

Review!

Quick Review

- **What did we do this semester?**
 - We'll go over the topics and review some potential questions.
- **What should I expect in the exam?**
 - 10 short true/false + explain, 3 longer questions.
- **How should I study?**
 - Review the materials, "debug" your understanding, use the make up exam (without looking at the answers).
- **What can I bring?**
 - Easy – cheatsheet, calculator, pens, yourself.

How should I study?

- In general, we don't aim to trick you. We want to see if you internalized the key concepts, you feel comfortable with the formalism expressing these concepts, and understand the conditions where they are applicable and the tradeoffs between them.
- This is **not** a math test. No proofs, minimal calculations.
- This is **not** a memory competition. Don't memorize formulas, instead make sure you understand their "internals"
- **Actively** look for gaps in your understanding.

Part 1: Intro to NLP

- Text classification
 - What is text classification? What kind of problems can we model as classification?
 - What is a feature representation? What tradeoffs are involved?
 - What is a loss function? How do we optimize? Min or Max?
- Neural Networks basics
 - Linear vs. non-linear classifiers.
 - Multilayer NN, backprop
 - RNN, LSTM, attention, transformer
 - Basics of representation learning

Part 2: CL

- CL vs. NLP. Why should we care about CL?
- CL pipeline: morphology, syntax, semantics, pragmatics.
- Syntax, grammars, CFG, CNF, CYK
- Structured prediction basics, branch and bound, dynamic programming, sequence prediction, graph/tree prediction.
- Semantics, lexical and compositional
- Pragmatics and discourse structure, coreference and coherence.

Part 3: LLM/Agents

- LLM: what is language modeling, N-gram vs. transformer based LMs
- Scaling
- Prompting, in-context learning, reasoning.
- RAG systems and Agent architectures
- Efficiency consideration, LoRA, quantization, pruning distillation
- RL