Machine Learning and Natural Language Processing Interview Questions

April 14, 2016

- What is POS (Part of Speech) tagging? What is the simplest approach to building a POS tagger that you can imagine?

- How would you train a model that identifies whether the word "Apple" in a sentence belongs to the fruit or the company?

- Design a recommendation system.

- What is latent semantic indexing and where can it be applied?

- How would you build a system that auto corrects text that has been generated by a speech recognition system?

- What is regular grammar? Does this differ in power to a regular expression and if so, in what way?

- How would you build a POS tagger from scratch given a corpus of annotated sentences? How would you deal with unknown words?

- How would you find all the occurrences of quoted text in a news article?

- How would you build a system to translate English text to Greek and vice-versa?

- How would you build a system that automatically groups news articles by subject?

- What are stop words? Describe an application in which stop words should be removed.

- How would you design a model to predict whether a movie review was positive or negative?

- What is entropy? How would you estimate the entropy of the English language?

- What is the TF-IDF score of a word and in what context is this useful?

- How does the PageRank algorithm work?

- What is dependency parsing?

- What are the difficulties in building and using an annotated corpus of text such as the Brown Corpus and what can be done to mitigate them?

- Are you familiar with WordNet or other related linguistic resources?

- Do you have any experience in building ontologies?

- Do you speak any foreign languages?

- What tools for training NLP models (nltk, Apache OpenNLP, GATE, MALLET etc.,) have you used?

- Describe approaches to data mining of customer logs.

- What is feature extraction and how would you perform it?

- How would you handle dirty data?

- How do you build a system similar to Netflix?

- How do you implement convolutional neural networks?

- Provide details about LDA and Gibbs Sampling?

- What do you mean by Rejection Sampling?

- Describe a Random Forest.

- Why does one use MSE as a measure of quality. What is the scientific/mathematical reason for the same?

- Given an objective function, calculate the range of its learning rate?

- What is deep learning and what are some of the main characteristics that distinguish it from traditional machine learning?

- What is linear in a generalized linear model?

- Give an example of an application of non-negative matrix factorization

- How would you evaluate the quality of the clusters that are generated by a run of K-means?

- Do you have experience with Spark ML or another platform for building machine learning models using very large datasets?

- What tools and environments have you used to train and assess models?

- What are some good ways for performing feature selection that do not involve exhaustive search?

- On what type of ensemble technique is a random forest based? What particular limitation does it try to address?

- What is a probabilistic graphic model? What is the difference between Markov networks and Bayesian networks?

- What is the EM algorithm? Give a couple of applications

- What methods for dimensionality reduction do you know and how do they compare with each other?

- Explain NLP to a non-technical person.

- What's the use of NLP in machine learning?

- Explain Vector Space Model and its use.

- Explain cosine similarity in a simple way.

- How unstructured text data can be converted into structured data for the purpose of ML models?

- Why and when stop words are removed? In which situation we do not remove them?

- Explain the distances and similarity measures that can be used to compare documents.

- Explain Linear and Logistic Regression and in which scenarios you can use them?