syllabus

schedule

WEEK 1

Introduction and Applications

January 9

- Language the most efficient and compact way to transfer knowledge is through words, where the window to AGI is through NLP. This lecture is an introduction that takes us through history of how we got to LLMs. We'll also review some applications of NLP, current industry standards, and some of the most impactful approaches and where they are being implemented. Finally, we'll preview what we'll be learning, the logistics of how we'll be doing so, and the expectations for your participation in this class.
- Applications Overview
 - Machine Translation (Baidu's Word-Word)
 - Summarization (Dialogues, Newspaper Articles, etc.)
 - Text Classification and Clustering (News Article Groupings, etc.)
 - Question and Answering (LLMs and Chatbots)
- Submissions
 - Laboratory Getting Started on Google Cloud with Your Credits
 - Assignment 0 is assigned A First Look at Processing Language

WEEK 2

ML Foundations and Software Engineering

January 16

- As NLP is a specific branch of machine learning, we will review some foundational knowledge that we'll utilize through the course of this class. We'll look at both machine learning and software engineering best practices that will help you build and scale NLP systems later in the course. Because most NLP algorithms today rely heavily on computing resources, we'll dive into distributed compution approaches and cloud-based operations.
- Lecturing Topics
 - Foundations of Machine Learning
 - Software Engineering Practices
- Required Keynote Reading
- <u>Submissions</u>
 - Laboratory Containerization in the Cloud
 - Assignment 1 is due
 - Assignment 2 is assigned Text Classification

WEEK 3

Language Classification

- Building upon our review of machine learning, we discuss strategies in feature extraction and generation. Particularly as creating a vocabulary can explode required memory space, our featurization includes NLP-specific techniques (e.g., tokenization, lemmatization, etc.).
- Lecturing Topics
 - Building Vocabulary with Stopwords and Stemming
 - Preprocessing Tokenization and Lemmatization
 - Logistic Regression Classifier
 - Naïve Bayes Classifiers
- Application Sentiment Analysis
- Submissions
 - Laboratory Naïve Bayes

WEEK 4

January 30

- One of the most widely used algorithms in practice today are autocorrecting algorithms that typically have on-device requirements. In this lecture, we'll review elements of dynamic programming, particularly with respect to the minimum edit distance algorithm, and how we can apply these concepts to the autocorrect and subsequently the autocomplete problem.
- Lecturing Topics
 - Representations of Language
 - Comparisons / Differences in Language
 - Minimum Edit Distance Algorithms
- Application Autocorrect in Practice
- Submissions
 - Laboratory Autocorrect Vocabulary Candidates
 - Assignment 2 is due
 - Assignment 3 is assigned Autocorrect and Minimum Edit Distances

WEEK 5

Introduction to Language Modeling

February 6

- Lecturing Topics
 - What is a language model? (Abstractive vs extractive approaches)
 - Overview of Basic Modeling Approaches
 - The N-Gram Model
 - Out of Vocabulary Words and Smoothing
 - Language Model Evaluation
- Application Autocompleting words and sentences
- Submissions
 - Laboratory Jupyter Notebooks with GPUs
 - Assignment 3 is due
 - Assignment 4 is assigned Autocomplete with Topical Information

WEEK 6

Unsupervised NLP - Topic Modeling

February 13

- This week, we will explore David Blei's contributions to the field, a set of concepts that indirectly attack the age-old question of **"what is** *k* **in the k-means clustering algorithm**. We will review the hierarchical nature of how to model natural language using Bayesian concepts, where our corpora is processed without preserving the order of words. This week also marks the first week of **required keynote paper** reading, where we will begin the tour of seminal papers that have revolutionized not only language processing but also machine learning and artificial intelligence writ large. This reading is perhaps the most difficult one that you'll read in this class, since it involves a heavy component of probability and statistics.
- Lecturing Topics
 - Parameter Estimation of a Distribution
 - The Dirichlet Distribution and its Attributes
 - Infinite Bayesian in Topic Models
 - Latent Semantic Indexing and Latent Dirichlet Allocation
 - (Collapsed) Gibbs Sampling, and Optimization
- Required Keynote Paper Latent Dirichlet Allocation
 - David Blei's Lecture
 - Introductory Blog to Topic Modeling
- Submissions
 - Laboratory Topic Modeling

WEEK 7

Word Modeling with Self-Supervision

February 20

Perhaps the most influential paper to have come out of the natural language community is the <u>word2vec paper</u> that most general machine learning practioners recognize. You'll find elements of its practice in communities from the information retrieval sciences to modern cyber applications to general ML problems. As it pertains to language models, modeling words is often the first stage in any system pipeline that you may design. This week's lecture reviews word models (including word2vec as well as continuous bags of words) and the embeddings / representations that they create
© 2025 Natural Language Processing CS 6120. Northeastern University, San Jose Campus, <u>Karl Ni</u>. Powered by <u>Jekyll</u> with <u>al-folio</u> theme.

- Embeddings with Continuous Bag of Words
- Intrinsic and Extrinsice Evaluation of Word Models
- Word Modeling in Practice
- The Skip-gram and Negative Sampling
- From Words to Sentences
- Required Keynote Paper Distributed Representations of Words
 - Blog Gentle Introduction to Negative Sampling
- Submissions
 - Assignment 4 is due
 - Assignment 5 is assigned Word2Vec Skipgram Implementation

WEEK 8

Introduction to Sequential Modeling

February 27

- Topics
 - Modeling with Hidden Markov Models
 - The Viterbi Algorithm Initialization, Forward, and Backward Passes
- Application Parts of Speech Tagging

WEEK 9

Recurrence and Neural Networks

March 6

- Lecturing Topics
 - Traditional Language Models vs Recurrent Models
 - The Recurrent Neural Network
 - Vanishing and Exploding Gradients
 - Memory Gating GRUs and LSTMs
 - Accuracy and Evaluation Perplexity
- Applications Named Entity Recognition and Machine Translation
- Required Keynote Paper Long Short Term Memory Networks
- Required Keynote Paper On the Difficulty of Training RNNs
 - Karpathy's Blog on Recurrent Networks
- <u>Submissions</u>
 - Assignment 5 is due
 - Assignment 6 is assigned Implement Your Own Recurrent Network

WEEK 10

No Instruction - Spring Break

March 13

• Have a nice holiday!

WEEK 11

Siamese Networks and Embedding Hashing

March 20

- Lecturing Topics
 - Siamese Architectures and Shared Parameter Networks
 - The Triplet Data Objective
 - Transforming and Retrieving Vectors
 - Locality Sensitive Hashing, Quantization, and Approximate Nearest Neighbors
- Application Sentence Comparisons and Document Retrieval
- Required Keynote Paper Text Similarity with Siamese RNNs
 - Video Andrew Ng's Siamese Networks (for Images)
- <u>Submissions</u>
 - Assignment 6 is due
 - Assignment 7 is assigned Coding Up Attention Networks

WEEK 12

Attention and the Transformer Model

March 27

- Attention models have been the leap forward that are the fundamental building blocks to modern machine learning today, including the essential ingredients for Large Language Models. We'll go deep into attention layers in neural networks, building our own from scratch.
- Lecturing Topics
 - Introduction to the Attention Modeling
 - The Self-Attention Mechanism
 - The Transformer Modeling Layer
 - Large Scale Attention Modeling
- Required Keynote Reading Attention is All You Need
- Required Keynote Reading BERT Pre-training Bidirectional Transformers
 - BERT Explained State of the art in NLP, Blog
 - Attention Paper Explained

WEEK 13

Introduction to Large Language Modeling (LLMs)

April 3

- The next three weeks are devoted to the state of the art in industry, and LLMs in practice, which may have changed in the time that you have started this course! This week, we introduce large language models using the fundamentals that you have learned, from perplexity in system design to transformer neural network layers for pre-training. We'll focus on techniques that large companies (or well-funded ones, at least) use to create *foundation* LLM models, taking training methods from OpenAI, Anthropic, Amazon, and Google.
- Lecturing Topics
 - Large Language Modeling (LLM) in Code
 - Deep Reinforcement Learning from Human Feedback
 - Alternatives to RLHF Partial Policy Optimization
 - Aligning LLMs in the Instruction Following Framework
 - Multimodal Large Language Models
- Required Keynote Reading Training to Instruct with Human Feedback
- Required Keynote Reading GPT-4 Technical Report from OpenAl
- <u>Submissions</u>
 - Laboratory Tuning LLMs
 - Assignment 7 is due
 - Assignment 8 is assigned Implementing Your RAG System

WEEK 14

Practically Leveraging Large Language Models

April 10

- Last week, we discussed how large companies might train LLMs. In contrast, this week's lecture is most useful for those interested in entering the industry at the mid- to startup levels, where we explore common approaches to optimally *leverage* large language models for your particular applications once the LLM has been created. These techniques additionally attack limitations in LLMs, such as knowledge gaps, hallucinations, and logical reasoning problems.
- Lecturing Topics
 - Prompt Engineering Query and Context
 - Retrieval Augmented Generation (RAG)
 - Tuning with Low Resources LoRA and Quantization
 - Intelligent Agents with Program-Aided LLMs
- Required Keynote Reading Retrieval Augmented Generation
- Required Keynote Reading A Survey of LLMs Including ChatGPT and GPT-4
- <u>Submissions</u>
 - Laboratory Instruction Following Tuning

WEEK 15

Language Modeling Systems Lifecycle

April 17

- You've learned about the inner workings of the LLM, the mechanisms that power it, how and when to tune it, the data collection processes that govern it, how it can be used practically, and the agents that can run on it. In this lecture, we explore the practical aspects of GenAI engineers when product managers ask them to design a system for them. More than the theory, we'll learn about the system itself, devoting time for *when* to focus on certain components of your LLM, and the life cycle of your system design.
- Lecturing Topics

- From Prompting to Fine Tuning to PEFT
- Applications Creating Your Own GenAl Smart Agents
- <u>Submissions</u>
 - Assignment 8 is due

WEEK 16

Demonstrations and Poster Sessions

April 24

• Deploy and show off your domain-specific LLM and pitch your startup idea!

grading criterion

Online Quizzes	10%
Reading and Discussion	15%
Labs	25%
LLM Deployment Project	25%
Assignments	25%

course meeting times

Lectures

- Thurs, 4pm-7:20pm
- Room TBD

Office Hours

- Professor, Thurs, 8:30-9:30pm
- TA, Date/Time TBD

suggested textbooks

Speech and Language Processing, 3rd Ed. Dan Jurafsky and James Martin, 2024

A Comprehensive Overview of Large Language Models, Naveed et. al., 2024