

CS 533: Natural Language Processing

# Logistics

Karl Stratos



Rutgers University

# Course Information

- ▶ **CS 533: Natural Language Processing (NLP)**
  - ▶ Wednesday 12:00–3:00pm at Beck Hall 252
  - ▶ All materials at course website: <http://karlstratos.com/teaching/cs533spring20/cs533spring20.html>
  
- ▶ Instructor: **Karl Stratos** (legal name: **Jang Sun Lee**, or **Jangsun Lee**)
  - ▶ Office Hours: Wednesday 3:20–4:30pm at Tillett 111H
  
- ▶ TA: **TBD**
  
- ▶ Use **Canvas**:  
<https://rutgers.instructure.com/courses/44246>
  1. To ask questions regarding lectures/homeworks/projects (and answer yourselves), discuss ideas, find collaborators, etc.
  2. To submit assignments
  3. To find announcements

# About

This course is *not* about

- ▶ Philosophy of language
- ▶ Linguistic phenomena
- ▶ Social impact of language

This course *is* about

- ▶ **Models, statistical techniques, and algorithms** for computationally processing language as **data**

# Grading

- ▶ Project: 40%
  - ▶ Written report: 30%
  - ▶ Presentation: 10%
- ▶ Exam (in-class and open book): 30%
- ▶ Assignments: 20%
- ▶ Participation: 10%

# Project

- ▶ Submit a **proposal** by later in the semester, do the work, and submit a **final report** and give an in-class **presentation**
- ▶ Has to be
  1. **Substantial**: cannot be done trivially in a few hours  
(e.g., nontrivial implementation and experiments)
  2. **Original**: has not been done already  
(e.g., new problem formulation, techniques, applications)

Ideally the quality of conference papers in NLP

- ▶ More information to come
  - ▶ But start thinking about projects early on

## ▶ The main way to learn

- ▶ Tentative plan:  $\approx$  4 assignments, each a mix of written problems and coding in Python
- ▶ A1 is already out (due 2/4 11:59pm)
  - ▶ If you cannot do A1 comfortably, you probably do not have the background needed for this course.
- ▶ Work individually (okay to discuss). Do not copy:
  1. Honor
  2. Meaningless (in-class exam)

# Exam

- ▶ Topics covered in lectures, readings, and assignments
- ▶ Tentative date: April 1
- ▶ You will be fine if you understand lectures and readings, and do well in assignments
  - ▶ I will tell you what will be on the exam
  - ▶ Definitely prioritize your research project over the exam

# Tentative Schedule

Date	Topics	Readings	Assignments
Week 1 (January 22)	Introduction to NLP, Language Modeling	Michael Collins notes on $n$ -gram models and log-linear models	A1 [code] (Due 1/31)
Week 2 (January 29)	Deep Learning for NLP: Neural Language Modeling	Colah's blogs on deep learning and LSTMs, NLM papers using feedforward (Bengio et al., 2003), recurrent (Mikolov et al., 2010; Melis et al., 2018), and attention-based (GPT-2) architectures	
Week 3 (February 5)	Deep Learning for NLP: Conditional Neural Language Modeling		
Week 4 (February 12)	Deep Learning for NLP: An Overview of Other Techniques and Applications		
Week 5 (February 19)	Structured Prediction in NLP: Tagging		
Week 6 (February 26)	Structured Prediction in NLP: Parsing		
Week 7 (March 4)	Unsupervised Learning in NLP: Latent-Variable Models and the EM Algorithm		
Week 8 (March 11)	Unsupervised Learning in NLP: Variational Autoencoders		
Spring Recess			
Week 9 (March 25)	Unsupervised Learning in NLP: Pretrained Neural Text Representations		
Week 10 (April 1)	Exam		
Week 11 (April 8)	Special Topics: TBD (Information Extraction, Question Answering)		Project proposal due
Week 12 (April 15)	Special Topics: TBD (Dialogue, Grounding)		
Week 13 (April 22)	Special Topics: TBD (Maximal Mutual Information Representation Learning)		
Week 14 (April 29)	Project Presentations		



## Other Requirements

- ▶ Use LaTeX to write up assignment and project reports
  - ▶ Template for assignment: [http://karlstratos.com/teaching/cs533spring20/assignment\\_template.tar.gz](http://karlstratos.com/teaching/cs533spring20/assignment_template.tar.gz)
  - ▶ Template for project: <http://karlstratos.com/teaching/cs533spring20/latex-example.tar.gz>
  
- ▶ Use Python and its libraries like NumPy, PyTorch for coding

QUESTIONS?