Large Language Models

Part 1: Background and Basics

Language models (LMs) aim to model the intrinsic patterns of human language to predict the probability of future (or missing) words or tokens in a sequence.

Statistical Language Models (SLMs): Emerged in 1990s, based on Markov assumptions to predict the next word using a fixed-length prefix, N-gram Model, w/ limitation in sparsity (smoothing) and long-range dependencies (curse of dimensionality). Key papers:

Jelinek, F. (1990). "Self-organized language modeling for speech recognition."

Katz, S. M. (1987). "Estimation of probabilities from sparse data for the language model component of a speech recognizer."

Neural Language Models (NLMs): Distributed Word Representation (word2vec), RNN, LSTM, GRU

Pre-trained Language Models (PLMs): Transformer, BERT, GPT

Large Language Models (LLMs)