MACHINE LEARNING

UNIT-1

(1) ~~Artificial Intelligence~~:-

(1) Machine Learning:

- Machine Learning is a subset of AI
- Machine Learning allows machines to learn the data without being explicitly programmed

$$E * T = P$$

E stands for Experience

T stands for Task

P stands for performance

(2) Types of Machine Learning:

There are mainly three types of Machine Learning.

(i) Supervised Learning

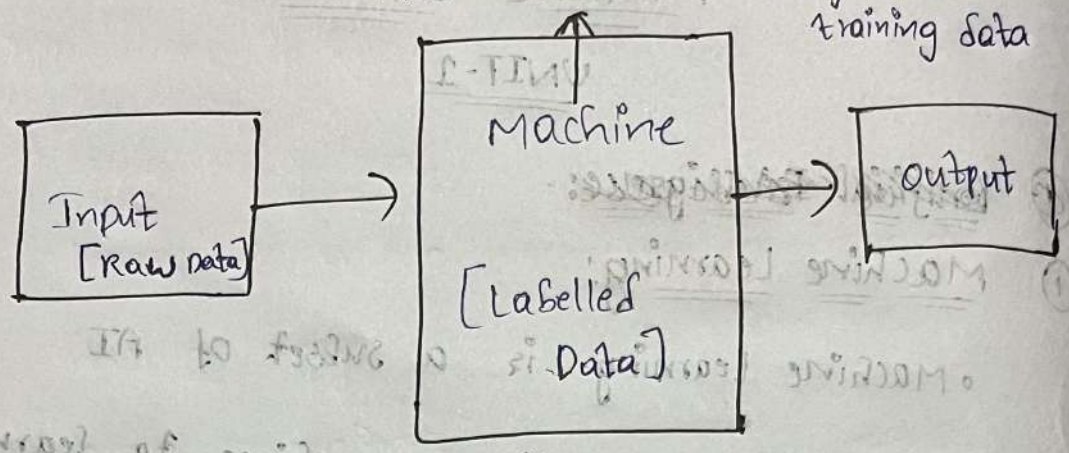
(ii) Unsupervised Learning

(iii) Reinforcement Learning.

(i) Supervised Learning:

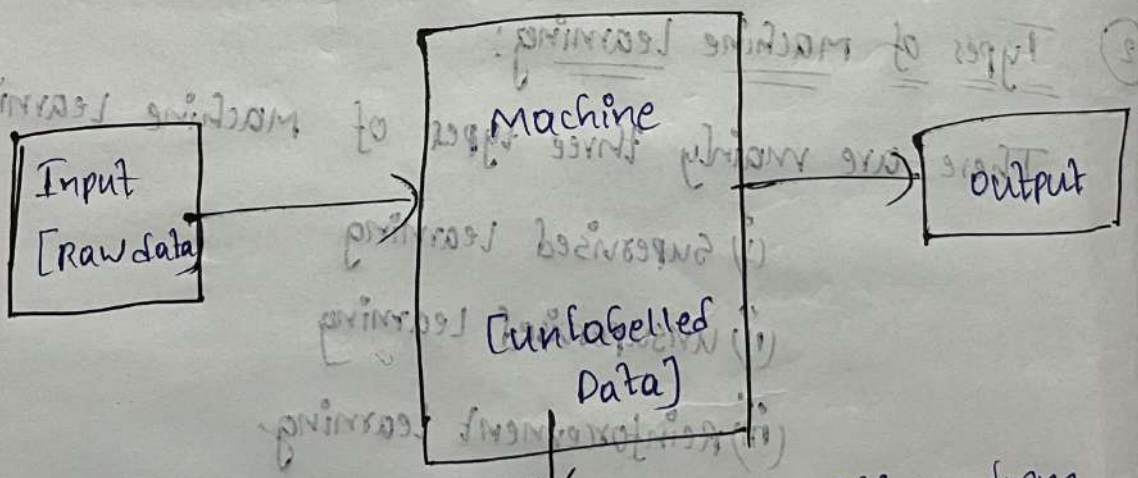
- Supervised learning is a technique for training the machine by using labelled data
- This technique maps inputs to the outputs

Ex: Identifying CPU and Ram by using labelled data



(ii) unsupervised Learning:

- unsupervised Learning is a technique for training the machine on unlabelled data
- This technique is used to identify patterns and relationship in the data

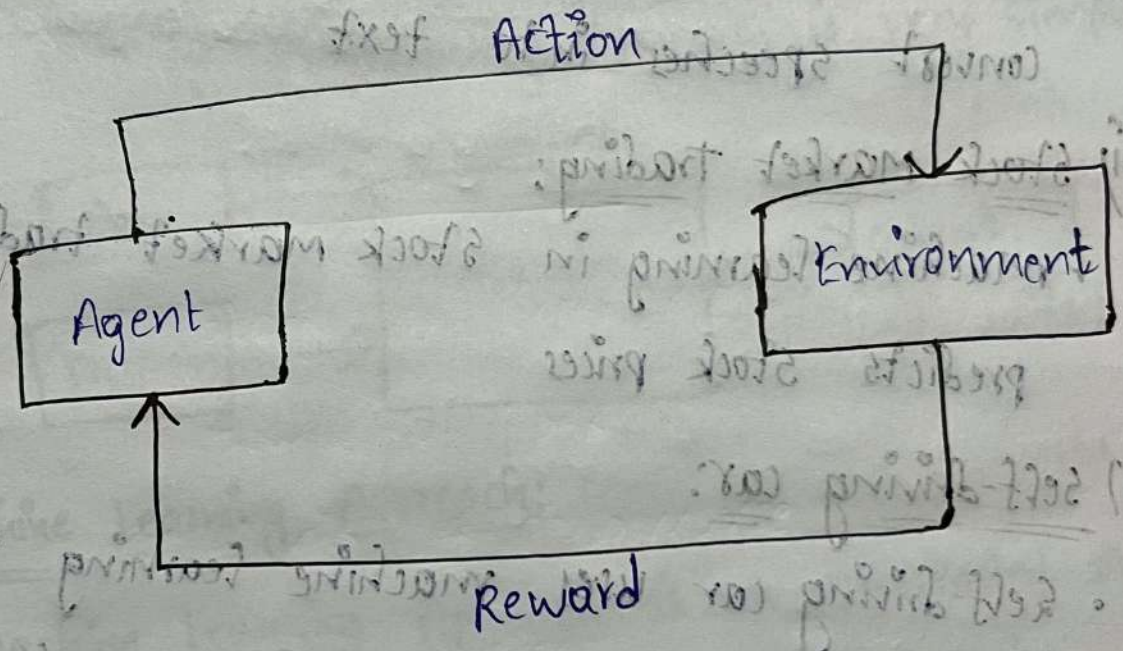


Machine do not learn the things from training data.

Ex: Identifying CPU and Ram based on their patterns and relationship.

(iii) Reinforcement Learning:

- Reinforcement learning is a technique for training the machine by using agent and an Environment
- This technique is used to produce actions & rewards in a given situation



Ex:

Chess game

(3) Applications of Machine Learning:

(i) Image Recognition:

- Image Recognition uses machine learning to identify images in photos or videos

(ii) Speech Recognition:

- Speech Recognition uses machine learning to convert speeches into text

(iii) Stock Market Trading:

- Machine learning in stock market trading predicts stock prices

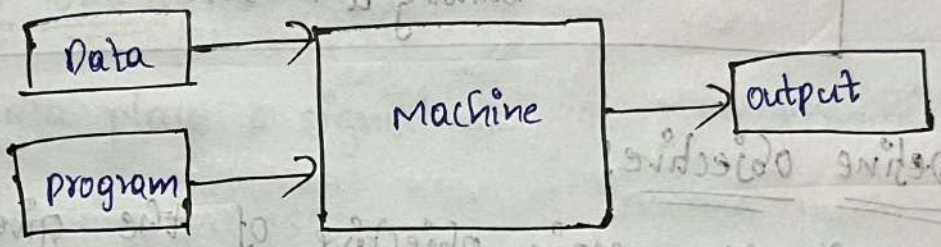
(iv) Self-driving car:

- Self-driving car uses machine learning to drive the car without any human efforts

(4) Difference between Traditional programming and Machine Learning Approach

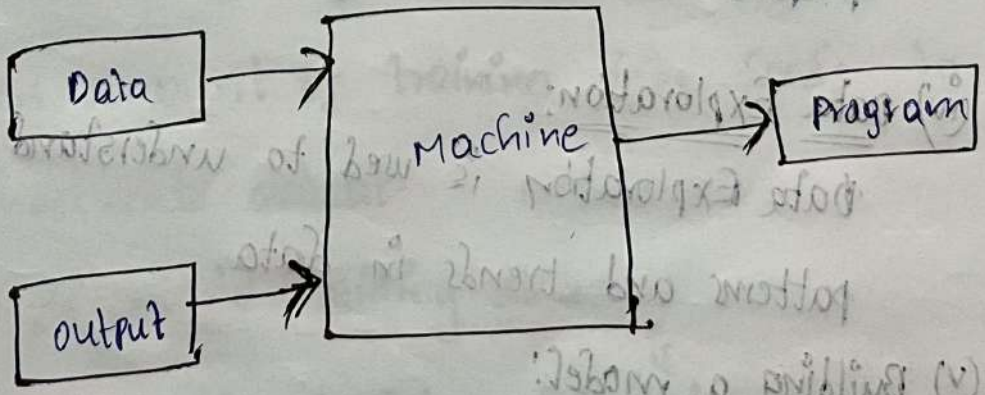
Traditional programming Approach:

- Traditional programming is a manual process.
- Data + program → output
- Generates output by explicitly programmed
- Traditional programming Approach is very simple

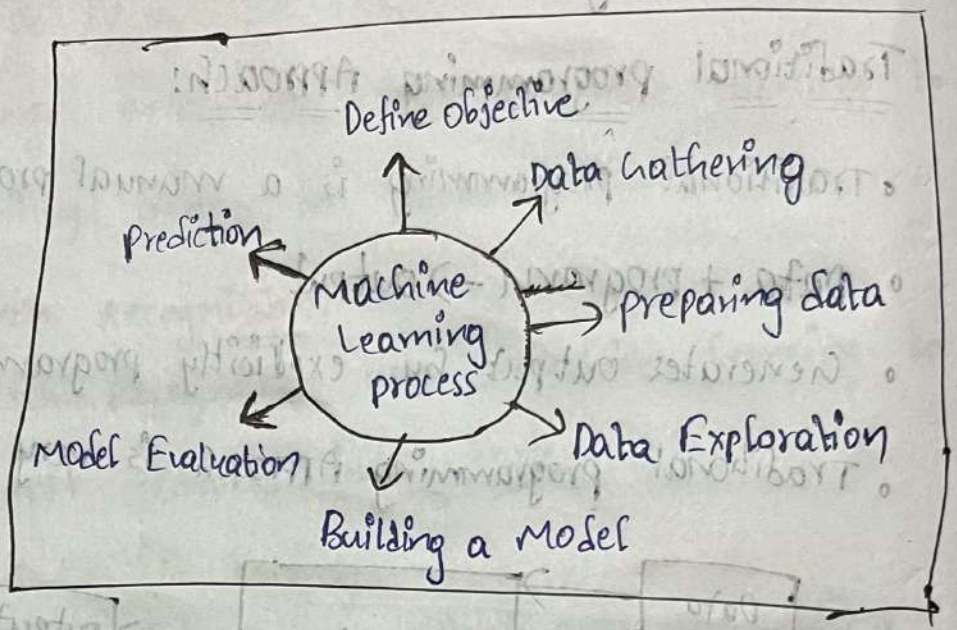


Machine Learning Approach:

- Machine Learning Approach is an automated process
- Data + output → program
- Learn's the data without being explicitly programmed
- Machine Learning Approach is very complex



(5) Machine Learning process:



(i) **Define objective:**

Identify the main objective of the given problem

(ii) **Data Gathering:**

collect the data that solves the given problem

(iii) **Preparing Data: [Data preparation]**

collected data have lot of inconsistencies and prepare the data in right format

(iv) **Data Exploration:**

Data Exploration is used to understand the patterns and trends in data

(v) **Building a model:**

patterns and trends in Data Exploration are used to build the Machine Learning Model.

(vi) Model Evaluation:

Model Evaluation is used to check the efficiency of a model.

(vii) prediction:

After the Evaluation, the model is used to make predictions

---

(6) Main challenges of Machine Learning:

(i) poor quality of data:

- Data plays a significant role in machine learning process
- poor quality of data leads to wrong predictions.
- Good quality of data is essential for Machine Learning process.

(ii) Not enough training data:

- Machine learning requires lot of training data to produce accurate output
- Less amount of training data leads to inaccurate output.

(iii) Non-representative Training data:

- It means the training data is incomplete and biased and produces inaccurate output

- Ensuring complete and unbiased training data in machine learning process.

(iv) underfitting of training data:

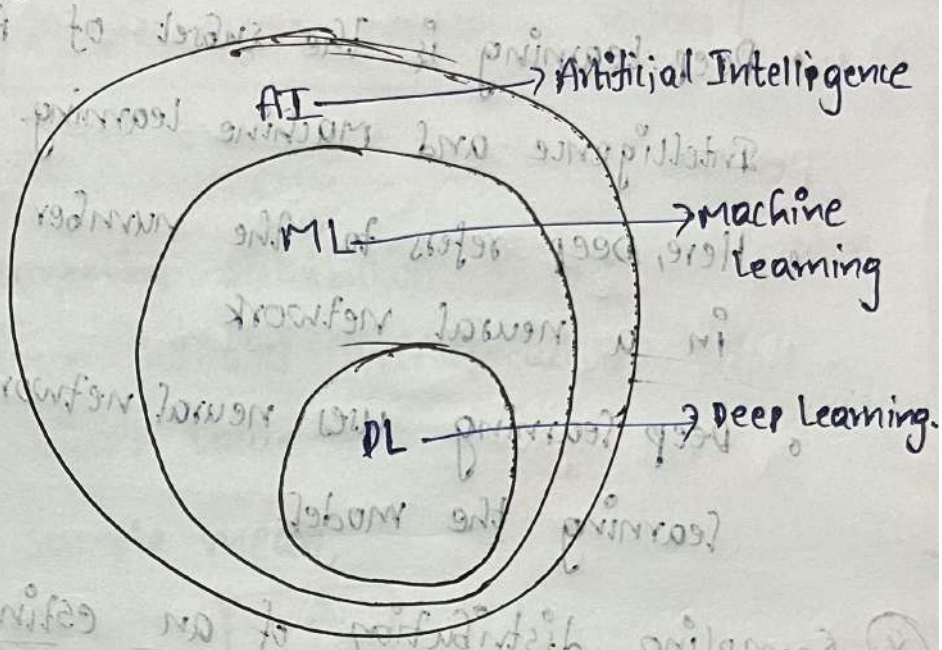
- underfitting is the opposite of overfitting.
- This model is very simple to learn from the training data

(v) overfitting of training data:

- overfitting is the opposite of underfitting.
- This model is very complex to learn from the training data

- underfit produces inaccurate results for both training data & test data
- overfit produces accurate results for training data but not for test data

(7) Differences between Artificial Intelligence, Machine Learning and Deep Learning.



(i) AI:

- AI stands for Artificial Intelligence.
- Artificial Intelligence consists of both Deep learning and machine learning.
- AI deals with making intelligent machines or computers to perform human-like activities

(ii) ML:

- ML stands for Machine learning.
- Machine learning is an subset of Artificial Intelligence.
- Machine learning allows machines to learn the data without being explicitly programmed.

(iii) DL:

- DL stands for Deep Learning.
- Deep learning is the subset of Artificial Intelligence and Machine learning.
- Here, Deep refers to the number of layers in a neural network
- Deep learning uses neural network for learning the model.

(8) Sampling distribution of an estimator

- Sampling Distribution of an estimator is the probability Distribution of an estimator
- Sampling Distribution of an estimator derives multiple samples from the same population.
- Sampling Distribution of an estimator helps in estimating the population parameters like mean, variance.

(i) mean:

- In Sampling Distribution of an estimator, mean indicates the expected value of

the estimator.

Here,

Sample mean = population mean

(ii) **variance:**

In sampling Distribution of an Estimator, variance indicates the spread of the estimator's value

For sample mean,

$$\frac{\sigma^2}{n}$$

where,

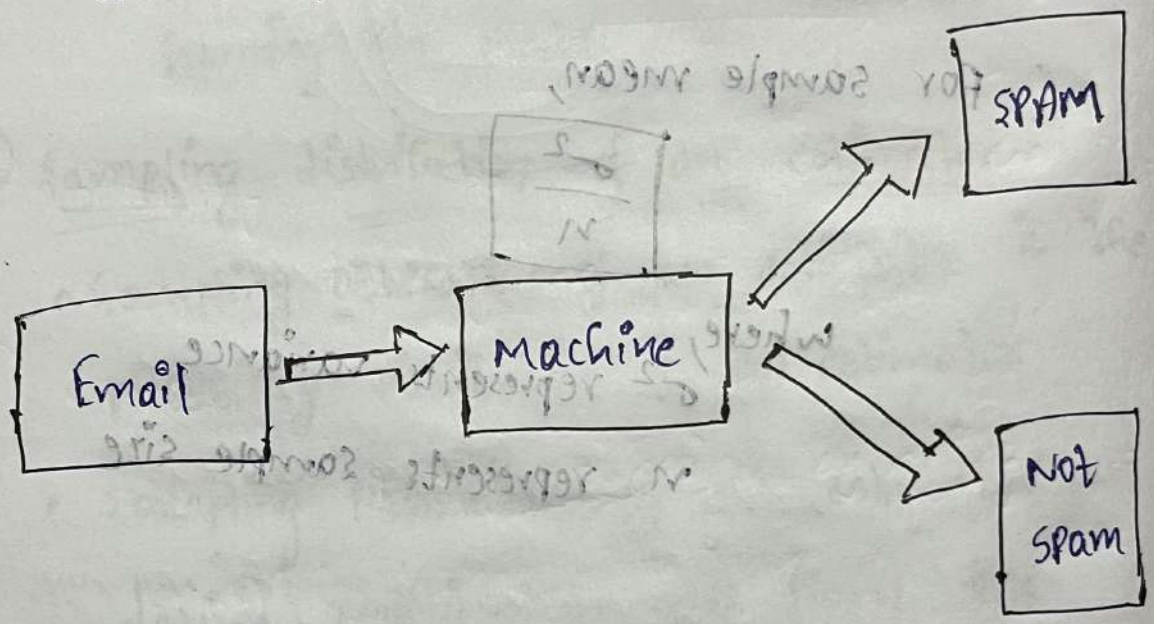
σ^2 represents variance

n represents sample size

(9) Spam Detection as a supervised or unsupervised learning problem

(A) • Spam Detection is a supervised learning problem because it trains the machine by using labelled data

• Here, emails are marked as either "spam" or "not spam"



- if no labelled data is available, an unsupervised learning problem groups the emails based on their patterns and relationships
- unsupervised learning problem do not explicitly identify the spam.

∴ Supervised learning problem is better at spam Detection when compared to unsupervised learning problem.

(10) Training & Test Loss:

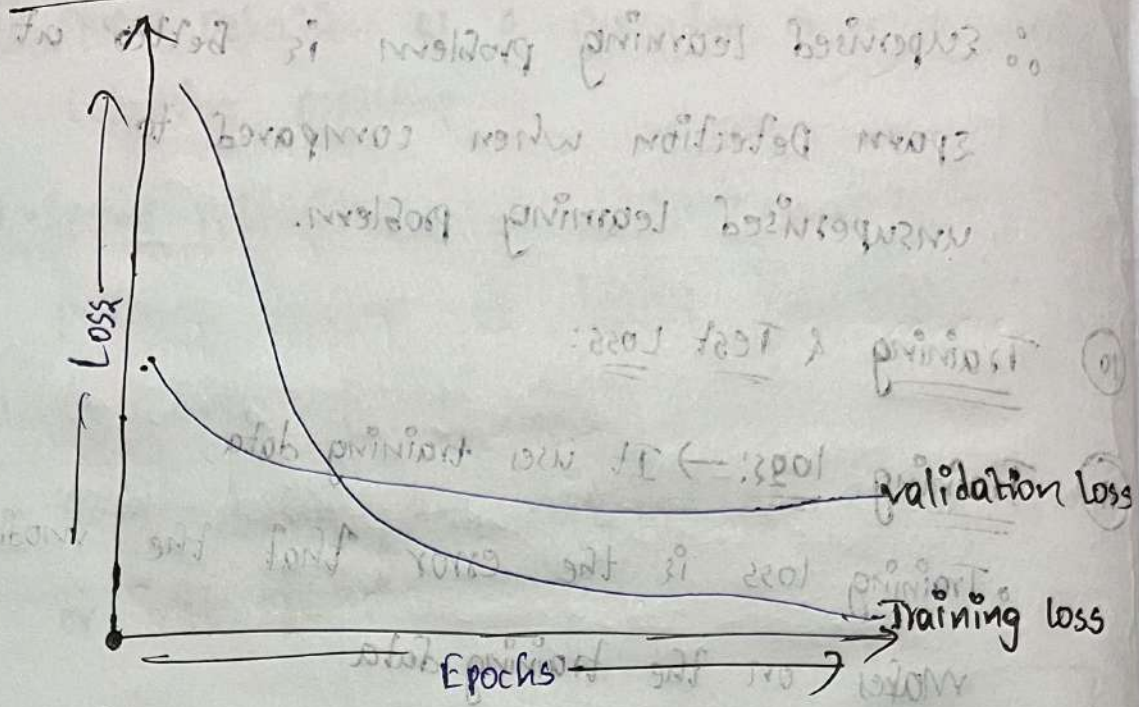
(A) Training loss: → It uses training data

- Training loss is the error that the model makes on the training data
- Lowering the training loss means the model is fitting the training data better

Test loss: → It uses unseen test data

- Test loss is the error that the model makes on the testing data
- Higher the testing loss means the model is overfitting the training data

- Training loss is used for optimizing the model during training
- Test loss is used for measuring the generalization performance



(11) Estimating the loss and Accuracy of the ML model for training and test cases

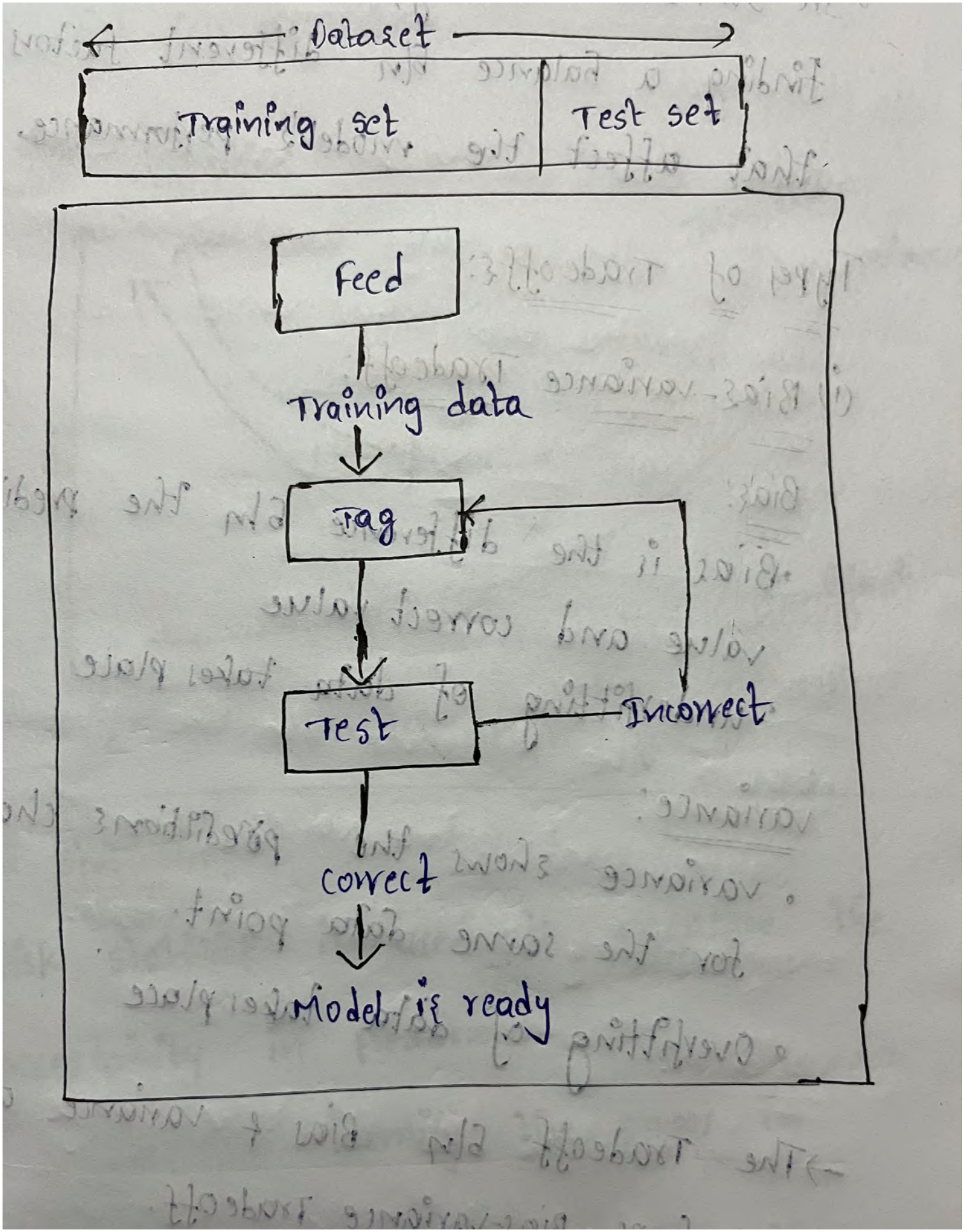
(A) (i) Loss:

- Loss is typically estimated by using a loss function.
- Loss calculates the error b/n the predicted values & actual values.

(ii) Accuracy:

Accuracy is the percentage of correct predictions made by the model

$$\text{Accuracy} = \frac{\text{No of correct predictions}}{\text{Total no of predictions}} \times 100$$


Dataset
Training set
Test set
Feed
Training data
Tag
Test
Incorrect
correct
Model is ready

(12) Tradeoffs in statistical Learning

- In statistical Learning, tradeoffs means finding a balance b/n different factors that affect the model's performance.

Types of Tradeoffs:

(i) Bias-variance Tradeoff:

Bias:

- Bias is the difference b/n the prediction value and correct value
- underfitting of data takes place

variance:

- variance shows the predictions change for the same data point.
- Overfitting of data takes place

→ The Tradeoff b/n Bias & variance are called as Bias-variance Tradeoff.

(ii) complex models vs simple models:

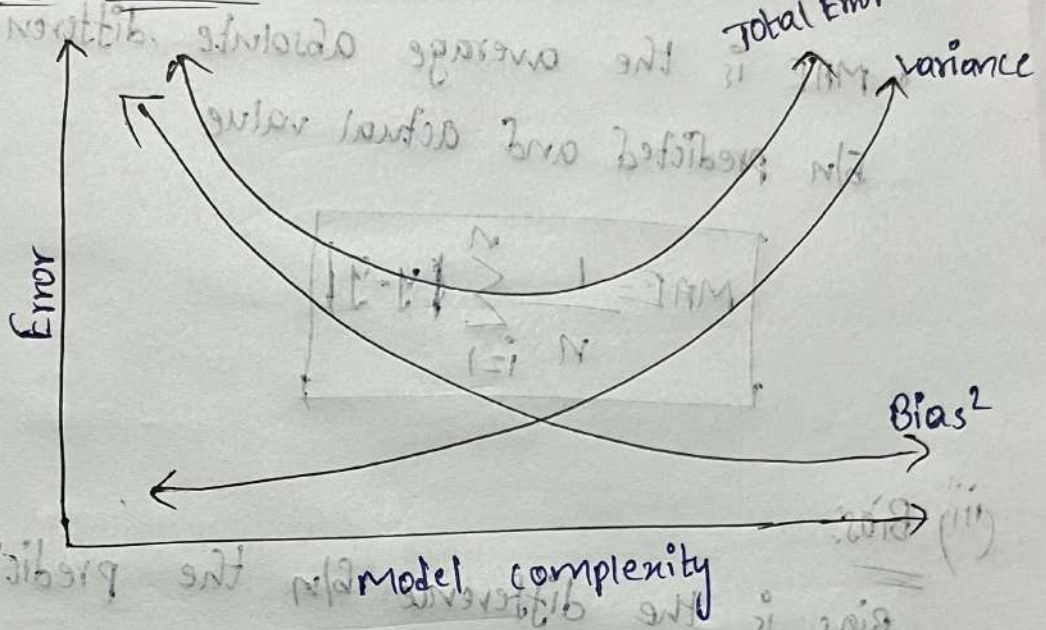
complex models:

- complex models are hard to understand but very accurate

Simple models:

- Simple models are easy to understand but less accurate.

(12) Bias Variance



(13) Risk Statistics:

- Risk statistics is used to measure the uncertainity in predictions
- Risk statistics identify the errors in predictions

(i) MSE:

- MSE stands for Mean squared Error
- MSE is the average squared difference bn predicted and actual value

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y-\bar{y})^2$$

(ii) MAE:

- MAE stands for Mean Absolute Error
- MAE is the average absolute difference b/n predicted and actual value

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y-\bar{y}|$$

(iii) Bias:

- Bias is the difference b/n the predicted value and actual value
- underfitting of data takes place

(iv) variance:

- variance shows the prediction change for the same data point
- Overfitting of data takes place

(v) Risk:

- Risk combines both Bias and variance to estimate the error.

$$Risk = E[L(y,\bar{y})]$$