Statistical Learning Methods

- Statistical Learning based on the Learning of uncertainty in real environments.

- The methods probability and decision theory are used to handle uncertainty by the Agents

- First the agent must learn its probabilistic theories of the world from **experience**.

- A Bayesian view of learning is extremely powerful, providing general solutions to the problems of noise, overfitting, and optimal prediction.
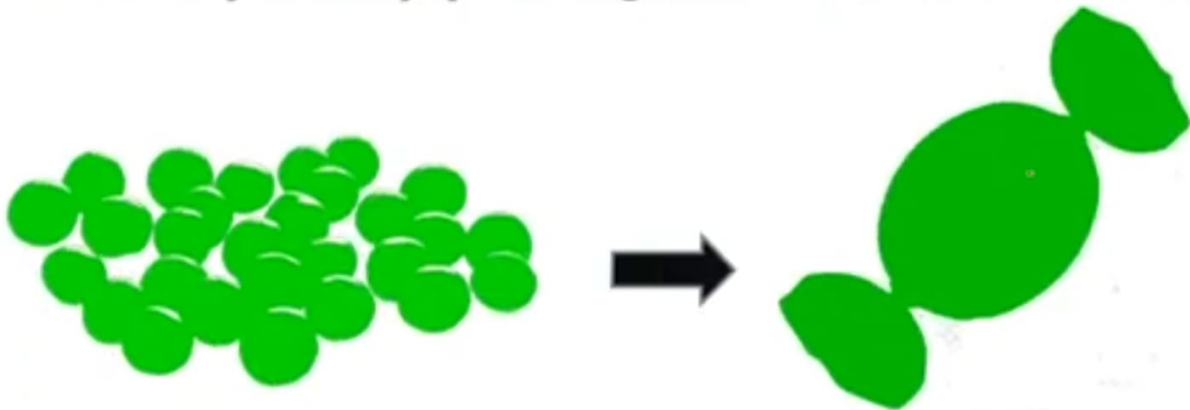
Candy bags Example
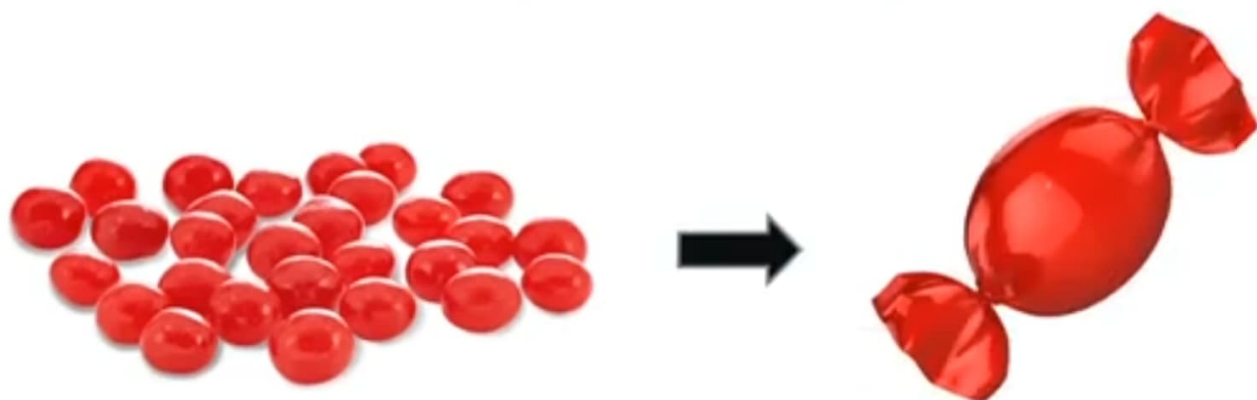
Normally candy packing is with same color wrapper

Candy bags Example

Normally candy packing is with same color wrapper

Candy bags Example

The <u>manufacturer wraps each piece</u>
of <u>candy</u> in the <u>same solid wrapper</u>, regardless of flavor

Candy bags Example

The candy is sold in very large bags
(Types of bags : 5)

Candy bags Example

| h1:
100% cherry | h2:
75% cherry
+
25% lime | h3:
50% cherry
+
50% lime | h4:
25% cherry
+
75% lime | h5:
100% lime |

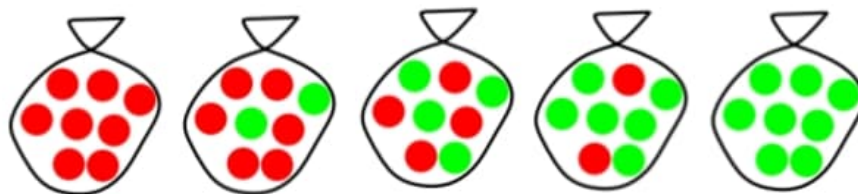Suppose there are five kinds of bags of candies:

   10% are h_1: 100% cherry candies
   20% are h_2: 75% cherry candies + 25% lime candies
   40% are h_3: 50% cherry candies + 50% lime candies
   20% are h_4: 25% cherry candies + 75% lime candies
   10% are h_5: 100% lime candies

Then we observe candies drawn from some bag: ● ● ● ● ● ● ● ● ● ●

- Expectation maximization (EM) algorithm simplifies difficult maximum likelihood problems. The EM algorithm can be used even for the variables whose value is not observable.

- Each iteration of the EM algorithm consists of two processes – the E-step, and the M-step.

- In the expectation, or E-step, the intention is to determine the missing values. The missing data is estimated based on the observed data. This is done given the observed data and the current estimate of the model parameters. For this expectation the conditional expectation explaining the choice of terminology is used.

- In M-step, the likelihood function is maximized under the assumption that the missing data is known. In place of missing data, the estimated data in E-step is used. Convergence is assured, since the algorithm is guaranteed to increase the likelihood during each iteration.

What is NLP?

NLP stands for **Natural Language Processing**, which is a part of **Computer Science**, **Human language**, and **Artificial Intelligence**. It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages. It helps developers to organize knowledge for performing tasks such as **translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction,** and **topic segmentation**.
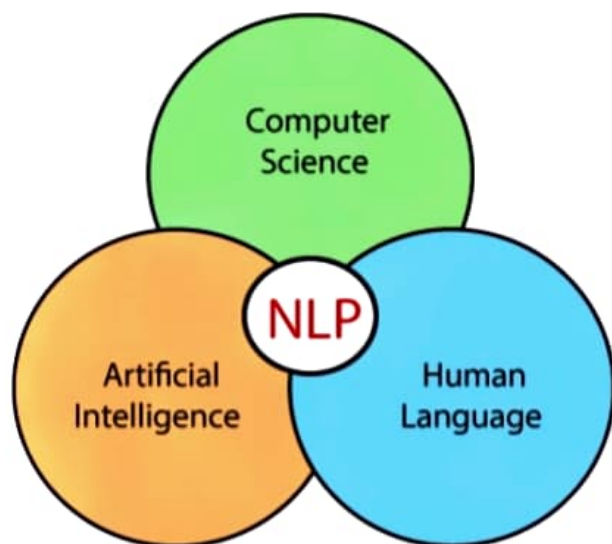
Components of NLP

There are the following two components of NLP -

1. Natural Language Understanding (NLU)

Natural Language Understanding (NLU) helps the machine to understand and analyse human language by extracting the metadata from content such as concepts, entities, keywords, emotion, relations, and semantic roles.

NLU mainly used in Business applications to understand the customer's problem in both spoken and written language.

2. Natural Language Generation (NLG)

Natural Language Generation (NLG) acts as a translator that converts the computerized data into natural language representation. It mainly involves Text planning, Sentence planning, and Text Realization.

> Note: The NLU is difficult than NLG.

Difference between NLU and NLG

NLU	NLG
NLU is the process of reading and interpreting language.	NLG is the process of writing or generating language.
It produces non-linguistic outputs from natural language inputs.	It produces constructing natural language outputs from non-linguistic inputs.

Applications of NLP
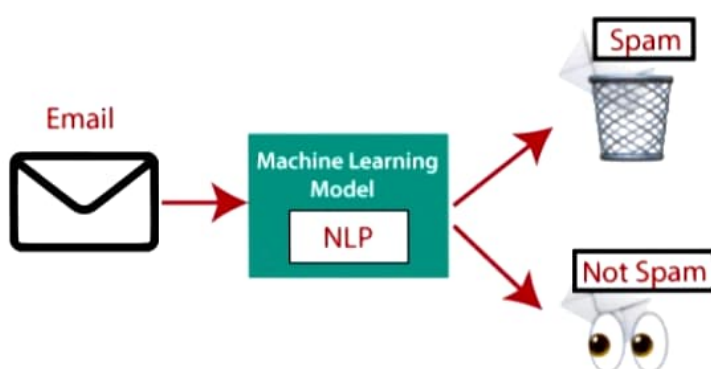
There are the following applications of NLP -

1. Question Answering

Question Answering focuses on building systems that automatically answer the questions asked by humans in a natural language.


2. Spam Detection

Spam detection is used to detect unwanted e-mails getting to a user's inbox.

3. Sentiment Analysis

Sentiment Analysis is also known as **opinion mining**. It is used on the web to analyse the attitude, behaviour, and emotional state of the sender. This application is implemented through a combination of NLP (Natural Language Processing) and statistics by assigning the values to the text (positive, negative, or natural), identify the mood of the context (happy, sad, angry, etc.)
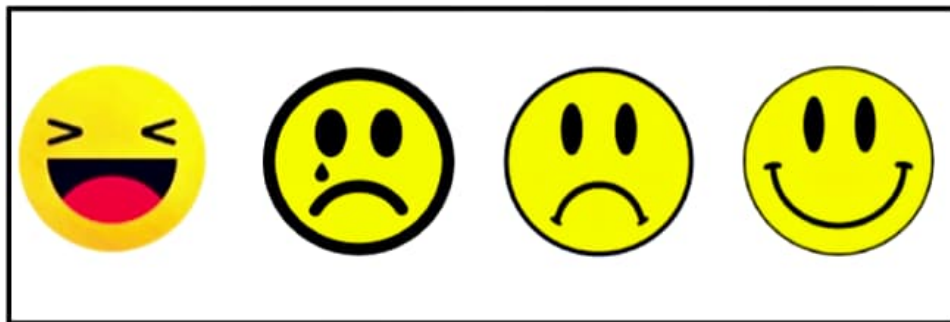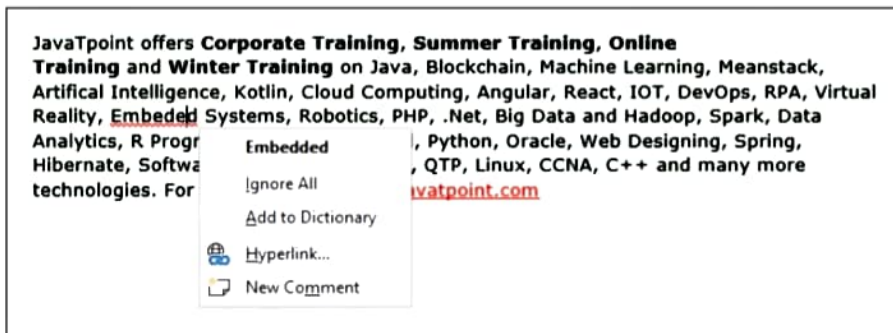

4. Machine Translation

Machine translation is used to translate text or speech from one natural language to another natural language.

5. Spelling correction

Microsoft Corporation provides word processor software like MS-word, PowerPoint for the spelling correction.


6. Speech Recognition

Speech recognition is used for converting spoken words into text. It is used in applications, such as mobile, home automation, video recovery, dictating to Microsoft Word, voice biometrics, voice user interface, and so on.

7. Chatbot

Implementing the Chatbot is one of the important applications of NLP. It is used by many companies to provide the customer's chat services.

Phases of NLP

There are the following five phases of NLP:
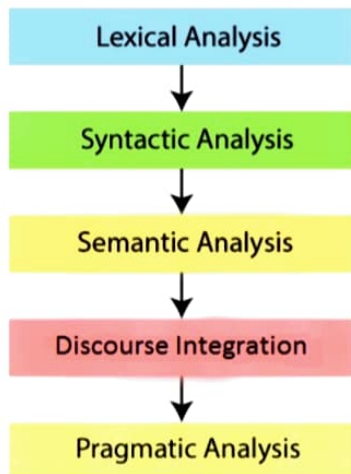
1. Lexical Analysis and Morphological

The first phase of NLP is the Lexical Analysis. This phase scans the source code as a stream of characters and converts it into meaningful lexemes. It divides the whole text into paragraphs, sentences, and words.

2. Syntactic Analysis (Parsing)

Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.

Example: Agra goes to the Poonam

In the real world, Agra goes to the Poonam, does not make any sense, so this sentence is rejected by the Syntactic analyzer.

3. Semantic Analysis

Semantic analysis is concerned with the meaning representation. It mainly focuses on the literal meaning of words, phrases, and sentences.

4. Discourse Integration

Discourse Integration depends upon the sentences that proceeds it and also invokes the meaning of the sentences that follow it.

5. Pragmatic Analysis

Pragmatic is the fifth and last phase of NLP. It helps you to discover the intended effect by applying a set of rules that characterize cooperative dialogues.

Advantages of NLP

- NLP helps users to ask questions about any subject and get a direct response within seconds.
- NLP offers exact answers to the question means it does not offer unnecessary and unwanted information.
- NLP helps computers to communicate with humans in their languages.
- It is very time efficient.
- Most of the companies use NLP to improve the efficiency of documentation processes, accuracy of documentation, and identify the information from large databases.

Disadvantages of NLP

A list of disadvantages of NLP is given below:

- NLP may not show context.
- NLP is unpredictable
- NLP may require more keystrokes.
- NLP is unable to adapt to the new domain, and it has a limited function that's why NLP is built for a single and specific task only.

Difference between Natural language and Computer Language

Natural Language	Computer Language
Natural language has a very large vocabulary.	Computer language has a very limited vocabulary.
Natural language is easily understood by humans.	Computer language is easily understood by the machines.
Natural language is ambiguous in nature.	Computer language is unambiguous.