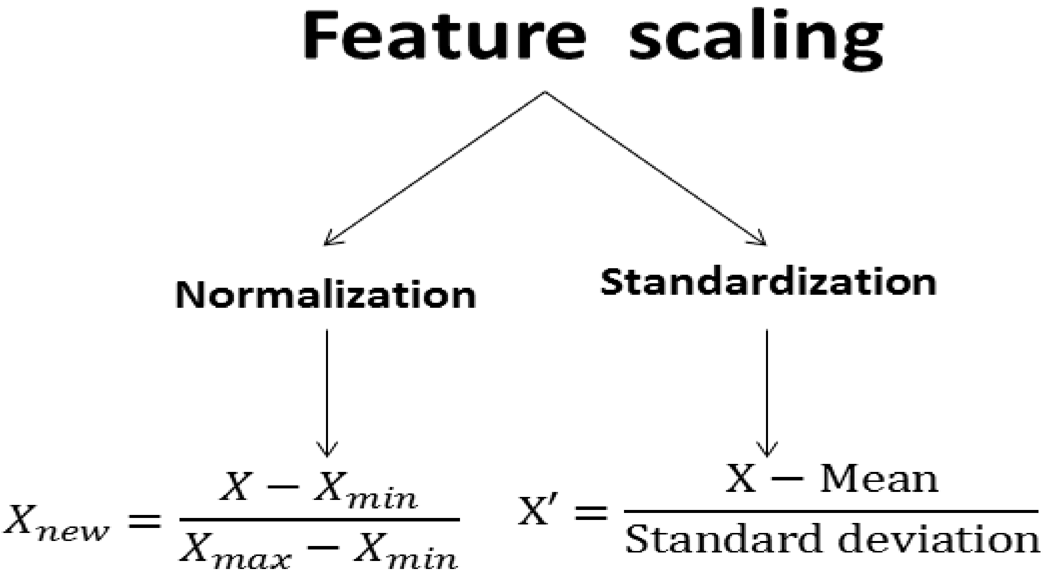


Geometrical representation:



We are pressuring the values to be fit in the 0 to 1 range. If it is 3D then we will pressure it to be in unit cube, if it's more then 3D then we will use unit hyper cube.

- Follow google colab.

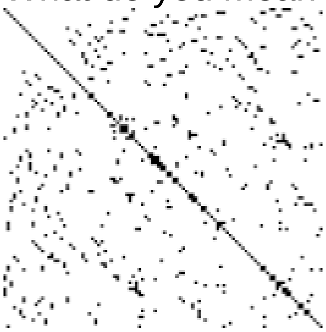


$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad X' = \frac{\mathrm{X - Mean}}{\mathrm{Standard\ deviation}}$$

When to use what:

1. Min-max scaler: when we know the value distribution range.
2. Standard scaling: when no idea about range.
3. Robust: if there is outliers.
4. Max absolute scaling: for sparse matrix.

What do you mean by sparse matrix?



In numerical analysis and scientific computing, a sparse matrix or sparse array is a matrix in which most of the elements are zero.

Encoding categorical data:

Machine learning models can only work with numerical values. For this reason, it is necessary to transform the categorical values of the relevant features into numerical ones. This process is called feature encoding.

Types of categorical data:

1. Nominal data
2. ordinal data

Nominal data, also called named data, is the type of data used to name variables, while ordinal data has a scale or order to it.

Ordinal encoding:

Involves mapping each unique label to an integer value. This type of encoding is really only appropriate if there is a known relationship between the categories. This relationship does exist for some of the variables in our dataset, and ideally, this should be harnessed when preparing the data.

What is meant by one-hot encoding?

One-hot encoding in machine learning is the conversion of categorical information into a format that may be fed into machine learning algorithms to improve prediction accuracy.

when to use one hot encoding?

One-Hot encoding technique is used when the features are nominal(do not have any order). In one hot encoding, for every categorical feature, a new variable is created. Categorical features are mapped with a binary variable containing either 0 or 1.

Column Transformer: is a sciket-learn class used to create and apply separate transformers for numerical and categorical data. To create transformers we need to specify the transformer object and pass the list of transformations inside a tuple along with the column on which you want to apply the transformation.

Pipeline: A machine learning pipeline is a sequence of data processing components that are assembled together to automate the workflow of a machine learning task. It streamlines the process of building, training, and deploying models by organizing and structuring the different steps involved. These steps can include data preprocessing, feature engineering, model selection, hyperparameter tuning, and model evaluation.

Strategy:

- missing values
- one hot encoding
- scaling
- feature selection
- data transformation
- hyperparameter tuning

Mathematical transformation:

In machine learning (ML), mathematical transformations refer to operations or functions applied to the input data to extract relevant features or to prepare the data for further processing by a machine learning algorithm. These transformations are crucial for enhancing the model's ability to learn patterns and relationships within the data.

Common mathematical transformations in ML include:

1. **Normalization/Standardization:**

   - **Normalization:** Scaling the values of features to a standard range (e.g., between 0 and 1).

   - **Standardization:** Transforming features to have a mean of 0 and a standard deviation of 1.

2. **Logarithmic Transformation:**

   - Applying the logarithmic function to the data. This is useful when the data spans several orders of magnitude, and you want to reduce the impact of large values.

3. **Polynomial Transformation:**

   - Creating polynomial features by raising existing features to higher powers. This is useful in capturing non-linear relationships.

4. **Box-Cox Transformation:**

   - A family of power transformations that includes the logarithm as a special case. It is useful for stabilizing variance and making the data more closely approximate a normal distribution.

5. **Fourier Transform:**

   - Used for transforming time-domain data into the frequency domain. This is common in signal processing and analyzing periodic patterns.

6. **Wavelet Transform:**

   - Used for analyzing signals in terms of localized patterns. It's often employed in image processing.

7. **PCA (Principal Component Analysis)

Types of mathematical transforms in sklearn:

- Functional transformation
- Power transformation
- Quartile transformation

Functional transformer:

- Log transformer
- Reciprocal transformer
- Square root transformer

Power transformer:

- Box-cox (can convert into normal distribution from any other form. Works only for positive values but can't work with zero and negative values)

- Yeo johnson (same as box-cox and can work with zero and negative values)

Encoding numerical features: convert numerical data into categorical data.

- Binning and Binarization (Binning or Discretization Binning method is used to smoothing data or to handle noisy data.

In this method, the data is first sorted and then the sorted values are distributed into a number of buckets or bins.)

- Discretization (Discretization simply entails transforming continuous values into discrete categories.

It's a common concept in statistics, often referred to as 'binning' or 'bucketing'. Discretization has numerous merits in machine learning and is easy to execute in Python, as will be explained in detail.)

- Quantile Binning

- KMeans Binning

What are three different types of binning?

- equal-width binning
- equal-frequency binning
- custom binning.

Handling mix variables:

Like cabin no 1A. Here we get numerical and categorical values together.

In two method we can fix it.

- Make then separate column
- Fill null value with numerical and categorical.

Missing value handling:

- Numerical data
- Categorical data
- Missing Indicator
- KNN imputer - Multivariate Imputation

Random imputation: Fill the missing values from other values randomly, using pandas. Sklearn doesn't have this feature.

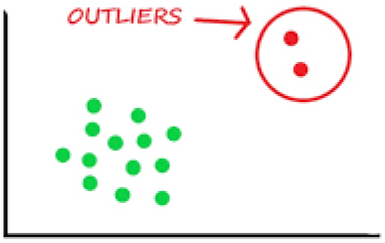
Missing Indicator: A missing indicator is a binary variable (0 or 1) that is used to indicate whether a particular value in the dataset is missing or not. In machine learning, dealing with missing data is a common preprocessing step, and missing indicators are one approach to handle this.

Multivariate Imputation: Multivariate imputation is a technique used to handle missing data in a dataset where there are missing values in multiple variables (features). Instead of imputing each variable independently, multivariate imputation considers the relationships between variables to impute missing values in a way that preserves the overall structure of the data. It is handled by MICE(Multivariate Imputation by chained equations)

Assumptions:

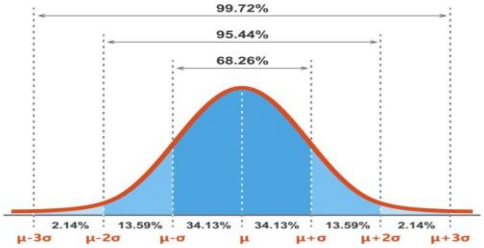
- MCAR (missing completely at random)
- MAR (missing at random)
- MNAR (missing not at random)

Outliers: Outliers are those data points that are significantly different from the rest of the dataset. They are often abnormal observations that skew the data distribution, and arise due to inconsistent data entry, or erroneous observations.



Outlier detection and removal technique: **Z-score method**

Data must be in normal distribution to apply z-score method. Means data will be centered mostly.



Outlier detection and removal technique: **IQR method**

It is applied when data is not normal distribution hence data is in skewed distribution. Using box-plot and IQR.

A skewed distribution: is neither symmetric nor normal because the data values trail off more sharply on one side than on the other. In business, you often find skewness in data sets that represent sizes using positive numbers (sales or assets).

Outlier detection and removal technique: **Percentile method**

Feature construction: is a process which builds intermediate features from the original descriptors in a dataset. The aim is to build more efficient features for a machine data mining task.

- Manually create features
- Split columns and create manual feature

Feature extraction: PCA(principal component analysis)

- PCA (To reduce curse of dimension)
- PCA tries to convert a higher dimension data set to a lower dimension data set.
- Faster execution
- Visualization (10D to 3D / 2D)

Date time handling:

date related task we will handle by order.csv

time related task we will handle by message.csv

Regression matrix: To find out how efficient our regression algorithm is. we use regression matrix.

Types:

- MAE (Mean absolute error)
- MSE(Mean square error)
- RMSE(Root Mean square error)
- R2 Score
- Adjusted R2 Score

Linear regression:

- Simple LR (1 input column and 1 output column)
- Multiple LR (multiple input column)
- Polynomial LR
- Regularization

MAE: In the context of machine learning, absolute error refers to the magnitude of difference between the prediction of an observation and the true value of that observation. MAE takes the average of absolute errors for a group of predictions and observations as a measurement of the magnitude of errors for the entire group.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i |$$

Advantages:

- Output is same unit as y. Input is CGPA and output is package then MAE will give package as output. So communication will be easy.

- Robust outliers means can handle outliers.

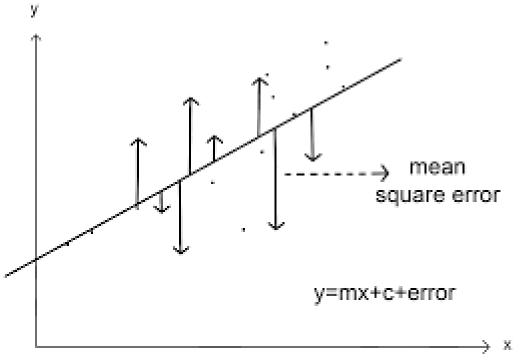
Disadvantages:

- Not differentiable at zero. So optimizations functions fails.

MSE: The Mean Squared Error (MSE) is perhaps the simplest and most common loss function, often taught in introductory Machine Learning courses. To calculate the MSE, you take the difference between your model's predictions and the ground truth, square it, and average it out across the whole dataset.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error
n = number of data points
Y_i = observed values
$\hat{Y}_i$ = predicted values



Advantages: we can use it as lose functions.

Disadvantages: y unit will be in square of it and can’t handle outliers.

RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

R2 Score: The R2 score is a very important metric that is used to evaluate the performance of a regression-based machine learning model. It is pronounced as R squared and is also known as the coefficient of determination. It works by measuring the amount of variance in the predictions explained by the dataset.

Formula

$$R^2 = 1 - \frac{RSS}{TSS}$$

R^2 = coefficient of determination

RSS = sum of squares of residuals

TSS = total sum of squares

$$\text{RSS} = \Sigma\left(y_i - \widehat{y}_i\right)^2$$

Where: y_i is the actual value and, $\widehat{y}_i$ is the predicted value.

$$\text{TSS} = \Sigma\left(y_i - \overline{y}\right)^2$$

Where: y_i is the actual value and $\overline{y}$ is the mean value of the variable/feature

Adjusted R2 Score: For multiple linear regression.

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where

R^2 Sample R-Squared

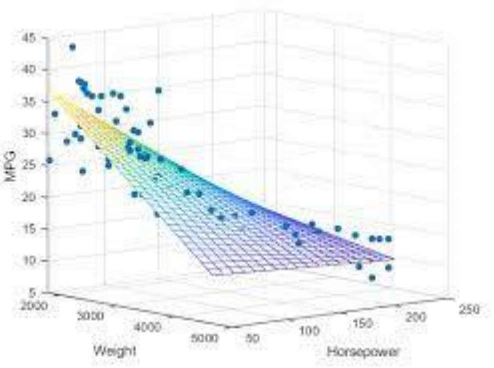
N Total Sample Size

p Number of independent variable

Multiple linear regression:

Multiple input column and single output column.

It can be 3D or more. In 3D data we plot a plane line and above of 3D we draw hyperplane.



$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ldots + \beta_i X_i}$$

Y : Dependent variable
β_0 : Intercept
β_i : Slope for X_i
X = Independent variable

Regularization: refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Types:

- Ridge
- Lasso
- Elastic net

Ridge/L2 regression: adding lambda with coef.

$$Cost\ (W) = RSS(W) + \lambda * (sum\ of\ squares\ of\ weights)$$

$$= \sum_{i=1}^{N} \left\{ y_i - \sum_{j=0}^{M} w_j \, x_{ij} \right\}^2 + \lambda \sum_{j=0}^{M} w_j^2$$

Lasso regression: adding lambda with absolute value of coef.

$$\sum_{i=1}^{n} (y_i - \sum_{j} x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^{p} | \beta_j |$$

Elastic-Net regression: combination of ridge and lasso.

$$\frac{\sum_{i=1}^{n} (y_i - x_i^J \hat{\beta})^2}{2n} + \lambda \left(\frac{1 - \alpha}{2} \sum_{j=1}^{m} \hat{\beta}_j^2 + \alpha \sum_{j=1}^{m} |\hat{\beta}_j| \right)$$