ML with Azure Schlorship

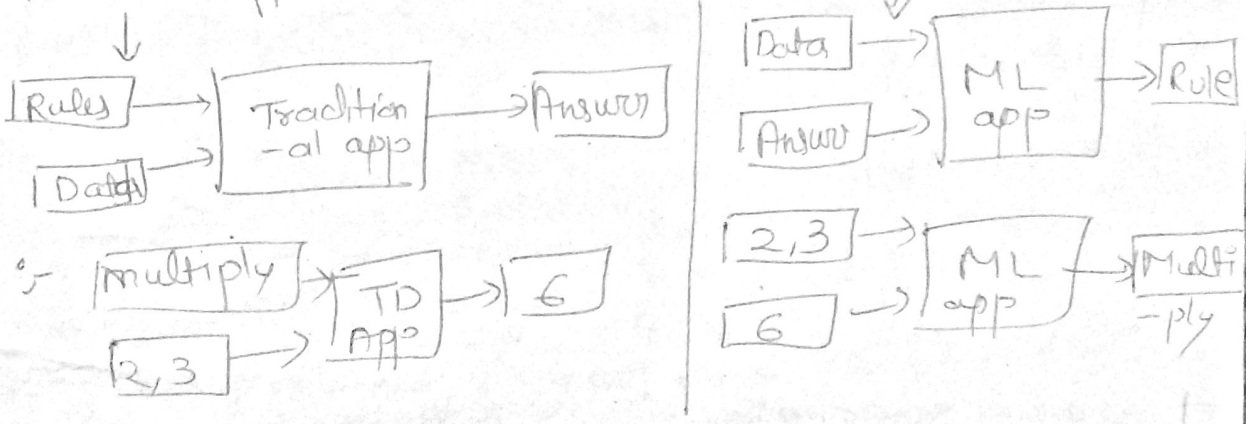
Lesson 1: Introduction to ML

1) What is ML?

→ It is a data science technique used to extract pattern from data, allowing computers to identify data & forecast future outcomes, behaviour & trend.

2) Application :- It is used in
1) NLP 2) Computer Vision 3) Data Anylysis
4) Decision Making
i) Automate the recognition of disease
ii) Recommended next best action for individual care plans
iii) chatbot
iv) Identify the next best action for the customer

3) Tradition approach Vs ML approach



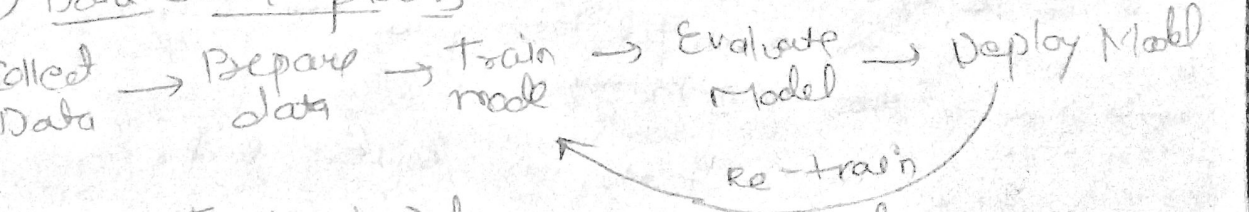
4) Diff Between

AI
→ A board term that refers to computers thinking more like human

ML
→ A subcatory of AI that involves learning from data without being explicitly programmed

Deep learning
→ A subcatory of ML that uses a layered neural network architecture originally inspired by the human brain

5) Data Science process



6) Common Types of Data Used in ML
- Numerical
- Time series

- Categorical
- Text
- Image

→ Its all numerical in the end

7) i) Tabular data : that is data arranged in a data table, it consist row, column & cell

8) ii) Scaling data :- It mean to transform it so that the value fit within some range or scale such as 0-100 or 0-10 or 0-1.
Eg: It used in image - RGB whose range is 0-255 & we can scale down it into range of 0-1. The scaling process not affect algorithm output. It speed up the training process, because now algorithm handle nos. less than or equal to 1

- Approach of Scaling

1) Standardization
→ Mean = 0 & Variance = 1
→ $(x-\mu)/\sigma$

2) Normalization
→ range [0, 1]
→ $(x - x_{min})/(x_{max} - x_{min})$

9) Encoding categorical data
i) Ordinal encoding 2) one hot encoding

→ In ordinal encoding, we convert categorical data into integer code range 0 to (no. of category - 1)

Eg:-

Colour	encoding		Colour
Blue	→ 0	⇒	0
green	→ 1		1
red	→ 2		2

But we can't use it becoz it make confussion that red is more powerful than green i.e why we get 2 no to red & 1 to green

→ One hot encoding :- It this we convert each categorical value into a column.

Eg:-

SKU	Red	Green	Blue
00721	0	0	0
00231	0	1	0
00441	1	0	1

- It increases no. of columns, it is the dimensionality

10) Image data
It is form of pixel & contain RGB or it can be different
- In grayscale image :- each pixel represent single no, in range of 0-255. (0 for black, 255 for bright; depth = 1
- In coloured image :- each pixel represent as vector of three for channel red, blue, Green (128, 0, 128), depth = 3

- Encoding of Image
(horizontal position of pixel, vertical position of pixel, color of each pixel)
Size of image = (height * width * depth)

11) Text Data
- Encoding of Text data by normalization (process of transforming a piece of text into a canonical form),
- We use lemmatization (eg of normalization) to reduce multiple inflections to that single dictionary form.

Eg:-

original word	Lemmatized word
is	be
are	be
am	be

- Stop word : which are high frequency word that are un necessary during the analysis. Eg:- If we write "which cookbook has the best pancake recipe" in search engin, the word "which" & "the" are less relevant than "cookbook", "pancake" & "recipe"
- tokenized are split each string of text into a list of small part of token.

Eg:

Original text	Normalized text
The quick fox	[quick, fox
The lazy dog	[lazy, dog]

- Vectorization : After apply normalization, next step to encode it. It turn a text into Vector
1) TF-IDF approach :- It give less importance to words that contain less inform & are common in document eg:- "this" & "the" & higher importance to word that contain relevant information & appear less. It assign in weighted form

Eg:-

quick	fox	lazy	dog
0.32	0.41	0.23	0.56

- Algorithm understand by if two vectors are closer to each other, we can say text represent by two vector have similar meaning.

Eg:-	quick	fox	lazy	dog
[quick fox]	0.32	0.23	0.0	0.0
[lazy dog]	0.0	0.0	0.12	0.23

- Whole pipeline to convert text
  1) normalization
  2) Feature Extraction (Vectorization)
  3) Algorithm

12) Two Perspective on ML

13) Computer science vs Statical perspective

i) Input feature & output feature.

ii) output = Program (Input Feature)

iii) A group of input variable is called input vector

iv) we use program from input data to make prediction

- Eg:- we if take data in tabular form, row are entities & colum are feature describe the property of entity

v) Output variable = f(Input variable)
Or
Dependent variable = f(Independent variable)
Or
$y = f(x)$

Eg:- price of a house is dependent variable that depend upon some variables like location, size

15) The tools for ML

- Libraries :-
  1) Scikit-learn
  - Tensorflow
  - Pytorch
  - Keras
- Development Enviroment -
  1) Jupyter notebook
  2) Azure notebook
  3) Azure Databricks
  4) visual Studio Code
  5) Visual Studio
- Cloud Service
  → Microsoft Azure ML

16) Libraries for ML

1) Core Framework & Tools - python, pandas, Numpy, jupitn
2) ML & Deep learn - learn, spark, tensorflow, Keras, pytorch

8) Visualization - seaborn, matplotlib, bokeh, plotly

17) Cloud service in ML

Core asset management
- → Dataset
- → Experiments/Runs
- → Pipeline
- → Models
- → Endpoint

Resource management
- → Compute
- → Environments
- → Datastores

- Dataset :- Define, version & monitor datasets used in ML
- Experiment/Runs - Organize ML workload & keep track each run task executed through the service.
- Pipeline :- Structured flows of tasks to make complex ML flow.
- Models :- Model registry with support for versioning & deploy -ment to production
- Endpt :- Expose real-time endpoints for scoring as well as pipeline for advanced automation.
- Compute :- Manage compute resource used by ML task.
- Environment :- Template for standarized environments used to create compute resources
- Datastore :- Data source connect to the service environment.

18) Model vs Algorithm

- Model are the specific representation learned from data
- Algorithm are the process of learning.

Model = Algorithm(Data)

20) Linear Regression :-

- Linear Regression is a algorithm that uses a straigh line (or plane) to describe relationship b/w variables.

1) Simple linear Regression

$y = mx + c$ → Intricept
↳ slope

in ML, c is bias.

$y = \beta_0 + \beta_1 * c$ (simple linear regression)

2) Multiple Linear Regression

$y = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \cdots + \beta_n c_n$

Instead of line a plane is form

- To train a linear regression model, we have to learn coeffs & bias that best fit the data & minimize the error during prediction. To calculate error, we use cost function.
- Most commonly used cost function for linear regression is root mean square error (RMSE)

i) The relation b/w input & output must be linear, if it is not linear we have to transform it.

ii) To remove highly correlated input variable & perform a correlation check among input variable.

iii) Gaussian distribution.

iv) Rescale data : one should first normalize & standardize the data

v) Remove noise : Linear regression is very sensitive to noise & outliers in the data.

These are the things imp for preparation data

- formula for slope

$$m = \frac{\sum_{i=1}^{m}(x_i - \text{mean}(x) \times (y_i - \text{mean}(y)))}{\sum_{i=1}^{m}(x_i - \text{mean}(x))^2}$$

- To get intercept,

$$x = \text{mean}(y) - m \times \text{mean}(x)$$

- root mean squared error

$$= \sqrt{\frac{\sum_{i=1}^{n}(p_i - y_i)^2}{n}}$$

- Sklearn & Pytorch library these inner working for linear reg is already present.

25. Learning function

$y = f(x) + e$ ← irreducible error (it is independent of $f(x)$ or input) (it is different by model error)

- f is done due to don't have enough data or don't have enough data feature.

→ Model error measure how much the prediction made by the model is different from true output.

26. Parametric Vs Non-Parametric

Based on assumption about shape & structure of func'n they try to learn, ML algorithms can be divided into two category:

1) Parametric

i) Simplify the mapping to a known functional form (eg linear regression)

ii) Benefit :- Simples, Faster, Less Training Data

iii) Limitation :- Highly constrained, limited complexity, poor fit

2) Non-parametric

- Not making assumption regarding the form of mapping b/w input & output (eg:- K Mean Neighbour (KNN))

ii) Benefit :- High Flexibility, power, high performance

iii) Limitation :- More Training Data, Slower, overfitting training data

27. Classical ML vs Deep Learning

- all deep learning algorithm are ML but not all ML algorithm are deep learning algorithm
- Deep learning algorithm are based on neural network & the classical ML algorithm are based on classical mathematical algorithm

Deep Learning :- advantages :-

1) Suitable for high complexity problem
2) Better accuracy, compared to classical ML
3) Better support for big data
4) Complex feature can be learned

Disadvantage

1) Difficult to explain trained data
2) Require significant computational power

Classical ML advantage

1) More suitable for small data
2) Easier to interpret outcome
3) Can run on low-end machine
4) Cheaper to perform
5) Does not require large computation power

Disadvantage

1) Difficult to learn large datasets
2) Require feature engineering
3) Difficult to learn complex func n

38. Approach to Machine Learning

1) Supervised learning
2) Unsupervised "
3) Reinforcement "

Supervised learning

Learn from data that contains both the inputs & expected output (eg labeled data). Common types are

i) classification :- output are categorical
ii) Regression :- Output are continuous & numerical
iii) Similarity learning :- Learn from eg using a similarity func that measure how similar two objects are
iv) Feature learning :- Learn to automatically discover the representations as features from raw data.
v) Anomaly detection - A special form of classification, which learn from data labelled as normal / abnormal.

Unsupervised learning

learn from input data only & find hidden structure in input data

1) Clustering : Assign entities to clusters or groups
2) Feature learning : Features are learned from unlabled data
3) Anomaly detection : Learn from unlabeled data using assumption that majority of entities are normal.