NLP Questions in Machine Learning

Student's Name

Institutional Affiliation

Course Number

Course Name

Due Date

**NLP Questions in Machine Learning**

**Question 7:** **What is the difference between generative and discriminative classification? Explain the main differences and present examples.**

In machine learning, classification is the predominant task. The category may be done using two models; generative and discriminative. Generative classifiers are models which try to classify how data is positioned through the space and rotates around dataset distribution to return a probability. The generative models give explanations about data generation. On the other hand, discriminative classifiers are models that emphasize predicting the data labels based on conditional probability or either used for regression or classification. Discriminative models also draw boundaries in data space. Discriminative models are best termed for administered machine learning because they use probability estimates to create new instances. The decisive objective of the discriminative model is to divide two classes. Discriminative models in machine learning have various types: logistic regression, support vector machine, decision tree, and random forest. On the other hand, generative models in machine learning include the following examples: latent Dirichlet allocation, Bayesian network, hidden Markov model, and the autoregressive model.

**Question 8: Describe the different levels of ambiguity in human language and include at least one NLP task (e.g., POS tagging) that addresses each of them).**

Languages play an essential role in human life as they allow humans to record and transfer information daily. Let us imagine a world without communication how everything would be difficult. Nevertheless, natural language processing is not designed so perfectly to disseminate information as many could imagine. Notably, they may lead to a different interpretation of information, thus giving the listeners further understanding. Human language has different levels such as lexical ambiguity, structural ambiguity, scope ambiguity, and a combination of lexical and structural ambiguity.

Lexical ambiguity is a level of ambiguity where written text may result in various word meanings. Also, the spoken words may have the same sounds which result from different word forms. Structural ambiguity is a level of ambiguity where a sentence may have different construals because the terms used in those sentences are closely correlated. Scope ambiguity is the level of human language caused by the likelihood of diverse semantic choices in a sentence. It means that there is uncertainty in surface structure or single words. The last type of ambiguity is brought by the combination of both lexical and structural ambiguity. This level of ambiguity has both structural and lexical properties. In other words, there are not only vague terms used, but the sentences have syntactic structure. NLP in lexical category disambiguation can be used to overcome Lexical ambiguity. Parts of speech tagger (POS Tagger) is a piece of software that can resolve structural and scope ambiguity because the software reads the text in some language and dispenses each word to part of speech.

**Question 9: What is the formal definition of *grammar* in the context of natural language parsing systems?**

Grammar in the context of a natural language parsing system is defined as a rule for establishing sentences that are well structured. Thus, grammar plays an essential role while describing the syntactic structure of well-structured programs. Furthermore, in natural language parsing systems, grammar outlines syntactic guidelines that ought to be used for conversation. Thus, operating statements and lists in the C programing language, grammar rules state how the machine's functions are structured.

**Question 10: Describe the main differences between *word embeddings*, *n-gram language models*, and *a bag of words.*Provide at least one NLP application in which each of these models would be best suited for (3 applications in total, one for each model)**

Word embeddings in Natural Language Processing represent words that allow words to make similar meanings to exhibit equal representation. Typically, this occurs in real-treasured vectors that translate close related words to get similar denotation. Word embeddings in Natural language processing are used in analyzing survey responses. N-gram language model in natural language processing is the model that envisages the prospect of a given N-gram in the most miniature system of words in the language. N-gram language models are used in natural statistical processing for speech recognition. Bag of words in natural language processing refers to models which represent simplified representation of terms and information. Technically, a bag of words can be referred to as a feature extraction text data model because the method discards documents with any information about the structure or order of words. In natural language processing, a bag of terms is used to retrieve data from records.